AN EFFICIENT INTELLIGENT ANALYSIS SYSTEM FOR CLASSIFYING X-RAY IMAGES OF BODY PARTS

Student Number: 2040636

Abstract: An X-ray classification algorithm aids clinical care and treatment by enhancing faster classification and error free interpretation of X-ray images, thereby conserves time for other important tasks. The challenge is to build a classifier that can classify any given X-ray image. The provided dataset consists of 6 categories of X-ray images namely: AbdomenCT, BreastMRI, CXR, ChestCT, Hand and HeadCT. Since the dataset is large enough and sufficient for training a Deep Learning model, a Deep Convolutional Neural Network that automatically extracts image features (with optimal representation) is used to reduce the model computational time, thus increasing efficiency. The built model classifies a given X-ray image of body parts into one of the 6 classes with up to 99.95 % accuracy on test set.

1. Introduction

The classification of digital images made possible by Artificial Intelligence has led to great and unlimited improvement in the health industry. Effective classification of medical images, for instance X-rays plays a major role in aiding clinical care and treatment since vast number of X-ray images are taken every day at hospitals around the world and there is not to recognise and differentiate them (WHO, 2001). This is because building a X-ray classification algorithm ensures faster and superior classification limiting human errors due to interpretation (Sirinukunwattana, 2016; Xie, 2017). This helps the clinicians to conserve time for other important tasks. X-rays are used to examine most body parts to detect defects in bones, joints or soft tissues such as internal organs.

The ability to classify different images as belonging to different parts of the body will bring about greater efficiency to clinicians in their ability to detect body defects or diagnose disease before greater damage is caused to a tissue or organ. Thus, this work attempts to build a deep learning model that identifies and classifies 6 classes of X -ray images from different body parts namely: AbdomenCT – the computed tomography of the abdominal cavity, BreastMRI – the MRI of the breast, CXR - the chest X-ray, ChestCT – the computed tomography of the chest, the Hand - hand X-ray and HeadCT - the computed tomography of the head.

The methodology involved is presented in chapter 2, followed by the model simulation detailing the algorithm used in chapter 3. The results and its critical analysis are presented in chapter 4 followed by conclusions and references in chapter 5 and 6 respectively.

2. Methodology

2.1 Neural Networks (NNs) and Deep Neural Networks (DNNs)

Neural Networks mimic the brain and its learning processes. Usually, Neural Networks have 3 layers; input layer that receives input ($28 \times 28 \times 1$ images), a middle-hidden layer that processes the input using certain weights and bias, and an output layer that generates the processed information as shown below (Figure 1a & 1b).

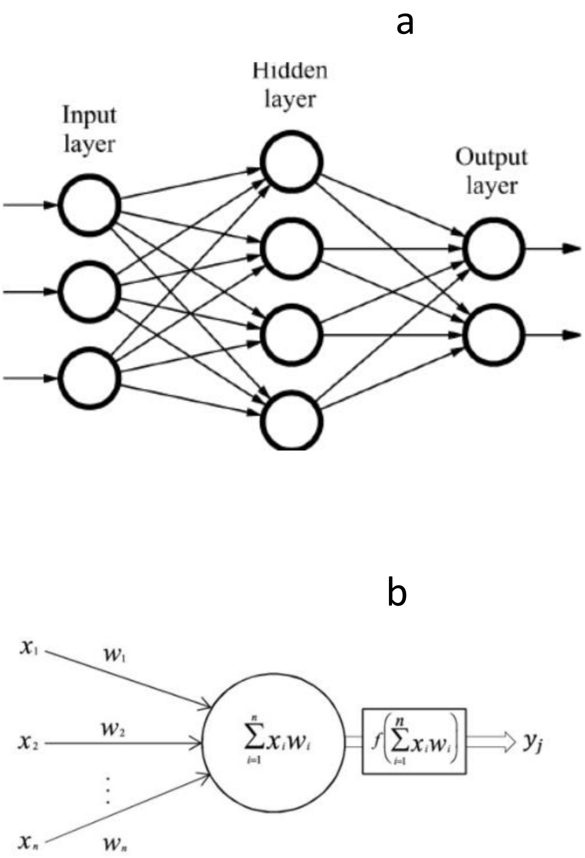

Figure 1: **a** - A structure of a Neural Network with 2 layers – excluding the input layer; **b** – Diagram and equations – how NN combine weight and bias with input to generate output in a 2 – layer NN (Source: Vieiraa, 2017).

