University of Reading

Department of Computer Science

Handling Class Imbalance Using Generative Adversarial Network (GAN) and Convolutional Neural Network (CNN)

Vijayakumar Asokan

Supervisor: Dr. Varun Ohja

September 24, 2021

Declaration

I, Vijayakumar Asokan, of the Department of Computer Science, University of Reading, confirm that this is my own work and figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

I give consent to a copy of my report being shared with future students as an exemplar.

I give consent for my work to be made available more widely to members of UoR and public with interest in teaching, learning and research.

<div align="right">

Vijayakumar   Asokan
September 24, 2021

</div>

Abstract

In recent times, the application of Deep Learning has widely been witnessed among medical imaging classification & segmentation, thereby facilitates towards successfully facilitating better diagnostic accuracy. In regards to the image classification, the datasets tends to often appear imbalanced, as its characteristic in negatively impacting the accuracy concerning with DL classifiers. In this thesis, the Balancing GAN (BAGAN) is proposed as an effective augmentation method for restoring balance among imbalanced datasets. For overcoming the issue via including adversarial training of the images comprising the minority and majority classes. This generative model facilitates in learning useful features via majority classes & uses for image generation for minority classes. The GAN generator was configured with an encoder module for an auto-encoder, allowing us to learn accurate class-conditioning. We compared proposed methodology state-of-the-art BAGANs as well as demonstrating BAGAN generated images of superior quality as it trained with the imbalanced datasets. In order for overcoming data constraints, the thesis showcased a more efficacious and novelised scheme comprising techniques that use an ensemble of convolution neural networks (CNN) with BAGAN based augmentation for improving diagnostic accuracy. Also, the implementation of the proposed technique on medical and benchmark datasets over limited size for image classification was presented in detail.

Keywords: Deep Learning, Computer Vision, CNN, GAN, BAGAN.

Report's total word count: 16860

Acknowledgements

Contents

vi

List of Figures

List of Tables

Chapter 1

Introduction

1.1   Overview

With recent and rapid improvements in convolutional neural networks (ConvNets) have significantly improved implementation within the PC Vision business [1], clinical imaging [2], manufacturing [5], and agribusiness [3, 4]. Nearly enhanced processing capabilities, such as access to big data [6], are one of the main explanations for the new achievements. Get a complete image as a major advance being developed for computer vision calculations. If it runs out of purchased photos, it may not be able to complete the ideal mission. Attributes properties, image classification [7], object position [8], and thus segmentation [9] are central structural blocks for advanced imaging tasks. These strategies use a deep web that spans huge layers with few constraints that need to be changed. Therefore, a large number of data agents are required to understand the provisioning and guessing capabilities. On the other hand, visual datasets are significantly extended with many datasets from global datasets and usually persist across different types of dislocations. One of the inevitable problems across digital image processing is the handling of class imbalance and functional anomalies in image datasets.

However, dataset inconsistencies when considering the classification of clinical images can lead to a more pronounced level in most histological presentations. Therefore, the main concern with nervous tissue was to reduce bad behaviour through a sharp decline in trends. After preparing within an imbalanced dataset, trends generally fall into the trap of expecting a larger portion. Aside from reducing larger class samples, the strong main order is assumed to be minority sample expansion. It is difficult to collect OCD cases from clinical images. The best placement is to create a new image of a sophisticated and diverse minority class. In this study, GAN focused on specific tasks by dispatching custom-made generation strategies, propagating real data to prepare datasets and legacy data (new data). Proven to be suitable for creating and mimicking the model was created in. We confirmed this result using a series of non-standard measurable significance tests in class imbalance. To deal with it, GAN is used as the most well-known generation model. In the model as one of the deepest neural tissues, GAN also requires a huge data set to prepare. In any case, a small number of subgroups are continually inadequate to prepare the appropriate GAN. Specifically, GAN changes [3] provide a way to deal with GAN training on imbalanced datasets and are explicit. To create minority class images importantly.

Given the complexity and different nature of datasets accessible today, our findings initiate later modified manufacturing datasets without attempting comprehensive device mapping of engineering data lifetimes. It has been suggested that this is the case. However, our findings in this work show that even in situations where it is important to create synthetic data, it needs to be continually blended with reality to make informed choices.

Overall, the GAN-designed data era is the focus of testing, and we are confident that further work and design improvements in the project could lead to a prestigious data era for various applications. The next step is to further work on the age model activity of the initiated data, investigate different variations of the GAN, investigate the GAN of the discrete data, and further quantitatively investigate the nature of the

initiated data. Finally, we adopt a "Manufactured Data Age" system to provide high-quality datasets for certified applications such as registration and clinical data [6].

Image categorizing is the assignment of classifying informational images as indicated by a set of possible categories. Classification calculations understand how to retain important individual data about an article in an image, such as shape and shading, and ignore non-essential parts of the image, such as planes and objects. Image information for image classification models known as LeNet [7], AlexNet [10], VGG-16 [11], GoogLeNet [12], ResNet [13], Inception-V3 [14], DenseNet [15] Capture and move to the Information via several layers of convolution and grouping. The convolution layer helps remove highlights from the info image, and the pooling layer reduces scaling. This may be followed by many convolution layers and incremental assembly, depending on the geometry and perspective.

The outcome is a series of element maps minimized in size from the first reaction, and the first image shows how to extract data about the material of the first image. All non-positional assignment maps are then changed into a single vector that the fully associated neural tissue progression can process to obtain the probability of mapping the degree of isolation. Informational images of the expected class can be output from this randomization.

While these models are intended to work well with regularly modified datasets, the typical problem with real global datasets is the observed category imbalance. The most common imbalance problem in image class assignments is class imbalance. The imbalances in the categories of photo realistic datasets are widespread and can adversely affect the display of ConvNet [16]. In principle, these datasets fall into four classify in terms of imbalance and size [17].

1. An ideal dataset contains a sufficient, equivalent, or substantially equal number of trials within each category. Similar probabilities are distributed to all groups as they update organizational boundaries and prepare to approach the core values of the wrong business. A wide range of standard AI calculations can be claimed to be ideal datasets.
2. A dataset containing a sufficient number of samples. Some categories are more eccentric than the various cases of categories that appear to be disproportionate datasets. Although these datasets contain a sufficient number of samples, they are expensive, and experts are looking for a large number of datasets that are not named for clarity. It may not apply to reviews.
3. Small datasets can be difficult to collect and are not effectively accessible. These datasets contain the same number of trials within each category but are difficult to collect due to various reasons and security limitations.
4. Rare datasets have a severe category imbalances and minimum number of samples. The description of categorical anomalies in these datasets is subject to change but typically causes problems for the following reasons: (A) Access a given number of specialists for an assortment of data. For the model, the lifetime of the clinical imaging dataset requires special equipment and all clinical professionals to obtain the data. (B) The enormous amount of manual work is required to label a dataset. (C) The lack of clear and candid samples leads to class imbalances. Therefore, the problem of dataset size and class imbalance becomes a bottleneck, making the true power of ConvNet unusable. Figure 1 shows different varieties of datasets in terms of imbalance and size.
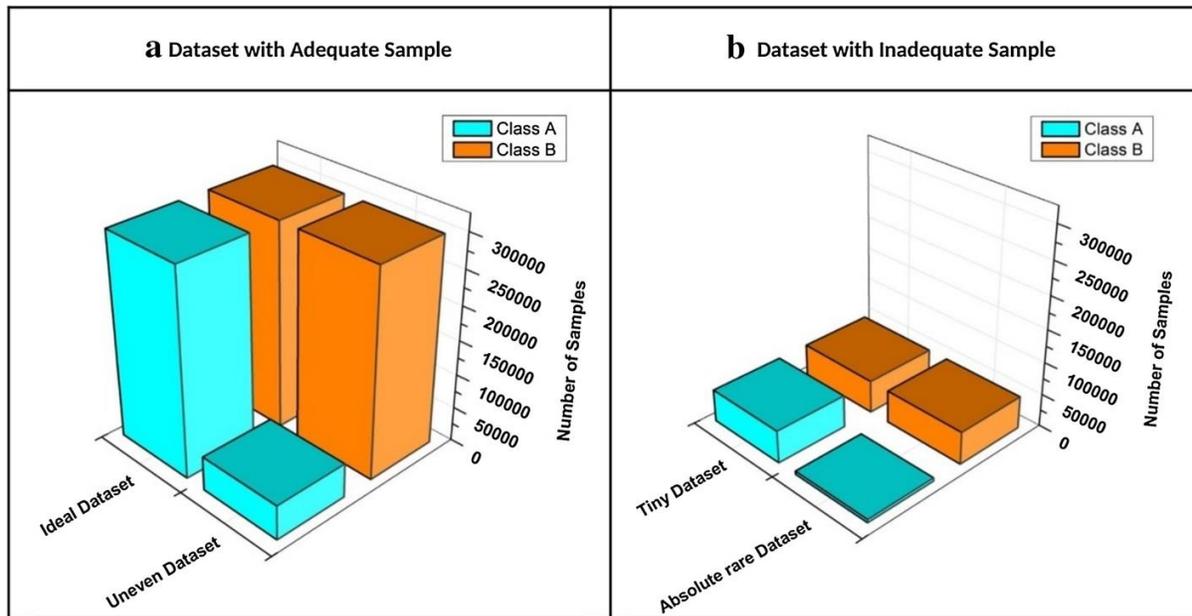
Figure 1.1. Different types of dataset assignments (A) Examples of incomplete datasets (B) Examples of pathological datasets [18]

Class imbalances within a dataset can occur between classes (cage class imbalances) or within classes (inner class imbalances). When a minority class is unable to compare a collective class to an existing event, an atom class imbalance occurs. Classifiers created with disproportionate datasets of dangerous categories may predict minority classes as anomalous events but are accepted as anomalies or confusion leading to misclassification of minority classes [18]. Minority groups are often more important and need to be approached with caution. For example, the rare clinical findings of disease urgently needed to identify a particularly rare disease in the normal population. All kinds of error detection are stressful and inconvenient for the patient. Therefore, deep learning models [19] built using these datasets must have the option of increasing the detection rate of minority groups.

Intra class Imbalances within dataset categories can also reduce the display of classifiers. Imbalances within a class can be seen as clear trends, such as fine visual classification imbalances. For example, dog sample categories can be further organized by dog status, current breed, and dog breed. Such class imbalances (inner class imbalances) are an unavoidable problem in datasets for many classification tasks. For example, methodology-based clinical imaging classification [19], fine-grained trait classification [20], personal identity reconstruction [21], age [22], and currently stable face recognition [23].

Improved GAN balancing: In contrast to the usual methodologies above, to create a composite image, the Generative Adversarial Network (GAN) captures the basic physical distribution of data from accessed restricted images (both minority and majority classes). It means to create using the obtained distribution. This raises the intriguing question of whether GAN can be used to generate composite images of minority groups from disparate imbalanced datasets. Undoubtedly, the continued development of GAN recommends the ability to process complex, high-dimensional data as an oversampling technique. GAN acknowledges the ability to harness the power of nervous tissue to bring the model's distribution as close to the actual distribution as possible. In particular, you can initiate composite images with high optical stability without relying on previous assumptions about data distribution. This important property makes GAN applicable to all kinds of imbalance problems in computer vision work.

GAN can reproduce the phantom image and allows you to change something in the first image. They can generally understand how to create integer categories (e.g., objects, characters, people, etc.) and many categories (views, lighting conditions, scales, exponents, etc.). Slightly there are many written declared GANs, each with its quality that alleviates computer vision task imbalances. For example, AttGAN [24], IcGAN [25], and ResAttr-GAN [26] are specific types of GAN commonly used in face-changing tasks. They have found ways to fine-tune the attributes of independent gels and various facial images with the desired properties. Recently, GAN has been combined with various existing element detection and image segmentation calculations to overcome imbalances and gallery work problems.

1.2 Problem statement

While considering the digital image processing field in recent times, the concerns regarding class imbalance remains to be one of the long-standing problems which is relevant across a wider range of real-world deep learning applications. For instance, oversampling techniques that happen to be effective to handle class imbalance under a classical learning systems, appears to be not directly applicable in case of an end-to-end DL systems. GAN could serve as an effective approach for managing and improving the accuracy of imbalanced datasets. However, there lies a problem concerning with generative adversarial models such that whilst learning to fool discriminator, they might tend to end up extracting or synthesizing one/ few irrelevant / foolish examples. This problem could be regarded to as mode collapse [8, 4, 5, 6]. For the present work, the aim intended to augment imbalanced dataset image classification for balance restoration. It serves to be one of the paramount importance such that, augmented dataset serves as a variable enough & does not include continuous repetition of example/ generated sample datasets, thereby avoiding mode collapse in this regard. In order for carrying out this, various approaches have been proposed. For the present investigation, intended to address this problem with a possible solutions using BAGAN.

1.3 Objectives and Goals

The objectives and goals of this thesis are:

1. To develop a novel technique to improve the overall accuracy of imbalanced medical datasets using BAGAN to produce synthetic images of minority classes for each dataset

2. To develop an efficient state-of-art CNN classifier algorithms for classification of medical images rendered from 9 different datasets.

3. To compare and evaluate the variations across 8 CNN classification models and classes for validating their overall accuracy of the rendered synthetic images.

1.4 Organization of Thesis

This thesis consists of seven chapters. It is organized as follows:

- Chapter 1 introduces medical imaging and the role of AI for improving medical diagnostics, problem statement, objective and goal and contribution of thesis in the field of medical diagnostics.

- Chapter 2 provides literature review on computer aided approach and other different existing works and techniques available that has been implemented for similar problems.

- Chapter 3 discuss about first proposed method; traditional augmentation method and its implementation in training CNN classifier & for Generative Adversarial Networks (GAN) and its implementation in this thesis.

- Chapter 4 presents results and discussions. It explains about evaluation metrics, results and performance comparison between all proposed methods.

- Chapter 5 is the conclusion where summary of thesis is provided along with problem encountered and suggestions for future work.

Chapter 2

Literature Review

2.1   Introduction

Deep learning is an intuitive process whose complexity of learning increaseswith the increase in the number of layers. Due to its high performance, it is regarded as a mature application for medical diagnostics [27]. However, limited data set creates tougher environment for the potential groundbreaking research in medical diagnostics with deep learning. One reason is dependency of the deep learning algorithm on training data size as it requires millions of parameters and large amount of labeled data to learn [28]. When limited data is used to train a deep learning model, it uses a large amount of its resources to train the model, creating over fitting issues. Over fitting issue refer to model's incapability to generalize on unseen data. A large number of research has been done to overcome challenges imposed by limited data on the training of deep learning models. It includes techniques like augmentation [29], transfer learning [30] and ensemble of classifiers [31].

Literature sources pertaining to Class Imbalance in several efforts have been made to get over the consequences of class imbalances using various methodologies. These methodologies can be categorized into data-level methodologies, computational-level strategies, and hybrid methods. Data-level methodologies change the distribution of sets of settings to re-establish harmony by removing or adding states from selected datasets, while computational-level strategies build the importance of minority groups. Change the target capacity of the workbook. Hybrid techniques integrate account-level strategies with data-level methodologies. The next two sections highlight to the reader some of the more accessible traditional ways to deal with the problem of class imbalances.

