Finding Location-Specific Aggregated Air Quality Index with Smartphone Images Using Deep Convolutional Neural Network

by

Joyanta Jyoti Mondal
19141016
Nowsin Kabir Rhidi
19101488

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science or Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
January 2023

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Nowsin

Joyanta Jyoti Mondal

19141016

Nowsin Kabir Rhidi

19101488

Approval

The thesis titled "Finding Location-Specific Aggregated Air Quality Index with Smartphone Images Using Deep Convolutional Neural Network" submitted by

1. Joyanta Jyoti Mondal (19141016)
2. Nowsin Kabir Rhidi (19101488)

of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science or B.Sc. in Computer Science and Engineering on January 19, 2023.

Examining Committee:

Supervisor:
(Member)

Jannatun Noor Mukta
Senior Lecturer
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Co-Supervisor:
(Member)

MEEM ARAFAT MANAB

Meem Arafat Manab
Lecturer
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Program Coordinator:
(Member)

Dr. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor and Chairperson
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Abstract

Although smartphones have already become the de facto tool for environmental health research for their ubiquity and portability, utilizing them in finding location-specific aggregated air quality index based on PM2.5 concentration is little explored in the literature to date. In this paper, therefore, we vigorously analyze the difficulties of predicting location-specific PM2.5 concentration from photos captured by smartphone cameras. Here, we particularly focus on Dhaka, the capital of Bangladesh, considering its very high level of air pollution exposure to a huge number of its dwellers. In our research, we develop a Deep Convolutional Neural Network (DCNN) and train it using more than a thousand outdoor photos captured and labeled by us. We capture the photos at various locations in Dhaka, Bangladesh, and label them based on PM2.5 concentration data extracted from the local US consulate as computed by the NowCast algorithm. During training with the dataset, our model learns a correlation index through supervised learning, which improves the model's ability to act as a Picture-based Predictor of PM2.5 Concentration (PPPC) making it capable of detecting comparable daily aggregated AQI index from a photo captured by a smartphone. Here, the computation necessary in our model is comparatively resource-efficient, as our model subsumes a much smaller number of parameters compared to most of the other alternatives. Moreover, our experimental results show that our model exhibits more robustness, for location-specific PM2.5 prediction than existing state-of-the-art models such as ViT (Vision Transformer) and INN (Involutional Neural Network) as well as other popular models that are created based on CNN, such as VGG19, ResNet50, or MobileNetV2.

Keywords: Air Quality Index, Picture-based Predictor of PM2.5 Concentration (PPPC), Deep Learning, Machine Learning.

Dedication

Firstly, we want to dedicate our work to our parents because without their support and well-wishes, we would not have come this far in our life.

Also, we want to dedicate our work to all the residents of Dhaka City.

Moreover, we want to dedicate all our efforts and challenges of our academic life to Rick Astley for his classic song: Never Gonna Give You Up.

And lastly, we would like to dedicate the work to Lionel Messi, Emiliano Martínez, and the whole Argentina Football Team, for giving the football fans the opportunity to get amazed by the phenomenal performance and for wrapping the G.O.A.T debate up by creating the golden moment of being the champion and lifting the FIFA World Cup 2022 in the hands of the G.O.A.T. As well as, for reminding us to never stop dreaming for the greater good.

(En Español: Y por último, nos gustaría dedicar el trabajo a Lionel Messi, Emiliano Martínez y toda la Selección de Fútbol de Argentina, por brindarle a los fanáticos del fútbol la oportunidad de asombrarse con la fenomenal actuación, por cerrar el debate sobre "el más grande de todos los tiempos", y por crear el momento dorado de ser el campeón y levantar la Copa Mundial de la FIFA 2022 en manos del "más grande de todos los tiempos". Y sobre todo, por recordarnos que nunca dejemos de soñar con el buenísimo.)

Acknowledgement

Firstly, all praise to the Almighty for whom our thesis is going smooth without any major interruption.

Secondly, to our advisors, Jannatun Noor and Meem Arafat Manab for their tremendous support in our work from the very beginning. Their constant help helped us throughout the whole journey.