Neural Network is a parallel, connected and tuneable processing computation system which consist of many processing units or nodes called neurons. A neural network is a linear combination of many layers, for instance, Multiple Layer Perceptron establishing a mathematical operation between the output of the previous layer and the weights (vectors) of the current layer. Afterwards, it passes the output of mathematical operation to the next layer through an activation function as shown below (Figure 1b & 2).

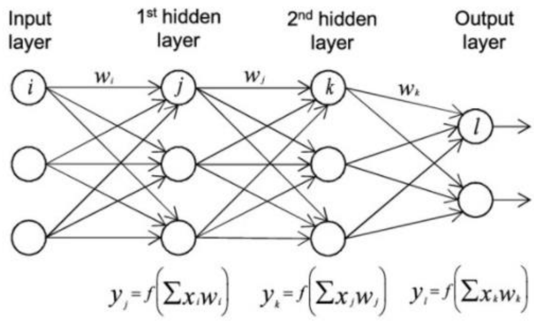

Figure 2: Structure of a Deep Neural Network (Source: Vieiraa, 2017)

The middle-hidden layer can have many layers, thus referred to as a Deep Neural Network (Hinton et al., 2006). The core of all Deep Learning (DL) algorithms or models are Neural Networks. DL involves training and testing of multi-layered neural networks that attain high-level abstraction by learning complex structures. DNN is a combination of non-linear processing layers which uses simple elements called ‘neurons' operating in parallel elements connected together in a specific way in order to perform a particular task.

Previously, for medical image classification, traditional machine learning methods such as Support Vector Machines (SVMs) are long used. Moshe, 2015 used SVM, K-Nearest Neighbour, Linear regression and Deep Belief Network to classify ImageCLEF dataset. However, they are slow and consume time during feature extraction, varying with different objects, thus perform below standard on medical image classification (Kerman, 2018). DNN, especially the convolutional neural networks have achieved significant performance in image classification since

2012, so now widely used (Rawat, 2017). Yadav, 2019 used CNN based algorithm on a chest X-ray dataset to classify pneumonia using 3 different techniques and found that data augmentation performed best, transfer learning better and SVMs least.

As a result, DL-based approaches surpass the state-of-the-art in a number of medical image analysis tasks in the field. Nevertheless, there exist great challenges and issues with small medical image datasets and anatomical variations (annotation) which restrict medical image classification effectiveness of classifying medical images (Weese, 2016). Improvement in easy access to massive datasets, computing power and availability of pre-trained models built by expert have dramatically improved deep learning models.

Most importantly, the classification and detection (Hinton, 2006; Bengio, 2009) as well as segmentation (Shen, 2017; Li, 2017) of medical image have a boosted performance because DL models have superior capabilities to automatically extract features from images, thus overcoming the need for manual feature. A study of colon cancer classification using histopathology images that adopts convolutional neural network to reduce manual annotation and produce good feature representations is presented (Xu, 2014). Vieiraa, 2017 supported DL as having promising potential in developing diagnostic and prognostic biomarkers of psychiatric and neurologic disorders.

A study shows Convolutional Neural Networks are widely used in the ImageNet Challenge with various combinations of datasets of sketches (Eitz, 2012). Yang, 2015 showed that CNN (74.9 % accuracy) outperformed average human being (73.1 % accuracy) on image datasets but not a trained network (64 % accuracy).

However, the deep learning model has not performed more than the best human expert. Thus, AI technology currently should complement clinicians and not used as a replacement

Therefore, the work on getting more quality larger data must be ongoing to make AI outperform the best human experts for medical image classification. This study attempt to work on more quality data to classify X-ray images of different body parts.