Two types of retesting can be applied to neutralize the problem of class imbalance remodeling. One is under sampling by scanning tests from the bulk class, and the other is by copying the tests from the minority class [24]. Re-scan adjusts the dataset but ignores the provision of additional data to the configuration set. The various limitations of this technique include: Oversampling causes fit problems during the screening procedure and results in significant data loss [31]. Under screening and oversampling were not well quantified using test strategies and were fully normalized [32]. The synthetic sampling strategy provides a new example rather than transcription for additional data in the preparation group to add consistency to the country's distribution. The Synthetic Minority Oversampling Technique (SMOTE) [33] is a well-known strategy for oversampling, with plans to initiate synthetic examples based on random selection near K-nearest neighbors. Destroyed does not evaluate the distribution of data between categories. The ADASYN approach [34] adaptively creates tests of synthetic data using even distributions of different minority groups, as demonstrated by the learning problem. The cluster-based over-clustering procedure [35] divides the information space into different clusters, after which the valves are checked (such as scaling). Many traditional synthetic extension techniques, such as SMOTE and ADASYN, are only suitable for low-dimensional basic data, limiting their application to high-dimensional image data [35]. In addition, all of the above technologies initiate data by erasing or averaging existing data. This can ignore further development of classification applications.

Augmentative oversampling: One of the most common methods for increasing the size of prepared datasets is data augmentation [36]. Increments such as interpretation, cropping, cushioning, rotation, and even permutations can cause slight changes in the age data of the image, but load adjustment does not work for classifier rendering. Until the end of the preparatory course, no standard technology allows to choose whether a particular augmentation therapy can further develop the outcome. Setting up ConvNets is a tedious interaction [31], so you can try a limited number of Penh boost steps before normal leveling. Similarly, the variety gained from slight adjustments to the image is fairly small. In addition to class changes due to oversampling, enhanced technology has also been introduced as a form of regulation in the design of deep nerve regulation to reduce future redundant snapshots. There is no consensus on the best way to standardize individual zoom procedures together. Additional extensions have been developed in line with these policies.

Semi-Structured Learning (SSL) [37] is one of the most attractive approaches to further developing classification implementations. It approaches a reasonable number of tests marked with x and a large unclassified example (dataset is non-uniform). The SSL protocol uses a combination of individual learning and structured learning techniques. Set up the models in sequence using several different examples, such as configuration sets, and use the prepared model to predict the rest of the dataset. The most common way to name each instance of unlabeled data with a unique expected return value is to use a prepared form. This is called pseudo-code. After labeling unlabeled data with pseudo-label interactions, set up a classification model for both real and pseudo-tagged data. Pseudo-labeled are an interesting point of view for describing various unlabeled data, and naming them can require a long period of human effort. However, SSL is based on the simple basic information distribution assumption of $p(x)$ and can have very few similar distributions in both named and unnamed examples. This marginal $p(x)$ distribution should contain data for posterior $p(y|x)$ distributions. A comprehensive summary of in-orbit learning is given point by point in [38].

Most cost-sensitive learning classification calculations expect the cost of misclassification in both classes and most minority classes to be very similar. Cost-sensitive learning [39] focuses on the cost of misclassifying minority groups through the cost grid.

The most obvious and widely used methodology in ConvNets is data-driven procedures. Deep networks with huge layers have so many limits to set that they tend to be customized when mapping to slightly estimated datasets. Data-level methodologies amplify the amount of setup data entered as regulation and reduces the possibility of future over-engineering of deep nervous tissue. Traditional data plane techniques face related drawbacks, especially when used in imbalances in high-dimensional image data classes.

    a.   Synthetic opportunities using regular data-level methodologies may not be a factor in your setup.

    b.   Either replication or direct interruption achieves the life of synthetic data. This does not create a new anomaly model or confuse the classifier's selection constraints, ignoring work in general implementations.

    c.   For medical images, magnifying techniques are limited to making small changes to the image to maintain exact parameters. Also, the type of reinforcement depends on the problem. Significant improvements, such as mathematical changes, arbitrary excision, and image mixing, can detract from the clinical picture.

d. Applying data extensions to a completely unfamiliar dataset may not give you the classes you need to create a particular example to add harmony to your gradient distribution.

e. Classroom imbalance management is tested in a strict visual order with little intra-class and inter-class variation.

f. Most methods are well-defined for double classification problems. The problem of multi-category imbalances is often considered more difficult than the double correlation for some reason. For example, some minority classes may be mixed. In most cases, for example, it includes 1. Few minorities-many big classes, 2. Many minorities-some big classes, 3. Minority Majority-Many classes are bulk.

Class imbalances in large image classification tasks were investigated and considered. Regardless of class perturbations, various types of perturbations can impair the performance of other PC vision tasks such as object selection and image segmentation [40]. Object detection, which manages constraints and classifies different articles in a particular image, is an additional test and a major task in computer vision. The most common way to constrain an element in an image is to draw a bounding box around it. This bounding box can be thought of as a set of coordinates defining a square. Currently, object detection algorithms are classified into two types: two-stage detectors and one-stage detectors.

On the other hand, two-stage detectors such as area-based convolutional neural network (R-CNN), high-speed R-CNN [40], high-speed R-CNN [41], and mask R-CNN [42]. Such use the Region Proposal Network (RPN) to examine the objects in the main stage and measure their locations of interest to characterize the regression in the regression box for the stages after the object. Again, single-stage detectors such as single-shot detection (SSD) [43], You Only Look Once (YOLO) [44] refrain from investing excessive power to initiate region width increase. It is also detected in the matrix focus on speed and perception rather than finding things exactly. As a result, one-stage object detectors are quick and simple, whereas two-stage detectors are more accurate.

Apply object detection calculations to existing real-world datasets, regardless of new developments such as in-vehicle video [45], sending surveillance images containing objects with large changes (object-scale imbalance)   [46] remains challenging. The actual size of the equivalent object will look different on cameras of different sizes. Singh et al. [47] have shown that the diversity of body-wide scales abnormally affects the overall exposure of the body to detectors. Several arrangements have been proposed to address the target scale imbalance. The R-CNN Rapid Scale [48] uses a combination of two object detectors. One distinguishes between large and medium-sized objects, the other distinguishes objects of limited size and combines them to create the final prediction. Multi-scale image pyramids such as SNIP [49] and SNIPER [50] use image pyramids to multiple aggregate metrics, including imaging. Hierarchical Component Network (FPN) [51] participates in prominent progressive systems at different levels and displays objects of different scales.

The objects in the actual dataset are only a small part of the image, but the rest is basic. Both one-step and two-step calculations evaluate approximately 104-105 areas per image [52], but only two areas have objects. Confusion between the foreground (object) and the base can also prevent the execution of object detection calculations. In addition, object detection calculations must be consistent, even in the event of object distortion or obstacles. For example, in the pedestrian detection dataset [53], more than 70% of people are blocked from walking around video overlays, and about 19% of pedestrians

are at all edges where obstacles are placed 50% of these cases. [54] The advantage, such as the dollar, is that the pedestrian detection range using standard detectors is significantly reduced under imperfect obstacles and certainly under dangerous obstacles. Information augmentation based on arbitrary deletion [55] is a commonly used strategy to allow detectors to focus on an entire object rather than part of it in the image but this strategy is not very valuable in all situations. Skewed distributions are unprecedented because it rarely occurs in real-world situations, and it also appears as part of obstacles and distortions within distorted or disturbed objects [56].

Image segmentation, which collects all the pixels in an image, causes non-uniform pixel levels, similar to other computer vision tasks Notable image hash calculations include Fully Associated Network, SegNet [57], U-Net [58], and ResNet [59]. Image segmentation is the basis for a variety of tasks, including [60], recent reviews [61], and malignant cell segmentation [62]. This task-loaded dataset experiences the negative effects of pixel-level imbalances. For example, in the city road scene dataset [63], the pixels compared to the sky, buildings, and streets are significantly different from the pixels of pedestrians and bicycles. This is because it is covered by the sky, structures and streets, not the pedestrians or bicycles in the image. Basically, in the mental development image segmentation dataset [64], MRI images contain healthier brain tissue pixels than prominent tissue pixels. Most of the time lost in image segmentation tasks is the loss of pixel perforation entropy [65]. This loss assigns an equal payload to each pixel, evaluates each pixel's predictions individually, and then evaluates the midpoint of all pixels. Some work has been done to fine-tune the effects of artificial space loss on a pixel-by-pixel basis to alleviate this problem. Standard entropy losses are weighted entropy [66], focal loss [53], template loss [67], Generalised Dice loss [68], Tversky loss [69], Lovász-Softmax [70], Median frequency balancing [57], to assign higher importance to rare pixels. Variable loss features are beneficial to some asymmetric practitioners, but these features are subject to serious problems with highly imbalanced datasets, as evidenced by clinical image segmentation.

2.2   General Computer Aided Approach

A general computer aided diagnosis system performs analysis using computer- based algorithms (Machine learning, Deep learning etc.) to detect medical conditions. These systems help human experts make accurate diagnostic decisions. In recent times, CAD systems have developed by integrating computer vision, machine learning, artificial intelligence, statistics and mathematics. A general overview of a CAD system for training deep learning model on limited dataset for classification is shown in figure 2-1. The general CAD system for classification on limited training.
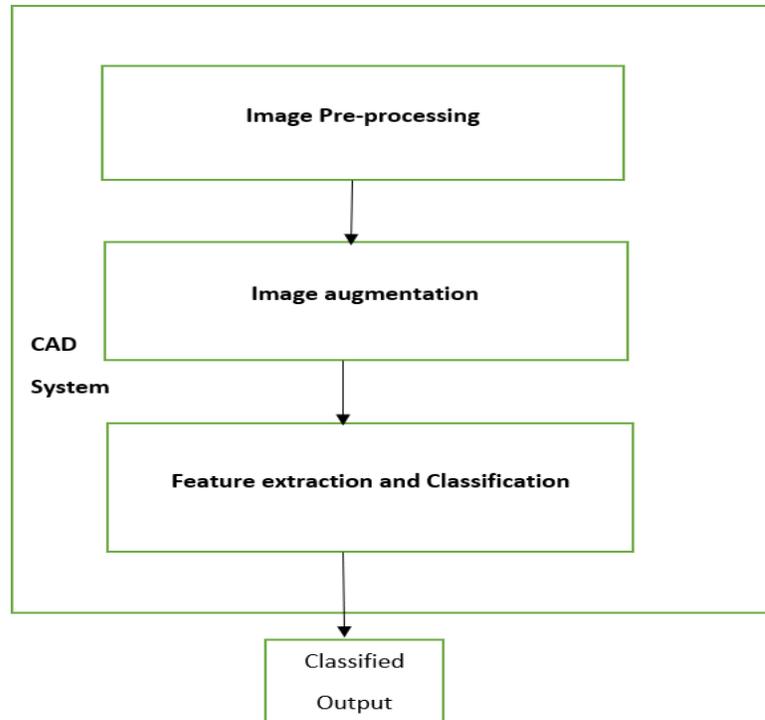
Figure 2.1. System Overview data has the following modules: (i) Image pre-processing, (ii) Image augmentation, (iii) Classification or (iv) Ensemble of Classification.

These modules along with the other existing methods on these modules are discussed in the subsections below:

2.2.1 Image Pre-processing

Before feeding input images into the neural network, images need to be pre- processed. Pre-processing is the first and basic step in image analysis. It is done to improve the quality of raw input images. Image is represented in the form of 2-D or 3-D array of pixels. In case of RGB images, it is a 3-D array of numbers (pixels). Commonly used pre-processing techniques include resizing of the image, de- noising, edge-smoothing, normalization and ROI (region of interest)/patch extraction. Image resizing refers to the process of converting all images to a specific size. It is essential to any system due to varying numbers of medium to capture images. Denoising is required to remove unwanted noise in images and is usually done using smoothing filters like Gaussian smoothing etc. [71]. Normalization centers the data and is done by subtracting the data from the mean in the first step and is then divided by standard deviation. This helps in normalizing the features. During the training in neural network, weights are multiplied, and biases are added to initial inputs and are passed through activation function. Further, back propagation occurs with gradients to train and optimize the model. Therefore, to have each feature in similar range while back propagating, normalization is needed. Normalization is done to input images at the initial stage. Pre-processing steps can also be executed by creating pipelines which can be done by loading each individual image, performing pre-processing and feeding through the network or by pre-processing the mini batches of images.

Four commonly used pre-processing techniques in deep learning are discussed in the sub- sections below:

2.2.2   Aspect Aware Pre-processing

When images are resized/rescaled into fixed size, the aspect ratio of the image is ignored. In image, aspect ratio is defined as the ratio of the width of an image to the height of an image. Ignoring aspect ratio creates compressed and distorted images. In medical imaging, it is safe to preserve the aspect ratio of an images while resizing. Aspect ratio pre-processing takes in aspect ratio of an image and resizes as per its height and width. For example, there are two steps involved in resizing an image to 64x64. In first step, image is resized along shorter dimension so that width is 64 pixels followed by second step where image is cropped along height in a way that height is 64 pixels.

2.2.3   Mean Pre-processing

Mean pre-processing is done by subtracting each input images with mean value of intensities and is done for data normalization. In cases of RGB images, such subtraction occurs for three different intensities; Red, Green and Blue. Initially, average of red, green and blue pixels is computed. Suppose, I is an input image with RGB channels, then mean subtraction is done by:

$R = R - \mu_r, G = G - \mu_g, B = B - \mu_b.$ [71] (1)

2.2.4   Patch Pre-processing

Patch pre-processing is applied when spatial dimensions of input images exceed the spatial dimension expected by CNN. Patch pre-processing is used to remove over-fitting by randomly sampling MxN regions of an image during training. By using patch pre-processing-instead of passing the entire image through the network, it passes the random cropped portion.

2.2.5   Crop Pre-processing

Crop pre-processing is used during the testing phase and is responsible for producing cropped images for oversampling. It is done by cropping four corners and the center region of test images. Further, horizontal flip of these five crops results in total of ten samples from a single test image.

2.3   Image Augmentation- GAN approach

Image augmentation is done with a purpose to expand limited data set or to pre- vent class imbalance in data set. It is the most popular strategy to compensate for insufficient training samples. It can be performed in numerous ways ranging from the traditional ways of using parameters like zoom, rotation, shear and other prepossessing functions to modern methods of using neural networks as image generators. Because of their simplicity and ability to increase data set size from 3x to 4x, traditional augmentation methods are widely used. Traditional augmentation includes parameters like rotation, flipping, shearing, translating, resizing/re-scaling and applying filters. The use of these parameters led data set to have these variations or attributes during training and thus, increasing the diversity to some extent. It is essential to understand about the parameters that can best capture medical image statistics as such parameters have great effect on classification [72]. According to research, traditional augmentation strategies differ in terms of image information retention as well as performance. [73].

However, in the limited dataset training cases, it always performs better when compared to the classification done without any augmentation and carries the potential to out-perform neural network based image augmentation methods like GAN [74, 75]. However, traditional augmentation is limited to certain limit. Beyond the limit, traditional augmentation reduces model's ability to generalize by reducing diversity in images by producing same images repeatedly [76].

Another approach to augmenting images is by using generative models, a type of unsupervised deep learning. Unsupervised deep learning understands the underlying structure of the data and is used to initiate similar types of data. One example about its application is the automatic poetry generator that studies underlying structure of poems from millions of poems to initiate new poem [77]. In the same way, it is being studied the domain of music, text and image to initiate the new data. In generative models, instead of directly modeling p (y/x), we model p (x/y) for each class 'y' and use this model to initiate samples. One of the simpler approach to initiating data is Bayes Classifier. It fits data distribution directly into the Gaussian distribution and gets the mean and the co-variance for sampling. But it is limited to the cases of single mode and fails to initiate data in cases of multi-modal distribution. To address multi modal distribution, generative models introduced latent variable (say 'z') to initiate samples and are capable of learning data distribution implicitly or explicitly. Gaussian Mixture Model (GMM) is one of the types of generative models that introduced latent variable 'z' to define the cluster 'x' that it belongs to, given label 'y' and its distribution p(y). Sampling is done by training the model using Expected-Maximization (EM). EM is an iterative process to find the maximum likelihood estimation of parameters. Despite being better solution to the Bayes classifier, samples initiated by GMM suffer from quality issues. Another issue with these models is that they explicitly model the probability distribution of each class, and their framework requires the output layer of the generator network to be in functional form, such as a Gaussian distribution. Generative adversarial network (GAN) is one of the promising generative models that has been proposed as a solution to the problems faced by other existing generative models [78], and GANs are able to produce exceptionally realistic images [79]. Figure 2.2 lists out different generative models.