Thirdly, we would like to express our gratitude towards Prof. Dr. A. B. M. Alim Al Islam, from Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET) for guiding us along with providing expert advice throughout the research.

Fourthly, we would like to thank Md. Farhadul Islam, and Raima Islam from School of Data and Sciences, Brac University, for being active contributors of this work alongwith us.

Fifthly, we would like to thank the anonymous reviewers of Association for the Advancement of Artificial Intelligence (AAAI) 2023 Conference for providing their expert opinion towards this work, to make the work more matured.

And, finally as well as most importantly, we want to express our gratitude to our families. Without their continuous support, it would not be possible to pursue this research work as well as this Undergraduate Degree.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ANN	Artificial Neural Network
API	Air Pollutant Index
AQI	Air Quality Index
CNN	Convolutional Neural Network
$DCNF$	Deep Convolutional Neural Field
$DCNN$	Deep Convolutional Neural Network
DCO	Dark Channel Operator
DCP	Dark Channel Prior
$DCWCN$	Double Channel Weighted Convolution Network
EPA	Environmental Protection Agency
GIS	Geographic Information System
INN	Involutional Neural Network
KNN	K-Nearest Neighbour
LBP	Local Binary Patterns
LDC	Least Developed Country
$LSTM$	Long Short-Term Memory
MAE	Mean Average Error
MSE	Mean Square Error
$NAAQS$	National Ambient Air Quality Standards
PM	Particulate Matter
$PM10$	Thoracic Particles
$PM2.5$	Fine Particles

$PPPC$	Picture-based Prediction of The PM2.5 Concentration
R^2	R-Squared
$ResNet$	Residual Network
$RMSE$	Root Mean Square Error
$RMSE$	Root Mean Square
ROI	Region of Interest
SDG	Sustainable Development Goals
SVR	Support Vector Regression
UN	United Nations
VGG	Visual Geometry Group
ViT	Vision Transformer
WHO	World Health Organization
WSN	Wireless Sensor Network
$YOLO$	You Only Look Once (An Object Detection Model)

Chapter 1

Introduction

1.1 Motivation and Research Problem

Air pollution is the flow of pollutants into the atmosphere; more specifically, these would be pollutants that are harmful to human health and the environment as a whole. It is a mixture of particulate matter (PM), gasses, and vapor-phase molecules. The most prevalent route of exposure to air pollution is the respiratory system. In the case of dust pollutants, the size of particulate matter plays a crucial role in determining the environmental health risk. PM is categorized on the basis of aerodynamic diameter. Thoracic particles (PM10) are defined as particles having a dimension of fewer than 10 micrometers, fine particles (2.5 micrometers or less), and ultra-fine particles (0.1 micrometers or less) [4]. Even millions of deaths are caused by air pollution each year, reported by the World Health Organization[1]. Nowadays, nine out of ten people breathe air that goes beyond the WHO's recommended levels for contaminants, and people in low and middle-income nations suffer the most due to this. Air pollution is caused by a combination of solid particles and gases in the atmosphere. Particles can include car emissions, toxins from industries, dust, pollen, and mold spores. Ozone, an inorganic gas, also contributes significantly to urban air pollution, leading to a smog.
Air pollutants having aerodynamic diameter of 2.5um are particularly much more dangerous as it can impair respiratory and cardiovascular functions which in the end causes serious life-threatening disease in human body. People with heart or lung conditions, especially older people and children, are particularly susceptible to air pollution. The air within buildings and other structures may also be polluted and have a negative effect on public health.

The 11th goal of the United Nations (UN) Sustainable Development Goals (SDG) [2] highlights the important role of having adequate, safe, and affordable living, and basic services as well, and air quality is one of the surest measures of living standard, especially in urban settings like Dhaka.
It is reported by the UN [17] that,

- Air pollution from different industries, power plants, waste plants, traffic, and others caused 4.2 millions of death count in 2019.