2.2 Learning Process of a Deep Neural Network

Each of the elements in the above layers are called neurons. The learning or training process of a neural network is an iterative process adjusting or updating interconnections between the artificial neurons within the network likened to human brain (Bengio, 2009).

The learning or training process of a neural network involves two steps which are forward and backward propagation. During forward propagation, images are fed into the input layer in the form of numerical numbers, denoting the intensity of pixels in the image. The backward propagation compares the output (generated prediction) with the actual value and gives an error. The calculated error is then used to update the values of the parameter and another forward propagation is repeated, followed by backward propagation until there is no error or drastically minimised.

3. Simulations

3.1. Dataset Description and Encoding

The dataset consists of X-ray images of 6 different body parts namely: AbdomenCT – the computed tomography of the abdominal cavity, BreastMRI – the MRI of the breast, CXR - the chest X-RAY, ChestCT – the computed tomography of the chest, the Hand - hand X-RAY and HeadCT - the computed tomography of the head. The dataset has 58 954 images in total and each class contains 10 000 images except the BreastMRI which has 8 954 images.

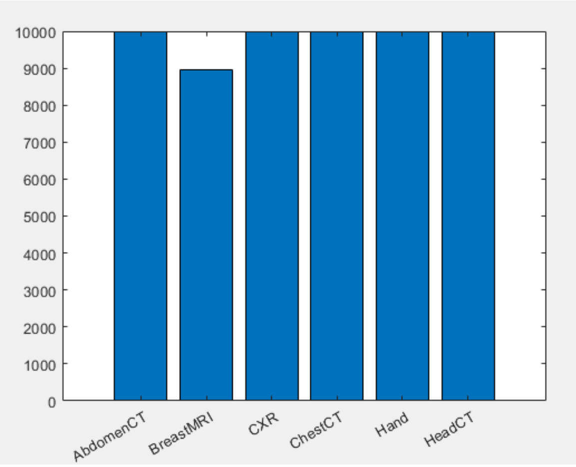

Figure 3: A bar chart plot showing the number of images in each class contained in the dataset.

The dataset has been gathered images from several sets at TCIA, the RSNA Bone Age Challenge, and the NIH Chest X-ray dataset. The image is a 2-dimensional greyscale image of different sizes. A plot of some of the images is shown below.

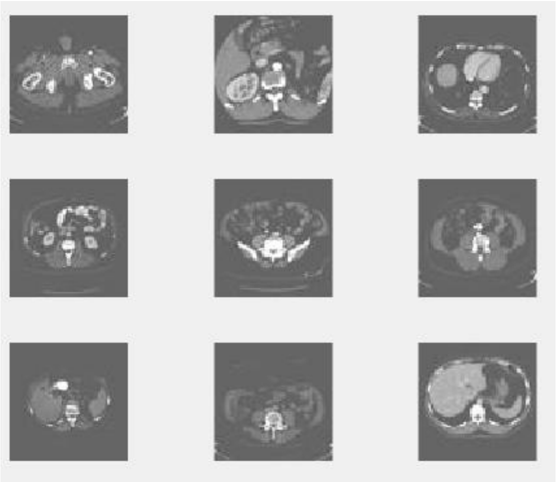

Figure 4: A plot of the resized X-ray images (28 by 28)

The dataset was loaded as an image datastore object and resized to 28 x 28 x 1-size the convolutional neural network takes as input. Next, the images are converted to greyscale since they show varying intensities.

3.2. Training, Validation and Test Sets

Overfitting is a global phenomenon that creates bias in a model when the training set has a large influence on the accuracy of a network (Lawrence, 2000). This is because a model may have millions of free parameters and quite possible for it to overfit to the data provided. This means that the model may adjust its weights to the precise values needed to predict every given image correctly but fail to recognize even slight variations on the original images not to talk of new images. Therefore, it is important building an architecture that supports maximum training as well as recognition performance.