The advantage of GANs over GMM is that they do not deal with explicit distributions, instead learn data distribution in implicit way. Main goal while working with GANs is to reach Nash's equilibrium of a game leading towards the generation of the exceptional sample quality. When compared to Variational Auto Encoder (VAE). GAN has no variational lower bound. VAE has some bias while GAN is asymptotically consistent. Restricted Boltzmann Machines can also generate samples, but they require Monte Carlo sampling, which takes 1000s of iterations. The Disadvantage of this method is that there is no way to understand how many iterations are enough. The same issue affects the Generative Stochastic Network (GSN). Compared to Boltzmann Machines and GSN, GAN has no variational lower bound and no intractable partition function and samples are initiated in one pass. There is no restriction on the size of latent variable in GAN as on generative models like Noise-Contrastive Estimation (NCE). GAN is better than Pixel RNN/CNN in terms of run time.
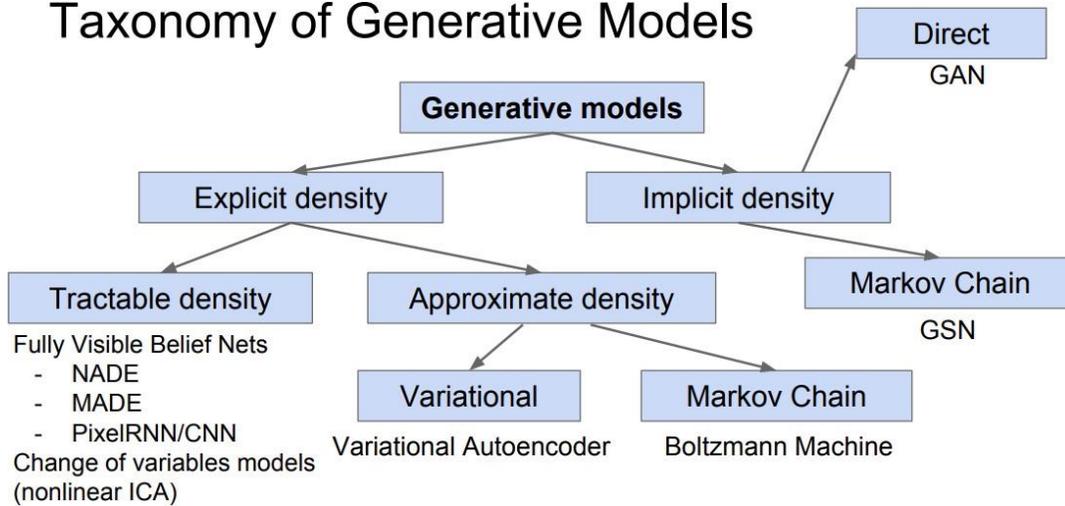
Taxonomy of Generative Models

Figure 2.2. Generative models [80]

Loss function is the indicator for model's performance. In the case of supervised learning, it is explicit and objective. We have metrics like R-score, accuracy etc. However, in generative models, assessing quality of sample is subjective and has no numerical measure. The Quality of sample is determined by how good it appears.

2.4   Image Classification

Classification problem in machine learning is a supervised learning that takes in a set of 'n' samples of data to predict properties of unseen data. It tries to learn from given observations to classify a new observation. Input dataset may have samples divided into two classes (binary classification) or may be into multiple classes. Commonly used classifiers are support vector machines, decision trees, random forest, linear classifiers and neural networks. When an input is in the form of images, it is referred to as an image classification problem. The traditional way to deal with image classification problems was to focus on feature extraction algorithms like BoW (Bag of Words) , SURF (Speeded up robust features) etc., to reduce an image to set of features and followed by the additional processing before finally, feeding to the classifiers[81] . ImageNet challenge in 2012 demonstrated ability of deep convolutional neural network to find out complicated structure in high dimensional data and thus, contributed to application of deep learning for image classification [82]. CNNs are neural networks arranged in hierarchical layers where lower layers extract basic information and higher layers extract complex information about an image. These extractions take place by using convolution, filters and by adding weights and biases to the features on the basis of their importance. CNN and its components are discussed in detail in the next chapter. In this chapter, we discuss about numerous alternatives that have been proposed to train CNN classifiers on limited dataset. Two popular approaches are either using transfer learning or building CNN classifier from scratch [83]. The latter approach is not applicable efficiently without augmentation. However, transfer learning is possible without data augmentation. Transfer learning is a deep learning method in which a model developed for one task is reused as the foundation for a model on a different task. Most commonly used approaches are; develop model approach and pre-trained model approach [84] [85].

Develop model approach focuses on the first step of developing a successful source model on a similar/related problem where the data is available in a large amount and feature learning is possible. In the second step, source model can be used asstarting point for model in the second task. Choice can be made on whether to reuse all parts or some parts of the source model. In the third step, model may require fine-tuning of parameters to adapt to the second task. Pre-trained model approach is similar to developed model approach. Here, one can select a source model from the available public models that are released by different researchers or institute. These publicly available models are usually trained on huge amounts of data with the best possible hardware. These pre-trained models can be used as a starting point for model in the second task of interest. Again, it follows the same steps as in develop model approach in using parts of pre-train model and in fine-tuning. As per review paper [86], most researchers use AlexNet, VGG (Visual Geometry Group)Net, GoogleNet or ResNet (Residual Neural Network) that are pre-trained on larger database such as ImageNet for transfer learning by modifying last fully connected layer of pre-trained model. It allows millions of parameters from pre-trained networks to be used for training on a smaller dataset. Transfer learning is improving the performance of CNN classifier for limited data sets [87] [88] without need for data augmentation or even with data augmentation[89].

2.5   Ensemble of CNN Classifiers

Another interesting alternative to the address data scarcity issue in deep learning is using the ensemble method in CNN. Ensemble method in deep learning is similar to traditional machine learning ensemble algorithms like random forest, decision tree where decision is finalized on the basis of voting. Improvement in performance is observed when the output predictions from different CNN classifiers are combined using averaging or voting. Using an ensemble of different CNN classifiers yields better or equal performance than using a single classifier [90]. There are two common ensemble methods that exist in deep learning. First is an ensemble of CNN classifiers using same training condition and the second is an ensemble of CNN classifiers with varying training conditions.

In the first method, numbers of similar CNN models are trained on the same data with a same training condition (like learning rate) varying the initial conditions. It makes an assumption that each model has different learning ability and if each of their output is averaged out, it produces better or equal result compared to single model and thus, overcomes problem of trapping into local maxima. Stacking is the method of aggregating a large number of CNN models and averaging their output based on Jensen's inequality. In the second method, it uses different networks by varying their capacity and training them under different conditions similar to the transfer learning. This creates lower correlation between models in the final predictions. Ensemble can also be done with different pre-trained networks[91].

Chapter 3

Methodology

3.1   Traditional Augmentation with CNN Classifier

The chapter provides a detailed emphasis on the training scheme on the key basis of effective data augmentation approach for increasing the overall performance of CNN classifiers. This is achieved through generation of training the sample images from that of the original ones without altering the class's labels present. Even though, data augmentation is widely been employed as 'regularizer', as the fundamental concern in this regard is towards increasing the size of the dataset via an effective way for overcoming over-fitting & generalization related issues. In case of the traditional augmentation serves as baseline approach for improving the CNN performance followed by furthermore providing a comparative analysis alongside with other approaches/methods involved. It could be effectively applied on a small amount via applying a simplistic geometric transformation to that of the original samples. These transformation involves majorly of translations, changes in scale, rotations, flipping & shearing in this case. Besides, the study proposes the generic architecture of CNN model for classification of medical datasets for the study. The key focus ensures to enhance generalizability concerning with CNN model via providing slightly modified form of version for the input data in order for the model to actually learn the robust features present.

From the proposed method could be basically divided under two major sections namely: augmentation & classification. For the dataset presented in the experiment, we implied different forms of traditional augmentation approaches to train the data and it appeared successful for ascertaining classification effectively. Then, the enlarged dataset was used to train the proposed CNN model. The experimental results showcased an improved accuracy by CNN classifiers, however could also overcome from over-fitting as well as generalization related issues, whilst comparing with that of a CNN model which is trained without augmentation. The below figure provide a detailed summary for the proposed method.
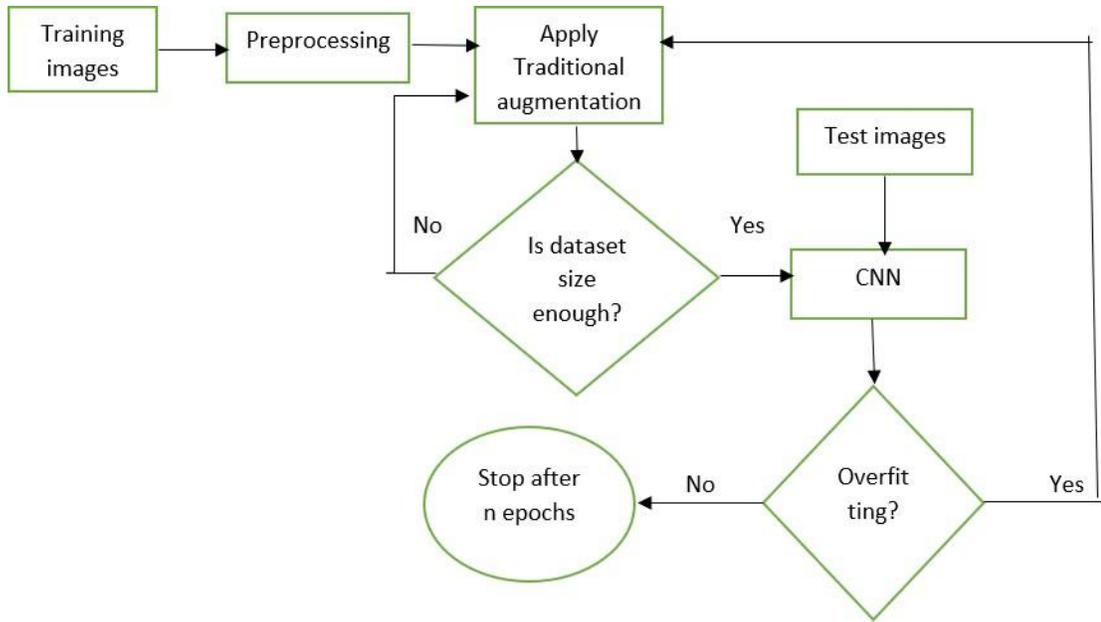
Figure 3.1- Flowchart of proposed method

3.2   CNN & their components- An introduction

As convolutional neural network is regarded to be one of versatile and also a popular DL algorithms especially in the computer vision field. Its state of art DL algorithms were developed via neural networks that are arranged in layers with its first layer extracting only the basic detail/ information pertaining to images namely- coloration, edges, etc. & output from the one layer is later the fed as input for the next subsequent layers concerned. Therefore, by increasing its learning complexity is achieved via increasing their layers. When compared with that of traditional ML algorithms, DL performs the automatic features pertaining to extraction & classification related processes. CNNs are generally comprised of various components. The foremost component involves with an input image which could be attributed as a matrix comprising of pixel values. For the investigation, RGB as well as grayscale images were employed. In case of RGB image could be separated via their color plane involving colors of namely- Red, Green & Blue and are represented under three dimensions: height x breadth x number of channels concerned. CNNs tend to reduce the images into suitable forms such that it appears easier for processing without losing their important as well as its underlying critical features.

When considering the second component of the CNN architecture comprises of convolution. In regards to the convolution, it extracts the high level features via addition of different layers wherein the lower layers appears to be solely responsible to extract features of low level as well as higher level in case of high level features & further then utilizes convolution via updating the weight & biases for assigning importance for the different features in this regard.
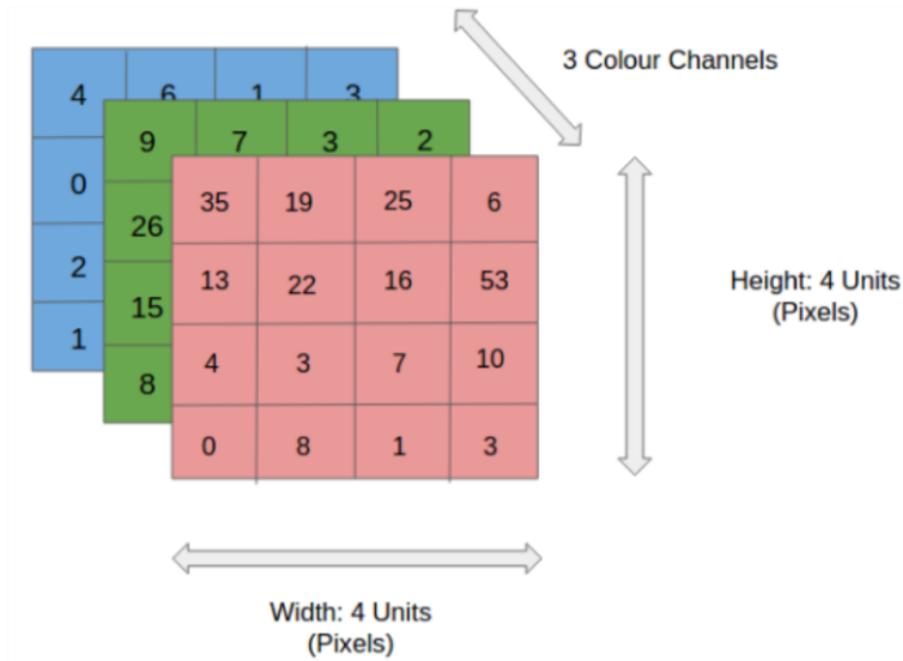
Figure 3.2. RGB Image of 4x4x3 Matrix [59]