[1] https://www.who.int/health-topics/air-pollution

[2] https://sdgs.un.org/goals/goal11

- Since 2015, more than 117 countries are focusing towards measuring air quality in a total of exceeding six thousands of cities.
- According to the new guidelines created by WHO, 99% of the people who live in cities breathe polluted air in 2021.
- People from least developed countries (LDCs) and middle-income countries suffer more by air pollution, which results 4.2 million deaths which are premature.

It is also mentioned, according to their target 11.6, it is targeted that by 2030, they will reduce the adverse per capita environmental impact of places where communities live through giving intense focus toward air quality as well as other waste management.
Every now and then, it becomes common knowledge for residents of one of the world's most densely populated cities, Dhaka, Bangladesh, that the city has been classified as the worst in the world in terms of air quality. [43] Furthermore, Dhaka was declared the most polluted city in the 2021 Air Quality Report [42], a continuation of its ranking as the worst in 2020.

Climate change plagues our world now more than ever. Some of its effects include extreme typhoons, drought, a rise in sea level, an increase in global temperature, and tsunamis. As humans have contributed alone to this drastic change in climate conditions, it bears on us to slow down this phenomenon. Our continuous usage of non-recyclable materials and excessive use of plastics contribute heavily to the rapid increase of climate change. The 13th goal[3] of UN SDG concentrates on actions that must be done in order to counter the effects of climate change, or at least slow down its effect.

It is also mentioned, according to their target 13.3, it is targeted to ameliorate awareness-raising on human and institutional capacity regarding the climate change, which includes getting aware regarding the atmosphere.
The globe as a whole has been a witness to a distressing environmental phenomenon known as urbanization and industrialization over the course of the most recent several decades. Dhaka is not an exception to this pattern at all. These are contributing to and hastening the deterioration of air quality, which has now reached the status of a global catastrophe. Dhaka, Bangladesh, which is one of the most densely populated cities in the world, has been ranked as the city with the worst air quality in the entire world [43]. This fact is occasionally brought to the attention of residents of Dhaka, and it eventually becomes common knowledge among them. In addition to this, Dhaka is ranked as the world's most polluting city in the 2021 World Air Quality Report [42], which is an extension of its status as the most polluted city in the 2020 World Air Quality Report. Dhaka has been struggling with issues related to air pollution for quite some time now. Dhaka is consistently ranked in the top 10 of the most polluted cities in the world, even if it does not maintain its position at the top of this list. It is common for air quality to degrade over the winter months, although it tends to improve during the monsoon season.