A common solution is to split into the training set, the validation test and the test set in the ratio 8:1:1 respectively so that more data is used for the training. The training set is used for training the model so that the model learns and extracts all the features in the image for each class separately and different from the other classes. The validation set is used to test the model during training to know when to stop training and prevent overfitting. The test set is used to test the model performance after training to know if the model would perform on unseen data.

3.3. Model Architecture

The Neural Network used here is a Convolutional Neural Network (CNN) is made up of many hidden layers (Deep Neural Network) which is inspired by the organisation of the visual cortex. It is chosen because it is sensitive to signal-translation problem by using its learnable kernel (also called filter or detector) to convolve each input signal. The layers present in the Convolutional Neural Network include the input layer, the feature detection layer and the classification layer sequentially (Figure 5). The feature detection layer consists of the convolutional layer, the pooling layer and the batch normaliser layer. The classification layer comprises the flatten layer, fully connected layer and the Softmax layer.

3.4. The Input Layer

The input layer takes in an input image. The resized image of dimension 28 × 28 ×1 is passed into the feedforward network. This implies that the network would use 28 × 28 numbers of neuron (784).

3.5. The Convolutional Layer

The convolutional layer is the first layer of the feature extraction layer. It applies a learnable filter or kernel on the input image to activate the images and learn the patterns in the images such as edges, corners, arcs or regions. The convolutional layer extracts the feature of image by assigning importance; learnable weights and biases to various aspects/objects in the image and convert it into lower dimension without losing its characteristics. This enables the CNN to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture fits better to the image dataset due to the reduction in the number of parameters involved and reusability of weights.

In this study, 3 convolutional layers are used with a 3 by 3 filter. The first, second and third convolutional layers have 8, 16 and 32 neurons respectively that learn the features of different regions of the images. The PaddingMode is ‘same’ so that the output of the convolutional layer has the same size as the input (output of input layer).

3.6 batch normalization and ReLu Layers

Batch normalisation layer scales the input from the convolutional layer. It normalizes a mini-batch of data across all observations for each channel independently. This is necessary to speed up training of the convolutional neural network and reduce the sensitivity to network initialization.
The next layer, the ReLu activation function makes all negative pixel values in the output of normalized image zero so

that the values range from zero to positive value.

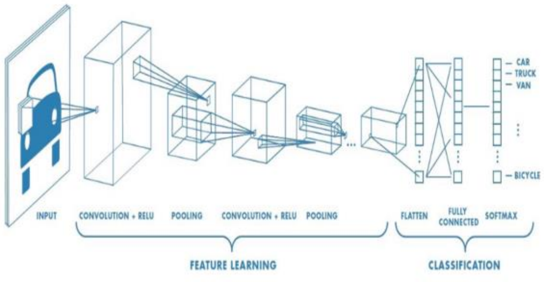

Figure 5: A diagram showing a Deep Convolutional Neural Network composed of the input layer, the feature extraction or leaning layer and the classification layer (Source: towardsdatascience.com)

3.7. The Pooling Layer

Pooling is the next process after the convolution and normalisation. It reduces the spatial size of the convolved feature to decrease the required computational power required to process the data using dimensionality reduction. Additionally, it is also used for extracting dominant rotational and positional invariant features, thereby maintaining model training process effectively. Maximum pooling (Sudholt, 2016) which returns the maximum value from the segment of the image masked by 2 x 2 filter is used. This is because maximum pooling also performs a noise suppressant.

Here, a 2 by 2 stride has been used which reduces the image to half of its size therefore reducing complexity of the images so that the training time and required number of neurons is also reduced considerably.

The next process is to flatten the final output and feed it to a regular Neural Network for classification purposes.

3.8. Fully Connected Layer (FC Layer)

Fully-Connected layer learns non-linear combinations of the high-level features from the downsized convoluted image - the output of the convolutional layer. Fully connected layer involves weights, biases, and neurons generating a vector of Z real numbers for each categories in the dataset.