In case of the convolutional layer, there exists several filters which is otherwise known to as Kernel which is responsible to perform convolution operation on the first layer of its network. Later on it defines filter size & then performs matrix multiplication via hovering the kernel with a size 'K' over the portion of the image P as indicated in figure 3-3. These stride values tend to be defined such that filter could move over its next direction via covering a defined length. As filter tends to move from left -to- right & then across top-to-the bottom, till it traverses the image entirely. Herein, kernel's depth appears to be same to that of input image. After employing matrix multiplication across the filter 'K' & images pixel 'I', the resultant outcomes are summed up with actual bias. With regards to the third component involves 'pooling'. In this case, pooling seems to be responsible in case of dimensionality reduction as well as for reducing the overall computational power which is required to data processing as well as for extracting rotational as well as position invariant dominant features via utilizing kernel that is similar with that of the convolutional layer. For the present experimental investigation, max pooling was employed mainly two purposes; firstly, it returns to the intensity value with maximum pixel across each time filter, thus covering the image at the P portion as presented in figure 3-4, and secondly, it carries out denoising through suppressing the noise alongside with reduction in dimensionality.
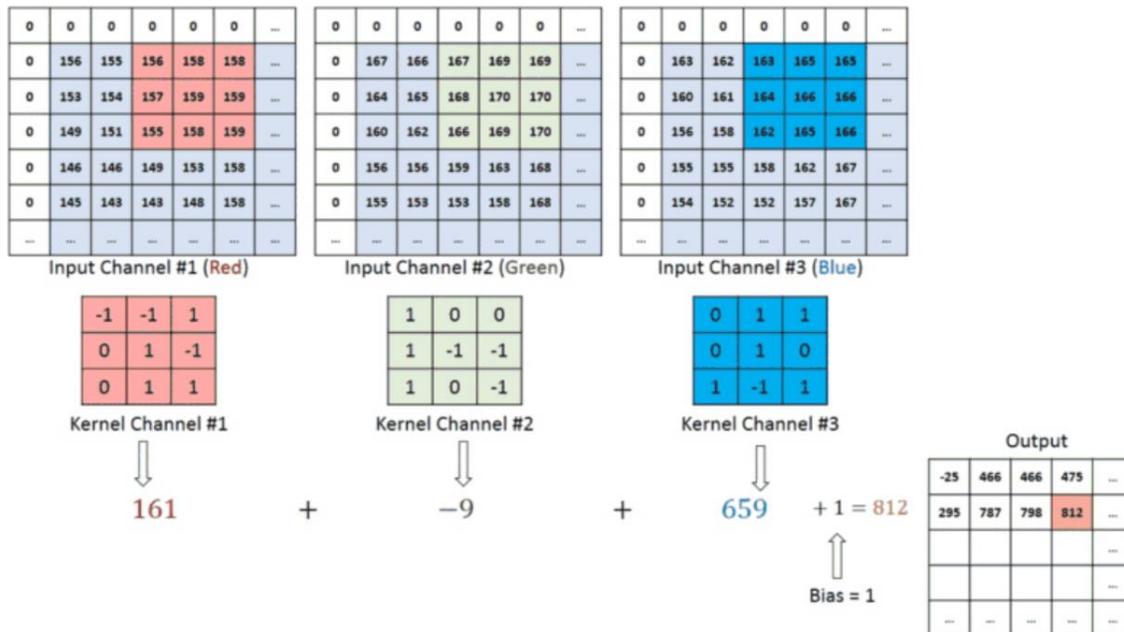
Figure 3.3 – Convolution with a 3x3x3 kernel on an MxNx3 image matrix [59]
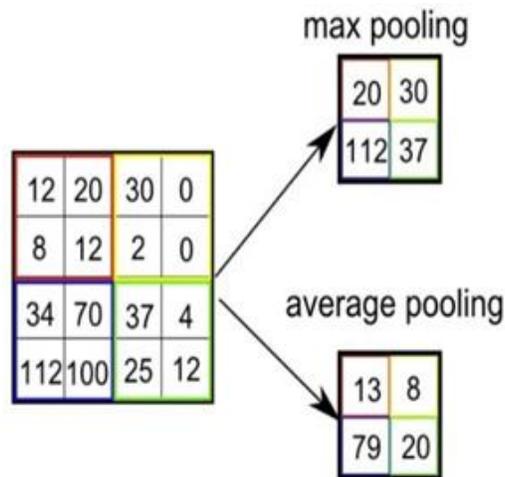
Figure 3.4 - Types of Pooling [59]

In case for formation of ith layer in case of CNN, both the convolutional & pooling (maxpooling) layer were stacked with one another.

Each of the outputs from neuron are subjected for the function known to as activation function. For this study, ReLU activation function were employed which are of non-linear and are defined A(x) = max (0, x). ReLU functional outputs x wherein x is considered positive & 0 for other cases. ReLU activation function is presented in figure 3-5.
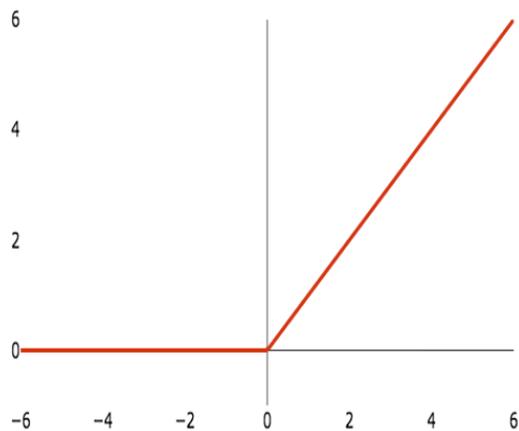
Figure 3.5 - ReLu Activation [59]

With regards to fifth component involves 'batch normalization' as it normalizes its output on previous activation layer. Then it then subtracts batch mean & divides via batch wise std. dev. It increases its stability of NN. Another key component involved in CNN is dropout as it involves regularization technique which are employed to reduce over-fitting layer as it drops out values over unnecessary nodes. It is implemented for any layers excepting that of output layer. As the Dropout value presents probability over which outputs layer that were dropped. Across numerous layers in convolution as its max- pooling, Fully Connected (FC) layers serves as other basic components involving CNN as it learns non-linear combination for higher levels of features presented by outputs in convolutional layers. The layer having connection for almost all activation fn. applied in entirety of the network. After image conversion into its suitable form, as it flattened to vector as they are later fed to feed forward NN & further then back-propagation was applied after its every iteration.

After training for majority epochs, CNN model are further then identified & distinguished the higher level as well as low level features of the images & then it classifies them via Softmax classifier. Using Softmax classifier facilitates cross entropy of loss fn. for classification & outputs probable class on key basis involved in probability distribution. Also it provides discussion concerning loss as well as optimization function in below section.

3.3   The Proposed CNN Architecture

From the proposed CNN architecture for skin lesion system of classification as showcased in table 3.1. As the Input size for the image specified for 64x64x3 pixels. As the architecture having three main blocks to convolution & its second & third block, multiple convolutional layers which are stacked together for deepening the NN and for prevention of over-fitting. As each of the activations were followed involved in batch normalization as well as dropout appeared to be applied after each of the pooling. The initial block involved with the architecture were comprised of convolutional layer as it comprises of 32 filters with size of 3x3. This is later then followed by ReLU activation & Batch Normalization (BN) respectively. Application of Max-Pooling 2D after BN for reducing the spatial dimension. Dropout in case of 0.25 were applied after its pooling tend to minimizes over-fitting. Second block having two convolution layers involving each of it consisting over 64 filters for a size of 3x3 stacked together under each of the layer followed via ReLU activation & BN afterwards. In case of after

19

the final BN for second block, dropout & Maxpooling2D were applied. As the two blocks learning the basic features for image whereas Deeper as well as richer features as it learned in 3rd block. In case of 3rd block that stacks with 3 convolutional layers & ReLU activation also with BN before the finalised pooling as witnessed from table 3.1. After completing 3rd block, standard feed- forward network were built employing fully connected for layers consisting over 256 nodes. After it is fully connected with layer are established-as we run via BN followed via dropout for 0.5. From the final block, SoftMax classifier are introduced. Training performed using the Adaptive gradient optimizer (Adagrad) with a learning rate for 0.001. Since, there lies a two-class classification problem, as the binary cross-entropy was applied for its cost function.

3.3.1   Implementation of Traditional Augmentation with CNN Classifier:

Traditional augmentation enlarged data set size and it has been implemented with CNN classifier to classify on the basis of categories or classes. It is detailed below:

- *Load the labelled training data.*

- *Resize training dataset size to 64X64X3 using aspect aware pre-processors.*

- *Write resized images into HDF5 (Hierarchical Data Format 5) format.*

- *Subtract the mean Red, Green, and Blue pixel intensities from an input to use the mean pre-processor for data normalisation.*

- *Apply traditional augmentation to generate four different images from a single image.*

- *Feed into CNN model.*

- *Apply patch pre-processor during training to extract M N pixel regions from image to avoid over-fitting.*

- *Train CNN model using binary cross entropy as cost function and adaptive gradient descent optimizer.*

- *Apply crop pre-processor to test images.*

- *Predict output of the CNN model on test images*

Table 3.1 - CNN Architecture

Layer	size/value
Input Image	64x64x3
Convolution	(32(3,3))
Activation	ReLU
BN	
Pooling	(2,2)
Dropout	0.25
Convolution	(64,(3,3))
Activation	ReLU
BN	
Convolution	(64,(3,3))
Activation	ReLU
BN	
Pooling	(2,2)
Dropout	0.25
Convolution	(128,(3,3))
Activation	ReLU
BN	
Convolution	(128,(3,3))
Activation	ReLU
BN	
Convolution	(128,(3,3))
Activation	ReLU
BN	
Pooling	(2,2)
Dropout	0.25
FC	256
Activation	ReLU
BN	
Dropout	0.5
FC	2
Softmax	2

3.3.2   List of CNN Models employed

1. *VGG16*
   VGG16 attributed as CNN model proposed by Simonyan & Zisserman for article published in "Very Deep Convolutional Networks for Large-Scale Image Recognition". As its model achieved to about 92.7% top-5 test accuracy for ImageNet, as it involves dataset for about 14 million images belonging to about 1000 classes.

2. *VGG19*
   VGG19 – it serves as a variant for VGG model as it presented in short comprised over 19 layers (16 conv. layers, 3 fully connected layer, 5 of the MaxPool layers & 1 SoftMax layer). Employed just as the good classification architecture for several other datasets as its authors made sure

that models available for the public as could be employed with modification for other similar tasks.

3. InceptionV3
Inception v3 is a widely-used image recognition approach/ model which is showcased to attain over greater than over 78.1% accuracy with ImageNet dataset. The model is a CNN to be of 48 layers deep.

4. InceptionResNetV2
It functions as a CNN that has been trained on over a million images from the ImageNet database. These networks have 164 layers because they classify images into 1000 object categories. In this regard, the network were learned with rich features that presents a wider image range.

5. DenseNet169
It is one of the most recent advances in neural networks for visual object recognition. DenseNet is very similar to ResNet, but there are some key differences. DenseNet is very similar to ResNet, with a few key differences. The DenseNet is divided into DenseBlocks, each of which has a different set of filters but the same dimensions. The Transition Layer employs batch normalisation via down-sampling; it is a necessary step for CNN.

6. DenseNet121
The model is regarded one among DenseNet group of models that are designed for image classification. The authors actually trained models for Torch*, however then converted it to Caffe* format. These DenseNet models were pre-trained via ImageNet imaging database.

7. MobileNetV2
It is a CNN architecture that aims to perform well on mobile devices. It is based on an inverted residual structure, with residual connections between bottleneck layers.

8. ResNet50
This CNN model has 50 layers deep. This makes it easier to load the pre-trained network version trained on over a million images from the ImageNet database [1]. This pre-trained network is capable of classifying images into 1000 different object categories.

3.4   GAN based Augmentation involving CNN based Classifier

As the following chapter presented the training scheme that first utilizes the traditional augmentation for enlarging as dataset & it applies GAN for generation of synthetic images. Initially, high quality synthetic imaging are generated using the Balanced GAN architecture (one variant for GAN) and later, wherein the novel scheme to skin lesion classification are presented via employing CNN. Furthermore, it was demonstrated that in regards to the generated images were employed for augmentation of synthetic data for improving CNN classifier performance. In this first section discussed about GAN as its components involved for second section, as it presented proposed model & the components as it final section presented the implementation steps to train CNN employing synthetic images. The experimental outcomes in the forthcoming chapter showed proposed approach as it not only prevented over-fitting oriented issues, however also improves the accuracy.

3.4.1   Introduction to GAN

Good fellow et.al 2014 proposed novelised GAN framework as the adversarial process for estimation of generative models [38]. GAN involving with two main components; namely- generator followed by discriminator. Generally, both of these involving NNs on the basis of generative & discriminative model respectively. In case of the former learns for generating realistic samples as well as samples generated via generator being regarded to be of negative training samples for discriminator. The discriminator learns towards distinguishing between the generator's fake & real samples & thus, penalizes the generator to produce unrealistic sample involved. During the training process, generator starts via production of unrealistic samples as its discriminator quickly learns enough to distinguish between the classes. Both generator as well as discriminator are trained as it is when generator being well-trained, as discriminator are found to be effectively classifying the fake samples in case of real. This further decreases discriminator's accuracy. There lies two famous analogies for explaining the GAN. When considering the first analogy involves counterfeiter vs that of police. In this case, the generator tends to present counterfeiter as it tries to produce the fake currencies as well as the discriminator further then presented by police personnel as it is against counterfeiter & it tries to detect the fake currencies. As both are found to be competing against each other as the competition facilitates further to improves the performances as it is counterfeiter produce realistic currencies that are indistinguishable from that of the real ones. Another analogy serves as a min-max game. From game theory, in a two-person, zero-sum game, a person is winner only after his competitor loses [91]. Till then both are competing against each other and improving from mistakes until one person loses.

Normally in GAN, the generator (G) is defined by a parameter ϑ and it takes input a random noise (z) and generates sample G (z;ϑ) which can be said to be derived from P_g distribution. The discriminator accepts actual data (x) from a real data distribution P_{data} and distinguishes between real and synthetic images [60]. General overview of GAN is presented in figure. Both of these networks are trained in a consecutive way. When the discriminator is being trained, the generator's weight remains constant and the generator keeps on producing fake samples for the discriminator to be trained on.
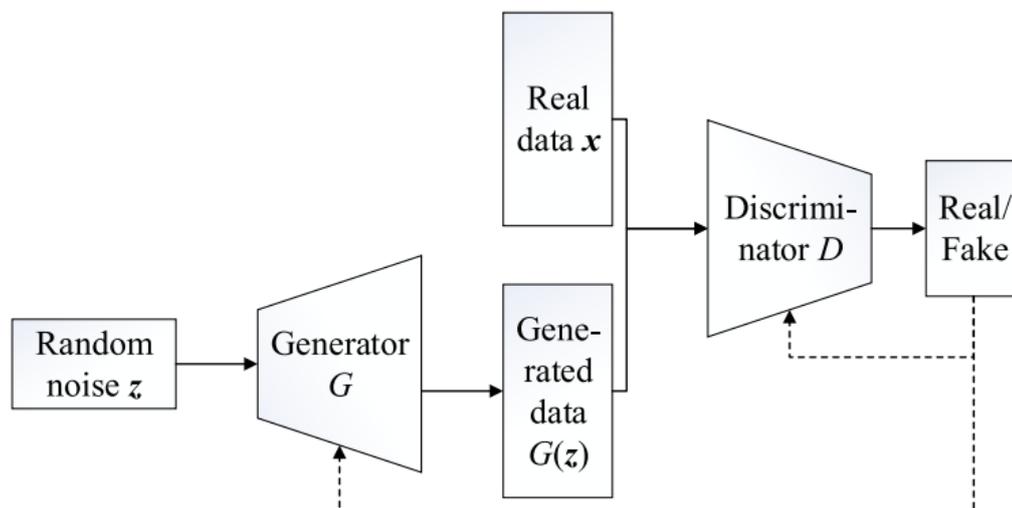
Figure 3.6 – General architecture of GAN

GAN has four major components and they are: generator, discriminator, training and loss function. It is a neural network based on unsupervised learning. It is a generative model and is responsible for generating samples that are similar to the training samples. This neural network estimates the probability distribution of the original samples and updates its parameters to generate samples with

similar distribution as real samples. The generator accepts random noise and produces meaningful an output. The choice for the distribution of noise is optional. It can have a uniform, normal or anyother distribution. In some versions of GAN, an output of the generator can be controlled by adding conditional input c to random noise z. Conditional input can beanything like class of image [92].