[3] https://sdgs.un.org/goals/goal13

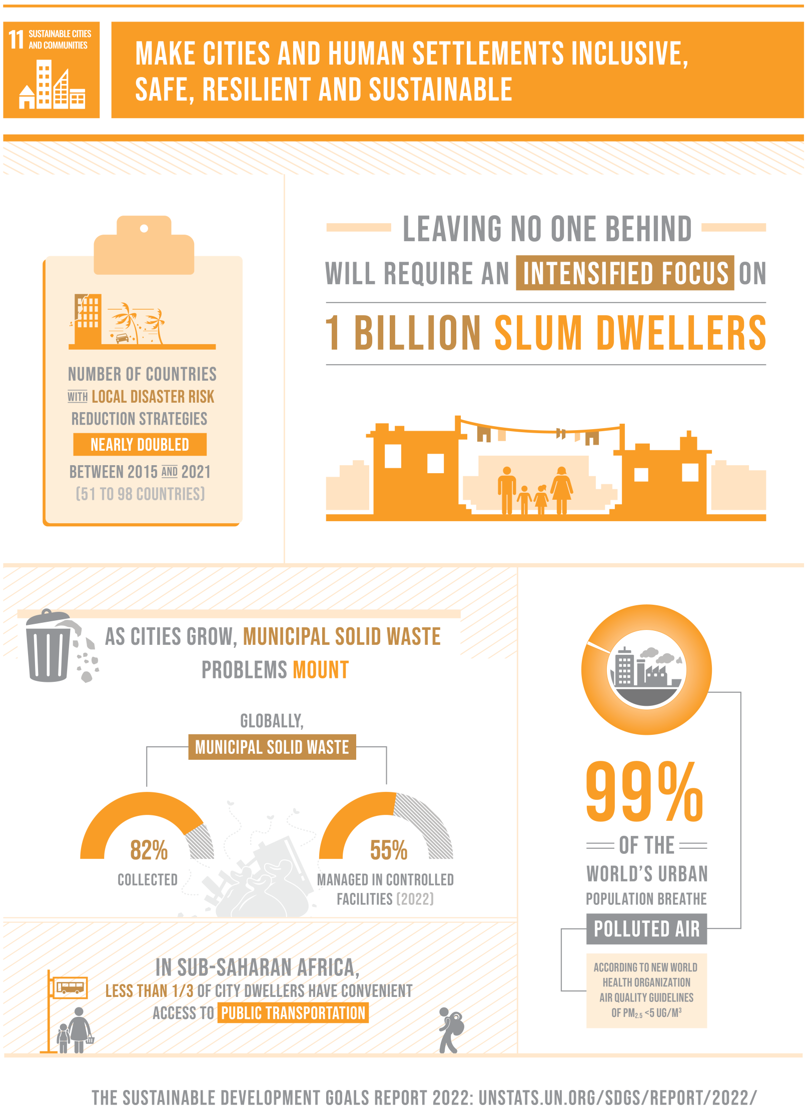
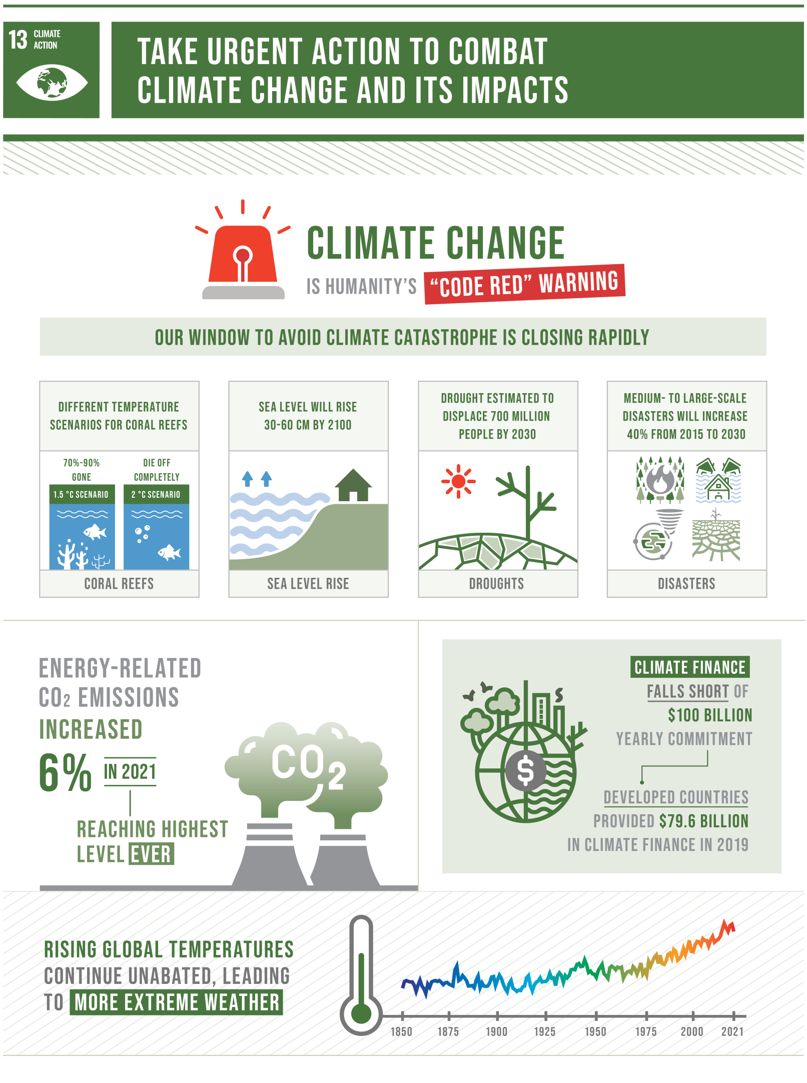