After the input image is converted into a suitable form (convoluted, normalised and downsized as many times as it works best), the image is flattened into a column vector. The flattened output is then inputted into the feed-forward neural network and every iteration of training is backpropagated. Because the datasets is large enough, the model is able to distinguish between dominating and certain low-level features in images and classify them using the Softmax Classification technique.

3.9. Softmax Activation Layer

This is the last layer of CNN that is used at the end of FC layer for multi-class classification of the image categories. It generalizes the logistic function to multiple dimensions. It is mostly used as the last activation function of a neural network that normalizes and turns the output of a network (a vector of Z real values) into a probability distribution of predicted output classes (Z real values that summing up to 1) according to equation 1 below:

$$\sigma\,(z)_i \;= \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Where K = The number of classes in the multi-class classifier, z = The input vector to the softmax function, made up of (z0, ... zK), and $\sum_{j=1}^{K} e^{z_j}$ = normalization term (at the bottom). It ensures that all the output values of the function will sum to 1 and each be in the range (0, 1), thus constituting a valid probability distribution.

3.10. Output Layer

The output layer is comprised of the image label categories. It maps each probability to the corresponding classes.

3.11. Training Options

The stochastic gradient descent method is used to train the deep neural network. Different values of learning rate is used but the value 0.01 performed best. Because the data is large enough, it was trained for only one epoch. The data is fed into the network in batches. For each iteration, a batch containing 163 training images is used such that after 638 iterations of trainings, the entire dataset is used. Each time a batch is selected from the entire datasets to be inputted into the algorithm, the dataset is reshuffled.

4.0 Results

After complete training of the dataset for an epoch, the training, validation and test sets have accuracies of 99.89 %, 99.86% and 99.95% respectively as shown in the charts below (Figure 6).

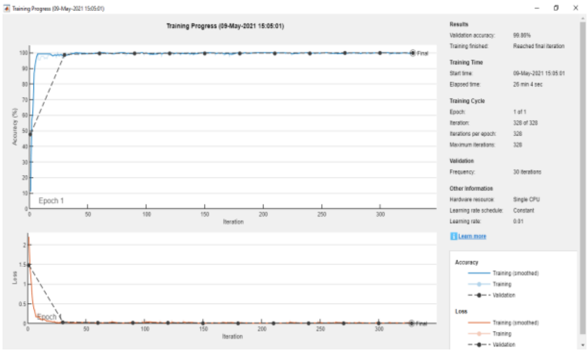

Figure 6: Accuracy curve for the training and validation results.

Besides, the confusion matrix was used to investigate model correct predictions and incorrect predictions (class that the model misclassifies as another class). Confusion matrix also called error matrix describes the performance of a classifier on test set with known true values.

The rows in the matrix represent the correct classifications, while the columns represent the predictions of the model. When the row and the column agree (i.e., along the diagonal), the model predicted correctly.

The model is able to classify the AbdomenCT, BreastMRI, ChestCT and HeadCT correctly when an unseen data is fed in (test set). However, it has mistakenly classified 2 CXR as Hand and only 1 Hand as HeadCT. Investigating the Hand and HeadCT shows that these images images are dissimilar and should be a great area to improve DL models.

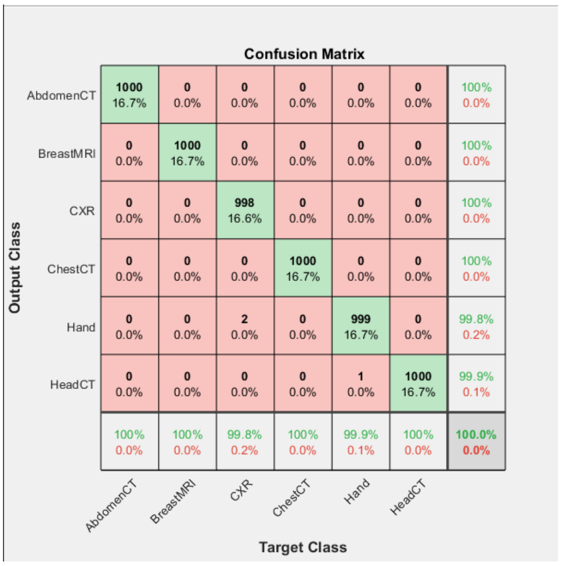

Figure 7: A plot of confusion matrix showing the performance of the classification model.