In a neural network, the training is done by updating the weights of network on the basis of error or loss of its output. In GAN, generator doesn't have direct connection to the discriminator loss. The generator feeds its output (fake data) into the discriminator network and the discriminator does its training and classifies whether input to the discriminator network is real or fake. Whenever discriminator classifies the generator's output as fake, the generator loss penalizes the generator. Here, the weight of the generator depends on the weight of the discriminator and thus, back propagation is done from the output to the discriminator into the generator as show in fig 3.6.

Here, the generator is trained by feeding in random noise and then thegenerator network produces the output. This output is fed into the discriminator where, the discriminator classifies the generated output into either real or fake. Furthermore, loss is calculated from the output of the discriminator, and gradients are obtained by performing back propagation through both the generator and the discriminator. These gradients are then used to change the weights of the generator in order to produce a sample that is better than the one produced in the previous iteration.

3.5   BAGAN- proposed framework

For the proposed study employed BAGAN methodology that aims for generation of realistic images of minority-class in case for imbalanced dataset. Furthermore it exploits on the all of the available information for specific classification issue via including BAGAN training majority as well as minority classes on a joint note. GAN as well as auto-encoding techniques involved in coupling in order for leveraging strengths for two of its approaches. GANs tend to generate high-quality images, wherein its auto-encoders converge for a better solution easily [76]. There are numerous authors suggesting for coupling the GANs as well as with auto-encoders [74]. Nonetheless such works appeared to be not directly meant for driving GAN generative process for a specific set of classes. Also it is not regarded easy for generalizing them for enabling GAN towards distinguishing between various classes. In our study in particular, the researcher applied class conditioning in accordance to Odena et al. [93] for embedding class knowledge on BAGAN. In this regard, it applied the pragmatic usage of auto-encoders for initializing GAN close for seeking good solution & to stay far off from undergoing mode collapse. In addition, the researcher has applied encoder portion for auto-encoder for inferring distribution across various classes within latent space. In case of auto-encoder-based GAN initialization could be achievable via utilizing same network topology on auto-encoder & GAN modules, For instance the following Figures 3.7(a) (b) and (c). Under the decoding stage pertaining to auto-encoder matches with the generator topology G. Also the encoding stage (E) in which the auto-encoder matches with first layers topology involving discriminator. For BAGAN, the knowledge pertaining to the auto-encoder are transferred to GAN modules via initializing parameter which weights correspondingly. Furthermore Figure 3.7(b). involves with completion of the discriminator, with final dense layer involving with softmax activation function translates latent features to probability with which the images are categorized either as fake or whether they belongs under one of the underlying problem classes represented from figure as c_1 - c_n.

As GAN modules are being initialized, with a class-conditional latent vector generator were set up via learning probability distribution between images in latent space for various classes. Furthermore, all weights of generator as well as discriminator were fine-tuned via performing traditional adversarial training, In case of representation from illustrated Figure 3.7(c) in which provides the entirety of the framework in which, the BAGAN training approach appears to be organized under three key steps

showcased from Figure 3.7: a) auto-encoder training, followed by b) GAN initialization, and ultimately c) adversarial training for the dataset images. Auto-encoder training. With regards to the auto-encoder which is trained via employing all images within that of the training dataset. The auto-encoder has no explicit class knowledge, it processes all images from majority and minority classes unconditionally. In this work we apply loss minimization for the auto-encoder training.
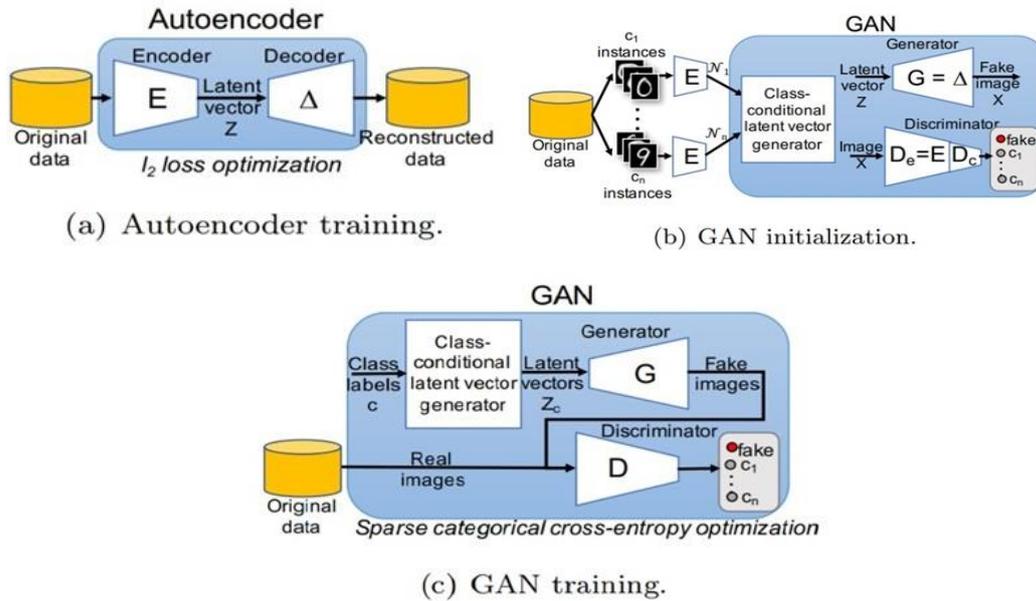

(a) Autoencoder training.

(b) GAN initialization.

(c) GAN training.

Figure 3.7 - Proposed BAGAN Architecture [76]

3.6   Datasets employed

We collected 9 publicly available anonymized image datasets from www.kaggle.com (A data repository database), which is a website that generates and collects sample notes and reports from different transcriptionists and clinical users. The dataset is divided into two parts: training data and testing data. The List of Datasets are:

1. MNIST
2. Fashion MNIST
3. CIFAR-10
4. Covid-19 X-Ray Images (Covid-19 and Non-Covid-19)
5. Covid-19 CT-Scan Images (Covid-19 and Non-Covid-19)
6. Brain Tumour (Tumour and No-Tumour)
7. Malaria (Parasitized and Uninfected cells)
8. Skin Cancer (Benign and Malignant)
9. Cell Nuclei (Marked border and cluster border)

Brain Tumour   Cell Nuclei   Chest X-ray

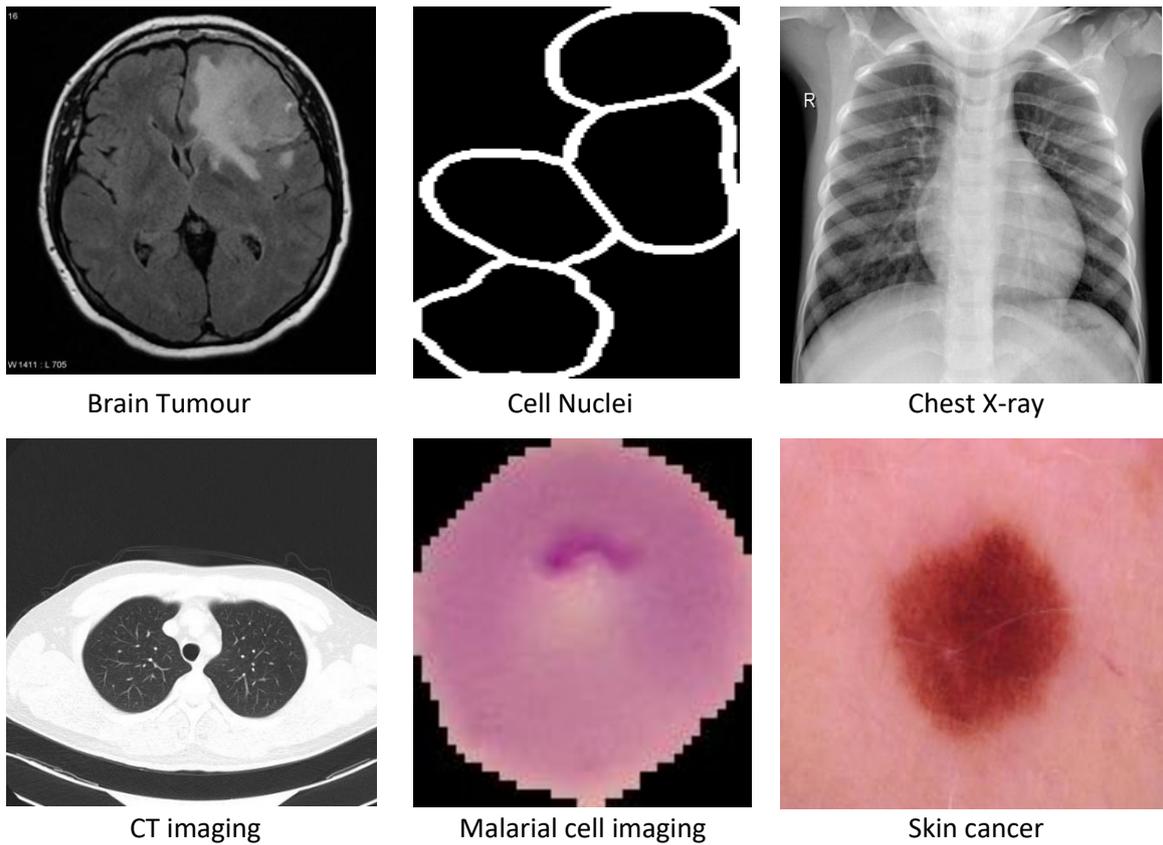CT imaging   Malarial cell imaging   Skin cancer

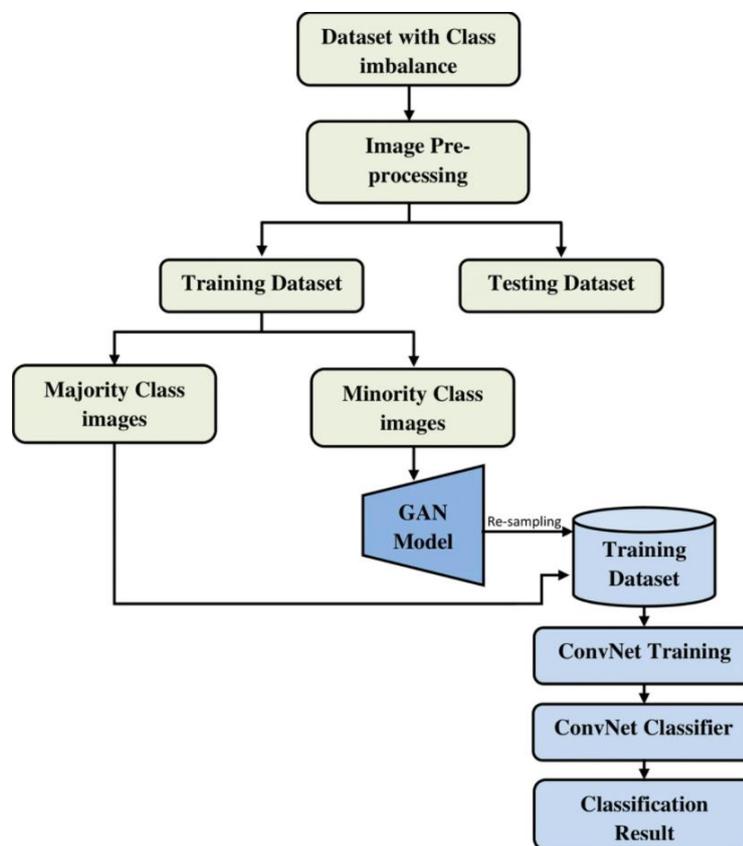Figure 3.8. Sample images from original datasets

Figure 3.9 - Flowchart presenting the complete GAN-based class imbalanced sampling

Chapter 4

Results

4.1   Introduction to this chapter

The chapter explores on handling the class imbalance datasets pertaining to medical images and was correspondingly verified using the benchmark datasets from the rendered accuracy of each datasets. Furthermore, the study validated a comparative assessment between the balanced datasets with that of imbalanced datasets. For the present experimental investigation, 'Purposeful Imbalancing' was carried out for our experiment in order for comparing the balance with that of imbalance. GAN was deployed to produce synthetic images of imbalanced classes of datasets. Furthermore, the comparison between balanced and imbalanced data was determined from the achieved accuracy from different CNN classifiers/ classification models.

For a well-throughput understanding of our experiment so far achieved, the chapter provides the achieved experimental outcomes comprising of intermediate steps for the present study alongside with final achieved values from each of the classification metrics that was achieved under each datasets that were utilized for the study. The achieved results are furthermore compared as well as critically determined/ analysed for deriving the necessary insights that addresses the significance of BAGAN in this particular regards was regarded as augmentation tool for restoration of balance within that of the imbalanced datasets. This particular task in image imbalance scenario is quite challenging as very few of the minority-class images cannot be considered enough for train GAN in general. In order for overcoming this issue the present experiment has included 6 medical datasets alongside for its adversarial training involving all the available images present in both majority & minority classes for this current experiment.

4.2   Classification Results Obtained

In this following section, we employed different data models which comprised a total of 6 different medical datasets along with 3 different benchmark datasets. Upon considering them, the training data sets comprised a total of 200 images under every single datasets of the 6 medical datasets. Of which, 80 images were subjected for training followed by 20 images were employed for both classes (i.e.) testing of each class.

Whilst considering the remaining three benchmark datasets namely- MNIST (60k images for training and 10 K for testing), Fashion MNIST and CIFAR-10 (50k for training and 10k for testing images). The validation accuracy of the CNN achieved from the performed results are presented in the forthcoming tables for both balanced and imbalanced datasets.

4.2.1   Classification outcomes for the Balanced Dataset

Before observing the accuracy of the classifier models pertaining to imbalanced dataset, we provide the achieved outcome for the balanced datasets in the initial course of the experiment. Compared to the validation accuracy achieved from the sample medical image datasets, benchmark datasets were presented with better accuracy outcomes, thereby validating the experimental process to proceed in correct pattern in this regard. Below table provides a detailed metrics of CNN accuracy for balanced datasets for all the 9 different types of datasets rendered.

Table 4.1- CNN accuracy of balanced datasets

Models/Datasets	VGG 16	VGG 19	InceptionV3	Inception ResNetV2	DenseNet 121	MobileNetV2	DenseNet 169	ResNet 50
MNIST	0.98	0.99	0.98	0.97	0.96	0.98	0.98	0.98
Fashion-MNIST	0.96	0.98	0.94	0.95	0.96	0.95	0.96	0.96
CIFAR-10	0.95	0.94	0.95	0.93	0.95	0.96	0.95	0.97
Chest X-Ray	0.80	0.97	0.68	0.78	0.60	0.89	0.80	0.88
Chest CT Scan	0.80	0.97	0.82	0.88	0.90	0.90	0.90	0.85
Skin Cancer	0.93	0.94	0.71	0.79	0.91	0.91	0.93	0.50
Brain Tumour	0.82	0.88	0.74	0.81	0.93	0.92	0.95	0.60
Malaria Cell	0.90	0.96	0.88	0.90	0.95	0.92	0.97	0.53
Cell Nucleus	0.93	0.95	0.86	0.89	0.94	0.90	0.97	0.90

4.2.2   Classification   outcomes   for   the   Imbalanced   Medical   Datasets

This section provided a well-throughput summary that is obtained pertaining with Validation Accuracy, of all the sample classifier that were employed for the variations among 6 different medical datasets that were considered for the present experiment. As discussed in previous chapter 3, the two variations of each dataset considered namely comprising of 'Original Dataset (Balanced)', 'Imbalanced' (after removing instances of some classes), and 'Balanced using BAGAN' (resulting from data augmentation in case for minority classes by images generated from BAGAN). The classifier were trained separately for all of these individual datasets as the trained models are employed to further evaluate. However, as previously mentioned in, no change were made to the test set of these datasets. The results achieved in regards from the benchmark datasets MNIST, FMNIST and CIFAR-10 were documented under separate sub-sections in the forthcoming experimental part, with an overall observations was carried out with analysis & discussion of the rendered outcomes were carried out in detail. The present section details on the imbalanced medical datasets that were utilised for this experiment. We performed two different variations for each data set, with variation-1 (class -I 80, class –II 20) and (20, 80) for variation-II.