11 SUSTAINABLE CITIES AND COMMUNITIES
MAKE CITIES AND HUMAN SETTLEMENTS INCLUSIVE, SAFE, RESILIENT AND SUSTAINABLE
NUMBER OF COUNTRIES WITH LOCAL DISASTER RISK REDUCTION STRATEGIES
NEARLY DOUBLED
BETWEEN 2015 AND 2021
(51 TO 98 COUNTRIES)
LEAVING NO ONE BEHIND
WILL REQUIRE AN INTENSIFIED FOCUS ON
1 BILLION SLUM DWELLERS
AS CITIES GROW, MUNICIPAL SOLID WASTE PROBLEMS MOUNT
GLOBALLY,
MUNICIPAL SOLID WASTE
82%
COLLECTED
55%
MANAGED IN CONTROLLED FACILITIES (2022)
99%
OF THE WORLD'S URBAN POPULATION BREATHE
POLLUTED AIR
ACCORDING TO NEW WORLD HEALTH ORGANIZATION AIR QUALITY GUIDELINES OF PM2.5 <5 UG/M³
IN SUB-SAHARAN AFRICA,
LESS THAN 1/3 OF CITY DWELLERS HAVE CONVENIENT ACCESS TO PUBLIC TRANSPORTATION
THE SUSTAINABLE DEVELOPMENT GOALS REPORT 2022: UNSTATS.UN.ORG/SDGS/REPORT/2022/

Figure 1.1: SDG Goal 11 [46]

1.2 Research Objective

From the last decade onward, smartphones have become one of the most important products in our everyday lives, which accompanies us wherever and whenever needed, thanks to their powerful capabilities and a wide section of applications. In some ways, smartphones have evolved into a true human sixth sense. We can now readily capture images anywhere at any moment, analyze and film what is happening all around us, and then send and store them in private or public clouds over commonly accessible connections (such as cellular data or WiFi) for personal access or sharing. Considering how affordable it will be to supervise air quality in real-time from an given image if the PM 2.5 can be detected rapidly as well as accurately. Despite its significance, very few research works have been dedicated to this sort of study up to this point.

Based on our study, we make these major contributions to this paper:


13 CLIMATE ACTION
TAKE URGENT ACTION TO COMBAT CLIMATE CHANGE AND ITS IMPACTS
CLIMATE CHANGE
IS HUMANITY'S "CODE RED" WARNING
OUR WINDOW TO AVOID CLIMATE CATASTROPHE IS CLOSING RAPIDLY
DIFFERENT TEMPERATURE SCENARIOS FOR CORAL REEFS
70%-90% GONE
DIE OFF COMPLETELY
1.5 °C SCENARIO
2 °C SCENARIO
CORAL REEFS
SEA LEVEL WILL RISE 30-60 CM BY 2100
SEA LEVEL RISE
DROUGHT ESTIMATED TO DISPLACE 700 MILLION PEOPLE BY 2030
DROUGHTS
MEDIUM- TO LARGE-SCALE DISASTERS WILL INCREASE 40% FROM 2015 TO 2030
DISASTERS
ENERGY-RELATED CO2 EMISSIONS INCREASED
6% IN 2021
REACHING HIGHEST LEVEL EVER
CO2
CLIMATE FINANCE
FALLS SHORT OF $100 BILLION YEARLY COMMITMENT
DEVELOPED COUNTRIES PROVIDED $79.6 BILLION IN CLIMATE FINANCE IN 2019
RISING GLOBAL TEMPERATURES CONTINUE UNABATED, LEADING TO MORE EXTREME WEATHER
1850
1875
1900
1925
1950
1975
2000
2021

THE SUSTAINABLE DEVELOPMENT GOALS REPORT 2022: UNSTATS.UN.ORG/SDGS/REPORT/2022/

Figure 1.2: SDG Goal 13 [47]

- We estimate the PM2.5 concentration based on regression through custom Deep CNN based architecture leveraging the power of smartphone images.