5.0 Conclusions

The built deep neural network is able to classify any given body-part X-ray images into 6 classes with a 99.95 percent accuracy. The improvement in data quality has increased the network recognition power. With improvement in image quality, Deep CNN will outperform the best human expert.

7.0 References

World Health Organization, "Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children", World Health Organization, WHO Press, Geneva, 2001, pp. S253–S261.

Rawat W, Wang Z., "Deep convolutional neural networks for image classifcation: a comprehensive review neural compu-Tation", Neural Computation, MIT Press, Massachusetts, 2017, pp. 2352-2449.

Eitz, M., Hays, J., & Alexa, M., (2012) "How do humans sketch objects?", ACM, ACM Transactions on Graphics, Association for Computing Machinery, Ne York 2012, Pp. 520-540

Yang, Y., & Hospedales, T. M., "Deep neural networks for sketch recognition", 2015

Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning", Europe PubMed Central, London, 2018, pp. 1122–1131.

Yadav, S.S., Jadhav, S.M., "Deep convolutional neural network based medical image classification for disease diagnosis", Spring Open, Springer Nature, New York, 2019, pp.

Moshe Aboud1, Assaf B. Spanier and Leo. Joskowicz, "Automatic Classification of Body Parts X-ray Images", Learn Modern CFML, Jerusalem, https://docplayer.net/50110329-Automatic-classification-of-body-parts-x-ray-images.html

Y. Bengio "Learning deep architectures for AI Found", Trends® Mach. Learn., 2009, pp. 1-127.

G.E. Hinton, S. Osindero, Y.W. Teh, "A fast learning algorithm for deep belief nets Neural Comput.", 2006, pp. 1527-1554.

Sudholt S. and G. A. Fink, “PHOCNet: A Deep Convolutional Neural Network for Word Spotting in Handwritten Documents” , ICFHR, , 2016, pp. 277–282.

K. Sirinukunwattana and S. Raza, “Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images’, IEEE, IEEE Press, New Jersey, 2016, pp. 1196–1206.

F. Xie, H. Fan, and Y. Li. Melanoma classification on dermoscopy images using a neural network ensemble mode’, IEEE, IEEE Press, New Jersey, Mar 2017, pp. 849–858.

W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang,Y. Zang, and J. Tian, “Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recognition”, Science Direct, Elsevier, Amsterdam, 2017, pp. 663–673.

R. Li, T. Zeng, H. Peng, and S. Ji., “Deep learning segmentation of optical microscopy images improves 3d neuron reconstruction”, IEEE Transactions on Medical Imaging, IEEE, Mar 2017.

Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I.C. Chang, “Deep learning of feature representation with multiple instance learning for medical image analysis”, Microsoft, IEEE International Conference on Acoustics, Speech and Signal Processing, Jul 2014, pp. 1626–1630.

J. Weese and C. Lorenz. Four challenges in medical image analysis from an industrial perspective. Medical Image Analysis, Science Direct, Elsevier, Amsterdam, Oct 2016, pp. 44–49.

Sandra Vieiraa Walter.L.Pinayab Andrea Mechelli, “Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders”, ScienceDirect, Elsevier, Amsterdam, 2017, pp. 58-75.

S. Lawrence and C. L. Giles, "Overfitting and neural networks: conjugate gradient and backpropagation", IEEE, IEEE Press, New Jersey, 2000, pp. 114-119.

https://www.researchgate.net/publication/312205163_Using_deep_learning_to_investigate_the_neuroimaging_correlates_of_psychiatric_and_neurological_disorders_Methods_and_applications/figures?lo=1