Table 4.2 - Description of the Imbalanced Medical Datasets

Training Set	Testing Set:
Variation 1: Class 1 – 80 Images Class 2 – 20 Images **Variation 2:** Class 1 – 20 Images Class 2 – 80 Images	**20 Images (under Each Class)** which appeared same for all the 6 medical datasets

The medical dataset appears to be a highly imbalanced sort of dataset derived from numerous medical images, beginning with chest x-ray images (Covid-19 & Non Covid-19); CT scan of the chest (Covid-19 & Non_Covid-19); Skin cancer (benign & malignant); brain tumour (tumour & non-tumour); malarial cells (infected and non-infected) and cellular nucleus imaging (marked border and cluster border). For this investigation for 9 different CNN model was subjected for determining the validation accuracy of the imbalanced dataset images were studied for both the variations and ascertained in the below table.

Table 4.3- CNN accuracy of imbalanced medical datasets

Models/Datasets	Variations	VGG16	VGG19	InceptionV3	InceptionResNet V2	MobileNet-V2	DenseNet121	DenseNet169	ResNet50
Chest X-Ray	Variation-1	0.84	0.74	0.68	0.66	0.75	0.82	0.62	0.74
	Variation-2	0.76	0.62	0.56	0.62	0.80	0.72	0.70	0.50
Chest CT Scan	Variation-1	0.50	0.70	0.62	0.72	0.80	0.73	0.65	0.88
	Variation-2	0.70	0.72	0.84	0.78	0.88	0.82	0.88	0.93
Skin Cancer	Variation-1	0.70	0.66	0.70	0.60	0.78	0.70	0.68	0.65
	Variation-2	0.57	0.72	0.65	0.67	0.76	0.71	0.82	0.50
Brain Tumour	Variation-1	0.50	0.71	0.68	0.70	0.75	0.69	0.90	0.60
	Variation-2	0.50	0.70	0.65	0.67	0.78	0.70	0.68	0.50
Malaria Cell	Variation-1	0.50	0.80	0.71	0.68	0.90	0.84	0.88	0.45
	Variation-2	0.82	0.86	0.75	0.70	0.82	0.80	0.90	0.73
Cell Nucleus	Variation-1	0.75	0.78	0.76	0.70	0.84	0.80	0.95	0.95
	Variation-2	0.70	0.80	0.78	0.66	0.86	0.78	1.0	0.88

4.2.3 Classification outcomes for the Imbalanced Benchmark Datasets

Table 4.4 - Description of the Imbalanced Benchmark Datasets

Training Set	Testing Set:
Variation 1: Class 1 – 80 Images Class 2 – 20 Images **Variation 2:** Class 1 – 20 Images Class 2 – 80 Images	**20 Images (under Each Class)** which appeared same for all the 6 medical datasets

For the imbalanced benchmark datasets pertaining to MNIST, FMIST, CIFAR-10. Since the actual training and testing images, is laborious for an academic setting, thus we carried out for only 2 different classes at a time instead of performing training classes simultaneously, for every variation only 2 classes were compared for a variation class (i.e.) a total of 9 different variations were performed for benchmark datasets (class-I- IX). The reason for comparison of only two classes at a time is primarily to avoid making randomly class imbalanced datasets, the present experiment performed benchmark datasets in binary classes (i.e.) 0*1; 1*2, 2*3,.......8*9, thereby comprising a total of 9 variations. The accuracy for the CNN classification 9 different variations performed for MNIST, Fashion MNIST & CIFAR-10 datasets were shown in the table 4.5

Table 4.5 - CNN accuracy of imbalanced benchmark datasets

MNIST									
Variations	**0 X 1**	**1 X 2**	**2 X 3**	**3 X 4**	**4 X 5**	**5 X 6**	**6 X 7**	**7 X 8**	**8 X 9**
VGG16	0.68	0.67	0.65	0.68	0.66	0.66	0.70	0.69	0.65
VGG19	0.69	0.68	0.66	0.67	0.67	0.65	0.69	0.70	0.66
InceptionV3	0.67	0.66	0.64	0.66	0.65	0.66	0.68	0.68	0.65
InceptionResNet V2	0.68	0.67	0.65	0.65	0.66	0.67	0.71	0.68	0.66
DenseNet169	0.68	0.66	0.70	0.69	0.65	0.66	0.72	0.73	0.65
DenseNet121	0.65	0.68	0.69	0.68	0.67	0.65	0.70	0.71	0.67
MobileNet-V2	0.66	0.67	0.68	0.66	0.65	0.67	0.69	0.69	0.70
ResNet50	0.67	0.68	0.66	0.68	0.65	0.66	0.71	0.68	0.69
Fashion MNIST									
Variations	**0 X 1** T-shirt/top X Trouser	**1 X 2** Trouser X Pullover	**2 X 3** Pullover X Dress	**3 X 4** Dress X Coat	**4 X 5** Coat X Sandal	**5 X 6** Sandal X Shirt	**6 X 7** Shirt X Sneaker	**7 X 8** Sneaker X Bag	**8 X 9** Bag X Ankle boot
VGG16	0.68	0.67	0.65	0.68	0.66	0.66	0.70	0.69	0.65

VGG19	0.69	0.68	0.66	0.67	0.67	0.65	0.69	0.70	0.66
InceptionV3	0.67	0.66	0.64	0.66	0.65	0.66	0.68	0.68	0.65
InceptionResNet V2	0.68	0.67	0.65	0.65	0.66	0.67	0.71	0.68	0.66
DenseNet169	0.67	0.68	0.66	0.68	0.68	0.66	0.70	0.67	0.65
DenseNet121	0.68	0.69	0.65	0.67	0.69	0.67	0.69	0.68	0.66
MobileNet-V2	0.66	0.67	0.66	0.68	0.70	0.68	0.67	0.67	0.65
ResNet50	0.68	0.66	0.67	0.69	0.68	0.70	0.66	0.68	0.68

CIFAR-10									
Variations	**Plane X Car**	**Car X Bird**	**Bird X Cat**	**Cat X Deer**	**Deer X Dog**	**Dog X Frog**	**Frog X Horse**	**Horse X Ship**	**Ship X Truck**
VGG16	0.66	0.67	0.70	0.69	0.68	0.68	0.70	0.68	0.72
VGG19	0.68	0.70	0.71	0.68	0.66	0.69	0.71	0.69	0.70
InceptionV3	0.65	0.68	0.66	0.67	0.64	0.65	0.69	0.67	0.68
InceptionResNet V2	0.66	0.69	0.68	0.66	0.65	0.67	0.68	0.67	0.67
DenseNet169	0.68	0.65	0.71	0.66	0.64	0.70	0.68	0.70	0.69
DenseNet121	0.67	0.68	0.70	0.65	0.66	0.69	0.67	0.71	0.74
MobileNet-V2	0.68	0.66	0.67	0.64	0.70	0.72	0.68	0.67	0.68
ResNet50	0.68	0.67	0.66	0.69	0.68	0.70	0.65	0.69	0.67

4.4 Visual Outcomes achieved via BAGAN Generated Balanced Dataset

From the previous sections presenting on performance of different CNN classifier for both medical and benchmark dataset and was further confirmed using quantitative values presented as accuracy. Result in this section provides a comparative analysis in discussion section confirmed the role of BAGAN based augmentation for the classification of imbalanced datasets. However, the visual quality of synthetic images generated by BAGAN can also support the importance of GAN in enhancing the accuracy of image datasets. Here, BAGAN, a version of GAN, produced visually synthetic images of medical images via conditional learning. These images that are generated and furthermore rendered by BAGAN is showcased in below figure 4.1. These generated images by this method overcame artifacts problem prevalent in other GANs version. Imbalance both medical or benchmark the variations are 80 * 20 in order for making it balance, as it require 60 more images of minority classes in order for balancing the class imbalanced datasets. In order for achieving this, BAGAN was performed and GAN generated images are presented in Figure below.
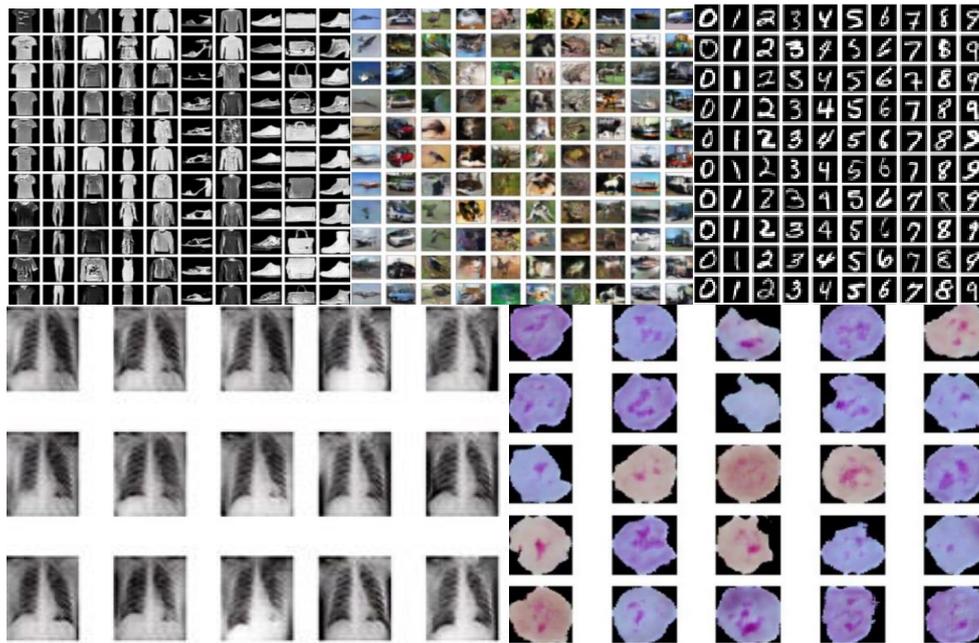
Figure 4.1. Sample BAGAN generated images

4.4.1 Classification outcomes for the BAGAN Generated Balanced Dataset

From the BAGAN generated synthetic datasets have a both of its classes comprising a total of 80 images for training and 20 images for testing in each class. The CNN performed efficiently when on comparison with that of balanced datasets that were generated by BAGAN in regards to the class-imbalanced datasets. For the benchmark the outcome which showed significant accuracy and were presented in the below table.

Table 4.6 - CNN accuracy of BAGAN generated balanced datasets

Models/Datasets	Variations	VGG16	VGG19	InceptionV3	InceptionResNetV2	MobileNet-V2	DenseNet121	DenseNet169	ResNet50
MNIST	*Variation-1 (2 x 3)*	0.93	0.95	0.92	0.92	0.95	0.90	0.93	0.95
	Variation-2 (6 x 7)	0.95	0.96	0.95	0.94	0.95	0.92	0.95	0.94
Fashion MNIST	*Variation-1 (1x2)*	0.93	0.92	0.91	0.89	0.94	0.90	0.91	0.91
	Variation-2 (6x7)	0.91	0.93	0.90	0.88	0.95	0.91	0.88	0.88
CIFAR-10	*Variation-1 (Dog X Frog)*	0.93	0.95	0.90	0.91	0.92	0.93	0.91	0.95

	Variation (Frog X Horse)	0.95	0.96	0.92	0.94	0.92	0.96	0.94	0.92
Chest X-Ray	*Variation-1*	0.88	0.94	0.75	0.72	0.80	0.89	0.71	0.75
	Variation-2	0.80	0.95	0.68	0.76	0.86	0.90	0.73	0.68
Chest CT Scan	*Variation-1*	0.78	0.80	0.71	0.80	0.91	0.83	0.92	0.94
	Variation-2	0.71	0.83	0.82	0.81	0.93	0.89	0.94	0.97
Skin Cancer	*Variation-1*	0.80	0.88	0.79	0.85	0.85	0.89	0.81	0.84
	Variation-2	0.83	0.90	0.84	0.79	0.84	0.87	0.83	0.80
Brain Tumour	*Variation-1*	0.74	0.88	0.81	0.83	0.84	0.84	0.81	0.82
	Variation-2	0.81	0.85	0.83	0.81	0.80	0.87	0.82	0.78
Malaria Cell	*Variation-1*	0.87	0.90	0.83	0.72	0.87	0.90	0.96	0.81
	Variation-2	0.90	0.95	0.88	0.89	0.82	0.85	0.95	0.87
Cell Nucleus	*Variation-1*	0.90	0.98	0.88	0.75	0.80	0.89	0.90	0.91
	Variation-2	0.93	0.94	0.91	0.81	0.89	0.82	0.92	0.90

From the achieved outcomes of the balanced datasets of GAN generated synthetic images, facilitated in improving the accuracy of the classifiers. This could be addressed from the above table showcasing a significant improvement in its accuracy across both medical and benchmark dataset images. Among the medical image datasets, the models of CNN classifiers were found to exhibit better outcomes on the basis of the nature of dataset images employed.

For instance, CNN classifier namely VGG19 classifier exhibited significant improvement in chest x-ray (Variation 1 - 0.94 & Variation 2 - 0.95); 2nd one (Variation 1- 0.88 and Variation 2- 0.85) for brain tumour images & 3rd one (Variation 1 - 0.98 and Variation 2 - 0.94) for cell nucleus images; likewise, ResNet50 classifier (Variation 1 - 0.94, 0.97) for CT scan image of chest, DenseNet121 (Variation 1 - 0.89 and Variation 2 - 0.87) for skin cancer images, Densenet169 (Variation 1 - 0.96 and Variation 2- 0.95) for malarial cell imaging. From the inferred data involving benchmark datasets appeared to have exhibited better accuracy in all the CNN classifier models observed from the experiment.

Chapter 5

Conclusions and Future Work

5.1   Conclusion

One of the contemplating issues in digital image processing, especially concerning with Image classification datasets appears to be most often concerned with imbalanced datasets, as this particular characteristic tends to negatively impact the accuracy of deep-learning classifier in this regard we employed CNN classifiers.

Class imbalanced learning on complex data has long been a source of difficulty in machine learning research. As a solution, we experimented with various oversampling strategies on a clinical dataset. The imbalance issue in the classification task of these imbalanced data sets was solved using Generative Adversarial Networks as an oversampling technique. Furthermore, GANs outperformed simple oversampling techniques such as random oversampling and SMOTE in learning the underlying distribution and thus contributing to the learning of machine learning models. One of the results of our experiments was that using the dimensionality reduction technique degraded the classifier's performance due to data under-representation. Furthermore, some learning models are less reliant on oversampling methods. This study limits the reduction in computational costs of GAN networks with dimensionality reduction tools that could help improve learning in classifiers and also only one combination of oversampling technique is explored.

In this paper, we proposed the Balancing GAN (BAGAN) as an augmentation tool for restoring balance from imbalanced datasets. This is quite a challenging investigation as the few minority-class images may not be enough for training GAN in the first place. We overcame the problem by incorporating all available images of majority and minority classes into the adversarial training. The generative model incorporates elements that are beneficial to majority classes and employs them to generate images for minority classes. In the latent space, we used class conditioning to direct the generation process toward a specific class. The encoder module of an auto-encoder is used to initialise the generator in the GAN, allowing us to learn an accurate class-conditioning in the latent space. We compare the proposed methodology to state-of-the-art GANs and show that when trained with an imbalanced dataset, BAGAN produces superior images. In the current experimental investigation involving with 9 different datasets comprising of 6 medical image datasets alongside with 3 benchmark datasets were studied using purposeful imbalancing for the sample images rendered. Furthermore, the study enhanced the accuracy of the CNN classifier when compared to their accuracy in imbalanced image datasets.