- We create a dataset of images and related PM2.5 based on Bangladesh, a South-Asian country where PM2.5 Concentration is usually higher than average in the world.

It should be noted that our work computes aggregated AQI for the entire city as the average difference between location-specific AQI, taken by a sensor and the average AQI published by the local US Consulate is within 5-10% value of location-specific AQI as reported by Auvee et al. [27]. Also, it addresses the 11th goal of SDG and fulfills one of the crucial targets of its 13th goal, which is the improvement of education, awareness-raising, and human and institutional capacity building on climate change mitigation, adaptation, impact reduction, and early warning.
The following is the framework for the remainder of our thesis. We begin with a

Figure 1.3: Pedestrians cover their noses as dust shrouds a dilapidated part of Rampura-Banasree Road from Dhaka

review of some of the most recent research studies related to this problem in Chapter 2. Then, we introduce and describe the dataset we work on in Chapter 3. Afterward, we discuss the background of the models in Chapter 4. Then, we present the deep CNN based process picture-based prediction of the PM2.5 concentration (PPPC) approach as well as illustrate the workflow proposed in Chapter 5. Then, we execute a major performance assessment in Chapter 6.

Chapter 2

Background

2.0.1 Related Works

There is very limited research on Picture-based Prediction of PM2.5 worldwide. More specifically, there are works related to AQI classification based on images.

In recent years, we see growth in the development of technology and the use of Machine Learning (ML). Alongside that, academics are increasingly using Artificial Neural Networks (ANNs). Additionally, CNN is mentioned in a variety of paper-related contexts. The authors continue by breaking down the DL design thoroughly and begin to push forward toward efficient outputs for diverse issues including AQI categorization.

Liu et al. [18] show that ML can be effectively used to predict PM2.5 using images. They start with learning six attributes derived from images. DCP is applied to forecast image transmission first, then picture contrast and entropy are analyzed as characteristics. Furthermore, the impacts of the color of the sky and the position of the sun on the estimate of PM2.5 concentration are addressed and utilized for features. They collect the PM2.5 data from their local US Consulate which is published on an hourly basis. However, there is a drawback on the dataset they worked on, where it can be recognized as a skewed one. Chang et al. [10] use a SVR to incorporate all of the information, for the purpose of forecasting PM2.5.

Nonetheless, the model is highly dependent on previously designated reference zones with varying depths, which severely limits its application possibilities. Moreover, the work is pretty limited to predicting from images taken by static cameras. In addition, Gu et al. [28] begin by extracting entropy characteristics in the spatial and transform domains. In addition, they develop naturalness statistics models for the aforementioned two entropy characteristics by using a large number of images recorded during a particular period of favorable weather. We can compute a relative value that reflects the likelihood that a given image has a low PM2.5 concentration by analyzing the probability of naturalness between the entropy features and the statistical models. The probability measurement is then transferred to the estimation of PM2.5 concentration using a simple non-linear logistic function. They also collect their data from the local US Consulate of Beijing. Moreover, their dataset is also seen as a skewed one.

Auvee et al. [27] present an approach related to Air Monitoring System which primarily relies on the Geographic Information System (GIS) and the Air Pollutant Index (API) where they collect area-based AQI around Dhaka City. While working on this, they report that the average value of PM 2.5, collected by the local US Consulate is 193ppm whereas the PM 2.5 value of their collected data in different places of Dhaka City is around 178ppm which is a difference of 7% which lends credibility to our work.

The local US Consulate collects the AQI data based on Nowcast Algorithm. Nowcast is an algorithm developed by the Environmental Protection Agency (EPA), United States, to calculate AQI based on a weighted average of air particulate condensation from the recent-most hours in a particular location [1]. This system is given in g/m3 or ppb, in the AQI scale ranging from 0 to 500. It is employed for all recorded AQI values on the airnow.gov website. The idea behind the nowcast is to provide the "24-hour average" that should be applied when converting concentrations to AQI. This is because the AQI scale indicates that each of the Health Concern Levels (i.e., Good, Moderate,... Unhealthy...) is valid under a 24-hour exposure. For instance, 188 AQI (Unhealthy) should be interpreted as "if I remain outside for 24 hours and the AQI is 188 throughout those 24 hours, then the health consequence is Unhealthy." This is completely distinct from the statement "if the current AQI is 188, it will be unhealthy for health." [2]

With the advent of time, the usage of Artificial Neural Networks (ANNs) and their offshoots play an increasingly important role in both regression and classification tasks. Additionally, Convolutional Neural Networks (CNNs) have found profound applications in the modern research arena.

Zhang et al. [35] present a deep convolutional neural network model AQI classification. Their model, an AQC-Net based on ResNet, extracts feature information from landscape images captured by a camera and then categorizes them to estimate air quality levels. A self-supervision module (SCA) is appended to this model and the global context information of the feature map is then employed for feature reconstruction by leveraging the interdependence between the channel maps to enhance the interdependent channel maps and improve the ability of feature representation. Moreover, they compile an outdoor air quality data set to facilitate the training and evaluation of the model's performance. As an extension of their previous work, [44] present a real-time and image-based deep learning model named YOLO-AQI which is a customized model based on YOLO.

Chakma et al. [21] present a fine-tuned deep convolutional neural network model, which also can categorize natural images into different classes based on their PM2.5 concentrations. They also present a dataset of 591 images, whrere PM2.5 data is collected from local US Consulate, in Beijing, China. They mention their model as "imagenet-matconvnet-verydeep". A drawback in their work is that they only classify 3 classes which are Good, Moderate, and Severe.

[1] https://cutt.ly/NXphsOi [Retrieved from EPA]

[2] https://cutt.ly/oXhiImf

Li et al. [16] come up with an approach where they classify air pollution from images based on haze. They also present a dataset of 8,761 images, whrere PM2.5 data is collected from local US Consulate, in Beijing, China. In their approach, they initially estimate the transmission matrix using the Dark Channel Prior (DCP) [8] method. In parallel, they estimate the depth map based on Deep Convolutional Neural Fields (DCNF) [19]. By combining the transmission matrix and the depth map, they approximate the haze level of the photo.

Rijal et al. [25] propose another algorithm to solve this problem. The solution is mainly an ensemble of deep neural networks-based regressors, which utilizes a feed-forward neural network to merge PM2.5 predictions yielded by three convolutional neural networks, VGG-16, Inception-v3, and ResNet50. They also present a dataset of 1,460 images, whrere PM2.5 data is collected from local US Consulate, in Beijing, China. They thereby arrive at the final PM2.5 prediction of the image. We also notice that this work has a really skewed dataset which may have contributed to creating biased output.

Sarkar et al. [40] investigate and highlight how the Covid 19 lockout affected PM2.5 pollution in the city utilizing "Ground based observation data." Basically, they look into the air quality in Dhaka from January 1, 2017, to August 1, 2017. Second, the research begins in January and continues through Pre Covid (March 23), Covid (March 24 to May 30), and Post Covid (May 31 to August 1). They demonstrate that the annual formula for PM2.5 pollution was the same from 2017 to 2020. While doing the study, a new concentration of -87.47 g/m^3 was developed. Here, when compared to earlier data, in this case, historical data showed that the significant drops in PM2.5 during the Covid shutdown were detected at a rate that was 11.31% lower than the absolute drop of 7.15 g/m^3. Air is such a crucial component of the environment; this study also alludes to health outcomes. But for that, more research ought to have been conducted. Since only three years' worth of data were used in this analysis, more catalysts may have an impact on the PM2.5 level.