5.2   Limitations

Since the experiments in this thesis are performed on a personal computer and not a high end server farm there are limitations with calculation speed and memory constrains which limits the size of the models. In Cho et al. argued that the size of their model greatly influenced the learning capabilities and discussed lower limits of their model sizes for the training to be effective on their dataset. This was the main contribution of Bahdanau, Cho and Bengio's attention model that managed to increase the learning capacity of the model without just scaling the model up. In this thesis the models were forced to be relatively small and thereby reducing the learning capacity. Because GAN training was utilised, the memory constraints were exacerbated. Since GAN training requires a discriminator to be

also fit into memory the generator have to be reduced to make room for the discriminator. One drawback in this thesis is that it has not made an exhaustive hyper parameter search. Some of the larger experiments take months of computing time to complete and it would be desirable but not feasible to repeat these experiments with different configurations of hyper parameters.

5.3   Future work

In this thesis GAN training showed great result over the trained datasets. However, there are plenty of additional steps to establish the effectiveness of GAN training in learning that this thesis did not cover. The experiments have to be repeated on multiple datasets to establish that this training method is not only effective on certain types of data and is a universally effective training method. There is even some indication in this thesis that certain training data effects the effectiveness of the training method. Even though this thesis demonstrated that GAN training can be effective it contains no evidence that it will be for all models. For future work multiple models needs to be tested to establish if the training method is only effective for certain models or can be effective for all types of Models. There also have to be further testing of how the discriminator effects the training. Additionally, GAN training is not one framework and there is multiple types of GANs, in this thesis the framework used was heavily influenced by the BAGAN framework but there are various frameworks that should be tested in future experiments. Furthermore, by deploying different GAN models could be beneficial for generating high quality synthetic images for datasets from a low resolution medical imaging data set. This ensures to improve the classification accuracy and improve the imbalance datasets accuracy.

Chapter 6

Reflection

In this thesis we studied classification tasks pertaining to the presence of class imbalanced datasets. The task in particular arises to be a general issues across various applications while researchers are indulged to explore more on the under-represented (or in other words minority) classes. For instance, this could be witnessed over a wider set of applications pertaining to fraud detection, medical monitoring as well as diagnosis, risk management, text categorization, information filtering & filtering. However, there lies numerous set of standardized approaches for classification for such task, majority cases were reported with poor generalisation performance when it concerns with that of the minority classes. From my personal reflection so far concerning the present investigation facilitated the researcher towards understanding the significance of GAN. We carried out a proposed balancing GAN (BAGAN) as augmentation approach/ tool for restoring balance on imbalanced datasets. From personal reflection of the study, it appeared quite a challenge since very few minority-class images tend to be not be enough for GAN training. This issue in particular was overcome at the time of adversarial training involving all of the available images on majority as well as minority classes. GAN facilitates in learning the useful features from that of the majority classes & later was employed for generating images for the minority classes. In addition, the generator utilized in the case of GAN is initialized with that of an encoder module involving auto-encoder which enables to learn as well as employ accurate class-conditioning over the latent space. Furthermore, the proposed methodology demonstrated BAGAN generated images showcasing better quality as well as accuracy upon training with imbalanced dataset among both medical and benchmark datasets.

References

[1]   B.T. Nugraha, S.F. Su, Fahmizal. Towards self-driving car using convolutional neural network and road lane detector. *Proceedings of the 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, ICACOMIT* 2017. 2017; 2018-January: 65–9.

[2]   S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis", *Journal of Big Data*, vol. 6, no. 1, 2019. Available: https://doi.org/10.1186/s40537-019-0276-2.

[3]   A. Gutierrez, A. Anzoátegui, L. Superego, C. Rubio, I. Rancic, L. Lenza. A Benchmarking of learning strategies for pest detection and identification on tomato plants for autonomous scouting robots using internal databases. *Journal of Sensors*. 2019. https://doi.org/10.1155/2019/5219471.

[4]   L. Santos, F.N Santos, P.M Oliveira, P. Shinde. Deep learning applications in agriculture: a short review. *Advances in intelligent systems and computing*; 2020. https://doi.org/10.1007/978-3-030-35990-4_12.

[5]   T. Wang, Y. Chen, M. Qiao and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control", *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3465-3471, 2017. Available: 10.1007/s00170-017-0882-0.

[6]   M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation", *Journal of Big Data*, vol. 6, no. 1, 2019. Available: 10.1186/s40537-019-0263-7.

[7]   Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*.1998. http://ieeexplore.ieee.org/document/726791/.

[8]   R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition. IEEE*; 2014. p. 580–7. http://ieeexplore.ieee.org/document/6909475/.

[9]   J. Long, E. Shelhamer, T. Darrell. Fully convolutional networks for semantic segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2015. p. 3431–40. http://arxiv.org/abs/1605.06211.

[10]   A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017. Available: 10.1145/3065386.

[11]   K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR* 2015–Conference Track Proceedings. 2015; 1–14.

[12]   C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al. Going Deeper with

Convolutions. *CoRR*. 2014; abs/1409.4. https://arxiv.org/abs/1409.4842.

[13]   K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*. 2016. p. 770–8. http://arxiv.org/abs/1512.03385.

[14]   C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the inception architecture for computer vision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE; 2016. p. 2818–26. http://arxiv.org/abs/1512.00567.

[15]   G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger. Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2017. p. 2261–9. http://arxiv.org/abs/1608.06993.

[16]   M. Buda, A. Maki and M. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks", *Neural Networks*, vol. 106, pp. 249-259, 2018. Available: 10.1016/j.neunet.2018.07.011.

[17]   S. Al-Stouhi and C. Reddy, "Transfer learning for class imbalance problems with inadequate data", *Knowledge and Information Systems*, vol. 48, no. 1, pp. 201-228, 2015. Available: 10.1007/s10115-015-0870-3.

[18]   A. Ali, S.M. Shamsuddin, A.L. Ralescu. Classification with class imbalance problem: a review. *International Journal of Advanced Software Computer Application*. 2015; 7:176–204.

[19]   J. Zhang, Y. Xia, Q. Wu, Y. Xie. Classification of medical images and illustrations in the biomedical literature using synergic deep learning. 2017. http://arxiv.org/abs/1706.09092.

[20]   Q. Dong, S. Gong, X. Zhu. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019; 41:1367–81. https://ieeexplore.ieee.org/document/8353718.

[21]   Y. Zhang, B. Li, H. Lu, A. Irie, X. Ruan. Sample-Specific SVM learning for person re-identification. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2016. p. 1278–87. http://ieeexplore.ieee.org/document/7780512/.

[22]   M.M Sawant, K.M Bhurchandi. Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging. *Artificial Intelligence Review*. 2019; 52:981–1008. https://doi.org/10.1007/s10462-018-9661-z.

[23]   E. Mostafa, A. Ali, N. Alajlan, A. Farag. Pose Invariant Approach for Face Recognition at Distance. *Springer*; 2012. p. 15–28.

[24]   Z. He, W. Zuo, M. Kan, S. Shan, X. Chen. AttGAN: Facial attribute editing by only changing what you want. *IEEE transactions on image processing.*2019; 28:5464–78. https://ieeexplore.ieee.org/document/8718508/65.

[25]   G. Perarnau, B. Raducanu, J.M. Alvarez. Invertible Conditional GANs for image editing. *Conference   on   Neural   Information   Processing*

Systems.http://arxiv.org/abs/1611.0635566.

[26]   R. Tao, Z. Li, R. Tao, and B. Li. ResAttr-GAN: Unpaired deep residual attributes learning for multi-domain face image translation. *IEEE Access*. 2019; 7:132594–608.

[27]   W. G. Hatcher and W. Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends," *IEEE Access on Human-Centered Smart Systems and Technologies*, vol.6, pp. 24411-24431, 2018.

[28]   B. Harangi, "Skin lesion detection based on an ensemble of deep convolutional neural network," *Journal of Bio-Medical Informatics*, vol. 86, pp.25-32, 2018.

[29]   W. Zhao , "Research on the Deep Learning of the Small Sample Data based on Transfer Learning," in *American Institute of Physics (AIP) conference*. *Proceedings. AIP*, 2017.

[30]   J. Wang and L. Perez, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," arXiv preprint arXiv:1712.04621 [cs.CV], 2017.

[31]   M. Olsen, R. Berk and A. J. Wyner, "Modern Neural Networks Generalize on Small Data Sets" in *32nd International Conference on Neural Information Processing Systems (NIPS).* Proceedings. NIPS, 2018, pp. 3623-3632.

[32]   "Perspectives on Image Processing," J. G Verne. [Online].   Available: https://www.semanticscholar.org/paper/Perspectives-on-Image-ProcessingVerne/cc43a71e05cfc49ab0777b82ca94d181f779149f.

[33]   Z. Hussain, F. Gimenez, D. Yi and D. Rubin, "Differential Data Augmentation Techniques for Medical Imaging Classification Tasks," *American Medical Informatics Association (AMIA) Annual Symposium Proceedings Archive*, vol. 2017, pp. 979-984, 2018.

[34]   P. Ganesan , S. Rajaraman, R. Long, B. Ghoraani and S. Antani, "Assessment of Data Augmentation Strategies Toward Performance Improvement of Abnormality Classification in Chest Radiographs" Proc. in *IEEE Engineering in Medicine and Biology Conference (EMBC), Berlin, Germany.* 23 – 27 July 2019.IEEE, 2019, pp. 841 – 844.

[35]   J. Wang and L. Perez, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," arXiv preprint arXiv:1712.04621 [cs.CV], 2017.

[36]   M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification",*Neurocomputing* vol. 321, pp.321-331, 2018.

[37]   "Poem Generator". Masterpiece Generator. [Online]. Available: https://www.poem-generator.org.uk/.

[38]   I. Goodfellow, A. J. Pouget, M. Mehdi, B. Xu, F. W.Farley, S. Ozair, A. Courviller and Y. Bengio, "Generative adversarial nets," in *Conference on Neural Information Processing Systems.NIPS*, 2014, pp. 2672- 2680.

[39]   T. Karras, S. Laine and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," arXiv preprint arXiv: 1812.04948 [cs.NE], 2018.

[40]   D. Kaeli, P. Mistry, D. Schaa and D. P. Zhang, "Image Classification" *Heterogeneous Computing with OpenCL 2.0*, chapter. 9, pp. 213-228, 2015.

[41]   A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks*," Magazine Communications of the ACM*, vol. 60, Issue. 6, pp. 84-90, 2017.

[42]   T. J.Brinker, A. Hekler, J. S.Utikel, N. Grabe, D. Schandendorf, J. Klode, C. Berking, T. Steeb, A. H.Enk and C. V.Kalle, "Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review," *Journal of Medical Internet Research*, vol. 20, no. 10, 2018.

[43]   "A Gentle Introduction to Transfer Learning for Deep Learning". *Machine Learning Mastery*. [Online]. Available: https://machinelearningmastery.com/transferlearning-for-deep-learning/.

[44]   S. Feng, H. Zhou and H. Dong, "Using deep neural network with small dataset to predict material defects," *Materials Design*, vol 162, pp.300-310, 2019.

[45]   H. W. Ng, V. D. Nguyen, V. Vonikakis and S. Winkler, "Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning," in *Proceedings 77 of the 2015 ACM on International Conference on Multimodal Interaction, ACM*, 2015, pp. 443-449.

[46]   W. Zhao , "Research on the Deep Learning of the Small Sample Data based on Transfer Learning," in *American Institute of Physics (AIP) conference. Proceedings. AIP*, 2017, vol. 1864.

[47]   L. Nanni, S. Ghidoni and S. Brahnam, "Ensemble of convolutional neural networks for bioimage classification."[online] Available: https://doi.org/10.1016/j.aci.2018.06.002

[48]   R. Minetto, M. P. Segundo and S. Sarkar, "Hydra: An Ensemble of Convolutional Neural Networks for Geospatial Land Classification,"*IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, pp. 6530 - 6541, 2019.

[49]   L. Taylor and G. Nitschke, "Improving Deep Learning using Generic Data Augmentation," arXiv preprint arXiv: 1708.06020v1 [cs.LG], 2017.

[50]   F. Perez , C. Vasconcelos, S. Avila and E. Valle, "Data Augmentation for Skin Lesion Analysis" 2.0 Context-Aware Operating Theatres, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis pp.303-311, 2018.

[51]   T. C. Pham, C. M. Luong, M. Visani, and V. D. Hoang, "Deep CNN and Data Augmentation for Skin Lesion Classification," *Lecture Notes in Computer Science*, pp. 573582, 2018.

[52]   A. Udrea and G. D. Mitra , "Generative Adversarial Neural Networks for Pigmented and Non-Pigmented Skin Lesions Detection in Clinical Images," in *International Conference on Control Systems and Computer Science(CSCS), IEEE*, 2017 , pp.364-368.

[53]   V. Sampath, I. Maurtua, J. Aguilar Martín and A. Gutierrez, "A survey on generative adversarial networks for imbalance problems in computer vision tasks", *Journal of Big Data*, vol. 8, no. 1, 2021. Available: 10.1186/s40537-021-00414-0.

[54]   A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman and P. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection", *IEEE Access*, vol. 8, pp. 91916-91923, 2020. Available: 10.1109/access.2020.2994762.

[55]   X. Yung, E. Walia and P. Babyn, "Unsupervised and semi-supervised learning with Categorical Generative Adversarial Networks assisted by Wasserstein distance for dermoscopy image Classification," arXiv preprint arXiv:1804.03700 [cs.CV], 2018.

[56]   D. Bisla, R. S. Berman, J. A. Stein and D. Polsky, "Towards Automated Melanoma Detection with Deep Learning: Data Purification and Augmentation," in the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, IEEE*, 2019.

[57]   B. Barz and J. Denzler, "Deep Learning on Small Datasets without Pre-Training using Cosine Loss," arXiv preprint arXiv: 1901.09054 [cs.LG], 2019.

[58]   M. A. Marchetti, N. C. F. Codella, S. W. Dusza, D. A. Gutman, B. Helba, A. Kalloo, N. Mishra, C. Carrera, M. E. Celebi, J. L. Defazio, N. Jaimes, A. A. Marghoob, E. Quigley, A. Scope, O. Yelmos and A. C. Halpern, "Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images," *Journal of the American Academy of Dermatology*, vol 78, pp.270-277, 2019.

[59]   "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way", *Medium*, 2021. [Online]. Available: https://towardsdatascience.com/a-comprehensiveguide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. [Accessed: 20- Sep- 2021].

[60]   M. J. Cooper, "A Deep Learning Prediction Model for Mortgage Default." [Online] Available: 10.13140/RG.2.2.21506.12487.

[61]   A. Agarwal, "Loss Functions and Optimization Algorithms. Demystified." [Online]. Available: https://medium.com/data-science-group-iitr/loss-functionsandoptimization-algorithms-demystified-bb92daff331c.

[62]   H. Heindenreich, "GAN Objective Functions: GANs and Their Variations", *Medium*, 2021. [Online]. Available: https://towardsdatascience.com/gan-objective-functionsgansand-their-variations-ad77340bce3c. [Accessed: 22- Sep- 2021].

[63]   H. Huan, P. S. Yu and C. Wang, "An Introduction to Image Synthesis with Generative Adversarial Nets," arXiv preprint arXiv: 1803.04469 [cs.CV], 2018.

[64]   "Discriminator". Machine Learning. [Online]. Available: https://developers.google.com/machine-learning/gan/discriminator.

[65]   "Generator". Machine Learning. [Online]. Available: https://developers.google.com/machine-learning/gan/generator.

[66]   "Loss Functions". Machine Learning. [Online]. Available: https://developers.google.com/machine-learning/gan/loss.

[67]   A. Ali-Gombe and E. Elyan, "MFC-GAN: Class-imbalanced dataset classification using Multiple Fake Class Generative Adversarial Network", *Neurocomputing*, vol. 361, pp. 212-221, 2019. Available: 10.1016/j.neucom.2019.06.043.

[68]   S. Wickramaratne and M. Mahmud, "Conditional-GAN Based Data Augmentation for Deep Learning Task Classifier Improvement Using fNIRS Data", *Frontiers in Big Data*, vol. 4, 2021. Available: 10.3389/fdata.2021.659146.

[69]   W. Fang, F. Zhang, V. S. Sheng and Y. Ding, "A Method for Improving CNN-Based Image Recognition Using DCGAN", *Computers, Materials & Continua*, vol. 57, no. 1, pp. 167-178, 2018. Available: 10.32604/cmc.2018.02356.

[70]   N. Chawla, K. Bowyer, L. Hall and W. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002. Available: 10.1613/jair.953.

[71]   A. Alhudhaif, "A novel multi-class imbalanced EEG signals classification based on the adaptive synthetic sampling (ADASYN) approach", *PeerJ Computer Science*, vol. 7, p. e523, 2021. Available: 10.7717/peerj-cs.523.

[72]   K. PUNTUMAPON, T. RAKTHAMAMON and K. WAIYAMAI, "Cluster-Based Minority Over-Sampling for Imbalanced Datasets", *IEICE Transactions on Information and Systems*, vol. 99, no. 12, pp. 3101-3109, 2016. Available: 10.1587/transinf.2016edp7130.

[73]   M. Pota, M. Esposito, G. De Pietro and H. Fujita, "Best Practices of Convolutional Neural Networks for Question Classification", *Applied Sciences*, vol. 10, no. 14, p. 4710, 2020. Available: 10.3390/app10144710.

[74]   I. Sutedja, "Imbalanced Data Classification Using Auxiliary Classifier Generative Adversarial Networks", *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 1068-1075, 2020. Available: 10.30534/ijatcse/2020/26922020.

[75]   C. Shorten and T. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, vol. 6, no. 1, 2019. Available: 10.1186/s40537-019-0197.

[76]   G. Huang and A. Jafari, "Enhanced balancing GAN: minority-class image generation", *Neural Computing and Applications*, 2021. Available: 10.1007/s00521-021-06163-8.

[77]   O. Elzeki, M. Shams, S. Sarhan, M. Abd Elfattah and A. Hassanien, "COVID-19: a new deep learning computer-aided model for classification", *PeerJ Computer Science*, vol. 7, p. e358, 2021. Available: 10.7717/peerj-cs.358.

[78]   A. Khan, F. Khan, S. Khan, I. Khan and M. Saeed, "Cost Sensitive Learning and SMOTE Methods for Imbalanced Data", *Journal of Applied and Emerging Sciences*, vol. 8, no. 1, p. 32, 2018. Available: 10.36785/jaes.81240.

[79]   Fanny and T. Cenggoro, "Deep Learning for Imbalance Data Classification using Class Expert Generative Adversarial Network", *Procedia Computer Science*, vol. 135, pp. 60-67, 2018. Available: 10.1016/j.procs.2018.08.150.

[80]   J. Cho and K. Yoon, "Conditional Activation GAN: Improved Auxiliary Classifier GAN", *IEEE Access*, vol. 8, pp. 216729-216740, 2020. Available: 10.1109/access.2020.3041480.

[81]   K. He, G. Gkioxari, P. Dollar, R. Girshick. Mask R-CNN. *IEEE Transactions on pattern analysis and machine intelligence*.2020.

[82]   Q. WU and S. LIAO, "Single Shot MultiBox Detector for Vehicles and Pedestrians Detection and Classification", DEStech Transactions on Economics, Business and Management, no., 2018. Available: 10.12783/dtem/apop2017/18546.

[83]   Redmon ,YOLO (YOU ONLY LOOK ONCE) Making Object detection work in Medical Imaging on Convolution detection System, *International Journal of Pharmaceutical Research*, vol. 12, no. 02, 2020. Available: 10.31838/ijpr/2020.12.02.0003.

[84]   X. Yan, H. Gong, Y. Jiang, S-T. Xia, F. Zheng, X. You, et al. Video scene parsing: an overview of deep learning methods and datasets. *Computer Vision and Image Understanding*.2018.

[85]   B. Singh, L.S. Davis. An analysis of scale invariance in object detection–SNIP. *Conference on computer vision and pattern recognition. IEEE*; 2018. p. 3578–87. https://ieeexplore.ieee.org/document/8578475/.

[86]   F. Yang, W. Choi, Y. Lin. Exploit All the Layers: Fast and Accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers. *Conference on Computer Vision and Pattern Recognition (CVPR). IEEE;* 2016. p. 2129–37. http://ieeexplore.ieee.org/document/7780603/.

[87]   O. Sevli, "A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation", *Neural Computing and Applications*, vol. 33, no. 18, pp. 12039-12050, 2021. Available: 10.1007/s00521-021-05929-4.

[88]   T. Lin, P. Goyal, R. Girshick, H.K. Dollar. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2020; 42:318–27. https://ieeexplore.ieee.org/document/8417976/.

[89]   Z. Zhong Z, L. Zheng, G Kang, S. Li, Y. Yang. Random Erasing Data Augmentation. 2017. http://arxiv.org/abs/1708.04896.

[90]   X. Wang, A. Shrivastava, A. Gupta. A-Fast-RCNN: Hard positive generation via adversary for object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE;* 2017. p. 3039–48. http://arxiv.org/abs/1704.03414.

[91]   M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet. Are GANs Created Equal? *A LargeScale Study.* ArXiv e-prints, 2017.

[92]   A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski. Plug & play generative networks: Conditional iterative generation of images in latent space. *IEEE Access;* ArXiv e-prints, November 2016.

[93]   A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier GANs. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research*, pages 2642– 2651, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

Appendix A

Performance curves of CNN models

Due to the huge number of variations and trails the performance curve of CNN Classification Models has been presented in the below figures for Balanced dataset of Chest X-Ray alone.
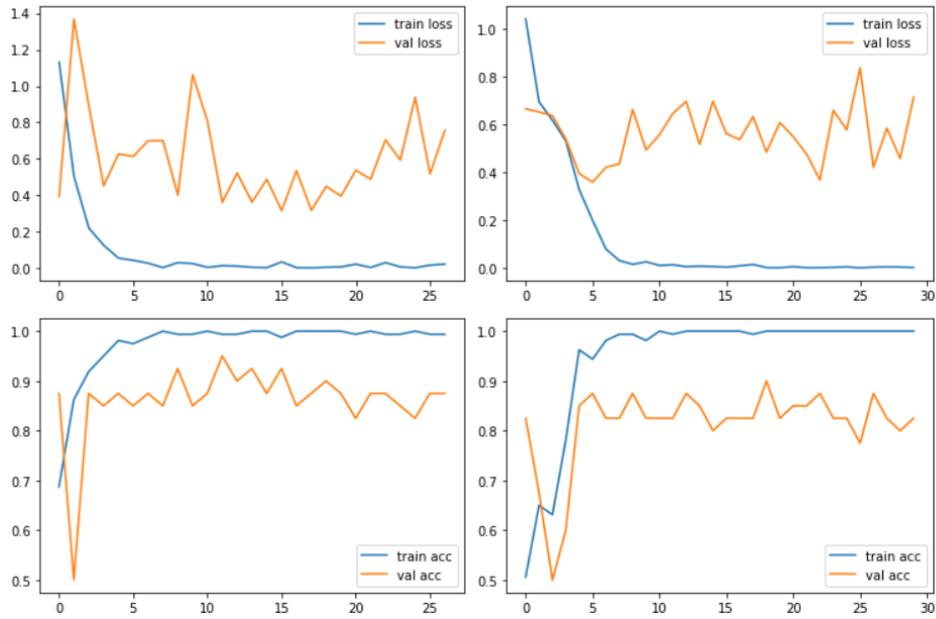

Figure A-1: Performance curve of MobileNetV2   Figure A-2: Performance curve of VGG16
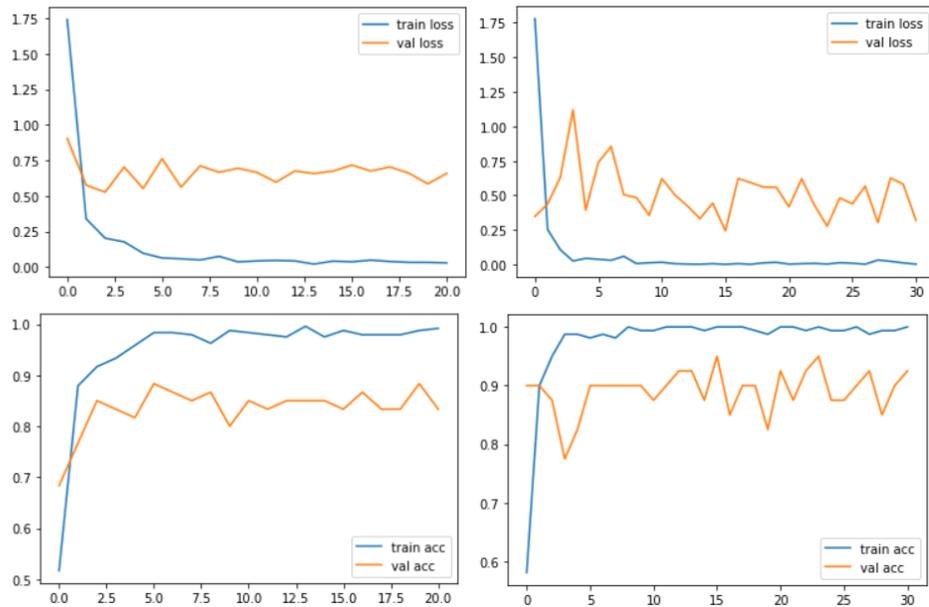

Figure A-3: Performance curve of InceptionV3   Figure A-4: Performance curve of ResNet50
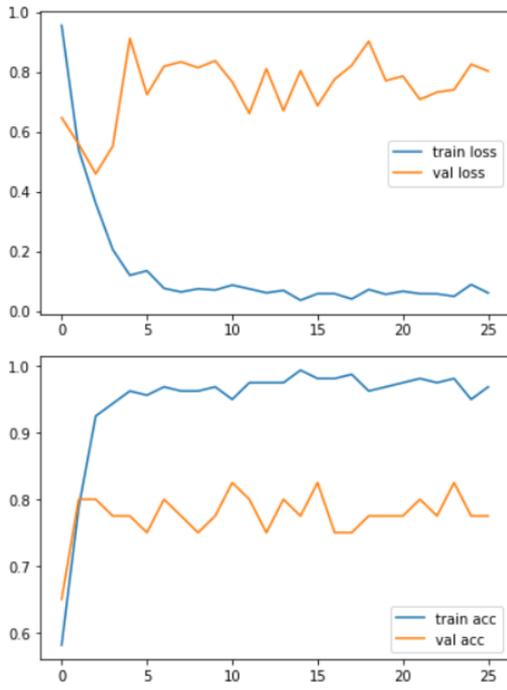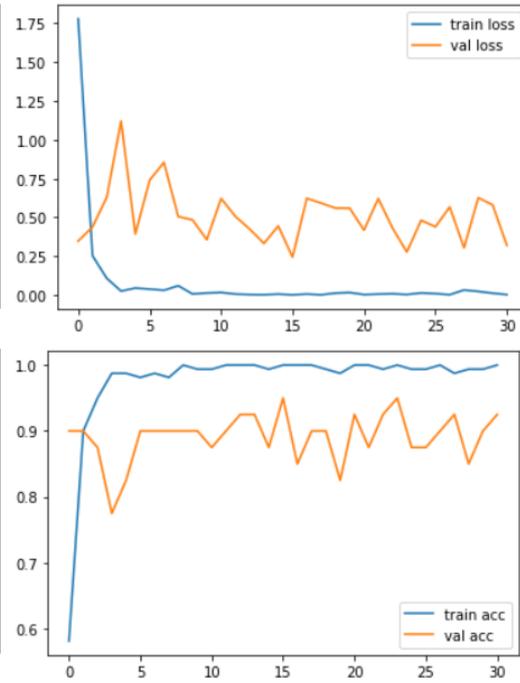
Figure A-5: Performance curve of DenseNet169   Figure A-6: Performance curve of InceptionResNetV2
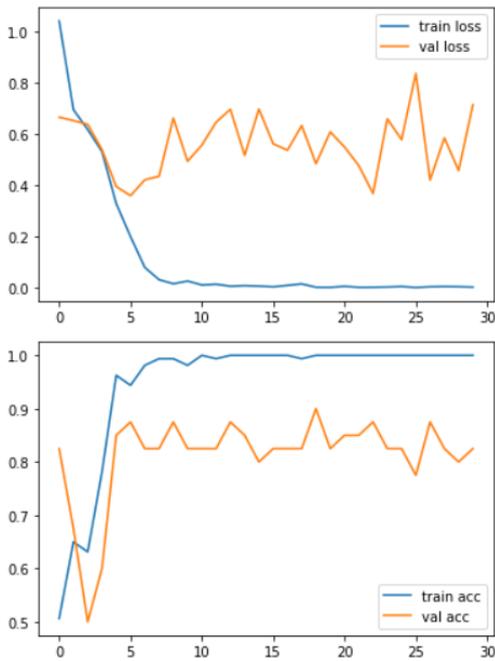

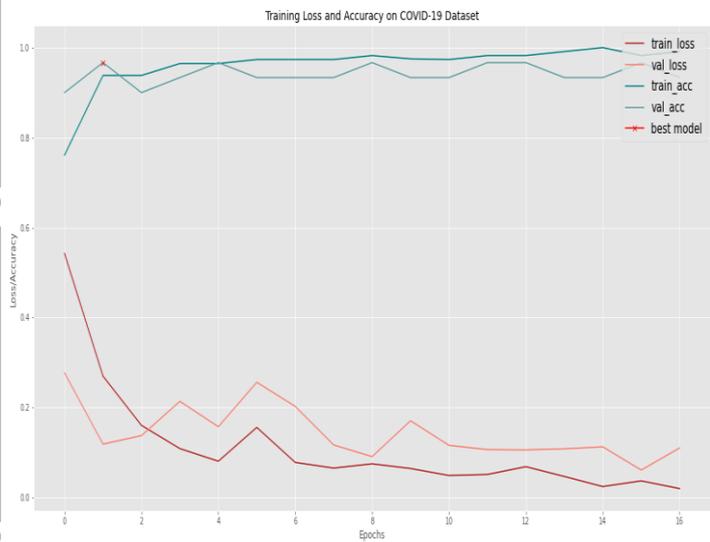Figure A-7: Performance curve of DenseNet121   Figure A-8: Performance curve of VGG19

Appendix B

The codes and scripts implemented for the purpose of this study have been made available at the following link:

https://drive.google.com/drive/folders/1wrP3NUT7R_w1Fpd3NvM9bZY4ewF21ddG?usp=sharing