Using the traffic volume list and Ecotec mini 2.5 sampler, this study by Hossain et al. [29] investigates the relationship between PM 2.5 concentration and mode of transportation in the city of Dhaka. Due to the weather, Dhaka is suffering from a significant amount of PM 2.5. Since the previous years, beginning in 2013, the concentration in Dhaka has increased six times higher than Bangladesh's NAAQS. According to the findings of this study, one of the major reasons for air pollution is increasing the number of automobiles in cities. Road categories that are mentioned here; Vehicle free roads, mixed-use roads, motorized vehicle-dominated roads, and predominant roads by non-motorized roads. Now, to organize this study 12 different locations have been identified for data collection and as the result of 12-hour concentration the air quality ratio scale of: Moderate>caution>unhealthy. Moreover, the research shows the PM2.5 concentrations in compared to mixed, motorized areas are on average higher than others with that, the maximum concentration was found in Mirpur-10 with 172.2 ug/m^3 whereas the minimum was in B.C Das street 40 ug/m^3.

Additionally, the study demonstrates that the average PM2.5 concentrations in motorized, mixed-use regions are greater than those in other locations. The highest concentration of ss was discovered in Mirpur-10 at 172.2 g/m^3, in contrast to other areas. This analysis essentially shows that Dhaka city is going to get worse. Due to the fact that, after categorizing the calculation, it shows that the concentration on motorized and mixed category roads is higher and is nearly 4 times higher than the WHO and Bangladesh national norm, with an AQI of 100% becoming unhealthy and others being labeled as caution and moderate. So now is the perfect moment to prepare for the issue.

Ma et al. [24] present a hybrid convolutional neural network approach that is used to estimate air pollution based on images. Moving forward, this method estimate air pollution based on a single smartphone photo. Here, they use both an RGB image and dark channel maps to simulate a hybrid Convolutional neural network. Basically, in order to move forward with the work, they take the collected image into consideration as their primary input. By doing so, they are able to address the negative impacts of increasing depth and enhancing the network's performance. This effort is for the primary network; going on to the secondary network, a dark channel map will be generated for the purpose of improving features with representation. This method has been compared to other techniques both qualitatively and numerically. In addition to image collection, data sets have been employed to conduct experiments, and the best results have been obtained. Additionally, 1575 photos with various PM 2.5 values were gathered in this location in order to build their network in fusion mode. In order to increase the accuracy of the classification, they introduce the relationship between hazy images and the air quality factor of various locations.

Bo et al. [23] present a technique for estimating PM2.5 from images that makes use of the CNN Model together with two weather features. SVR and Deep Learning algorithms are introduced in the study to blend picture and data to predict PM2.5 levels in the air. A ResNet model is utilized to predict PM2.5 levels for images, and humidity and wind speed are employed as weather variables. In order to assess PM2.5, the SVR model has now been coupled with Res-Net output. In contrast to other architectures, the CNN explicit end to end architecture enables automatic extraction of both low-level and high-level image features. This study's methodology demonstrates that extraction takes less time, and the stages of feature optimization don't result in any defects either. The experimental results demonstrate that the strategy stated performs better and that weather factors can contribute to improving PM2.5 estimation accuracy.

Samsami et al. [30] present a computer vision-based method for classifying the air quality level that is based on image processing and uses image analytics. Any smartphone with a camera may be used to accurately examine by taking a picture of the sky and collecting useful information from it. ImSkyset, which is essentially a dataset with 3422 pictures, is used for this paper's purposes. According to the article, environmental pollutants lead to deteriorating air quality, which has an adverse effect on the color of the sky as well. The three operators LBP, DCO, and Gabor filter are discussed in more detail here since they are able to extract images from photographs even when the variations in air quality are not very noticeable. Moreover, the KNN produced 7 sub-features. Additionally, the use of categorization resulted in the production of 7 sub features for the KNN. The verification and accuracy levels were assessed after completing all of them. And this result shows a performance accuracy rate of about 82.02%. The results thus support the benefit of utilizing image processing methods to measure air quality for the purposes of raising public awareness.

In order to precisely estimate the PM2.5 and PM10 values of air pollution, Song et al. [34] focus on the usage of a smartphone. As a result, they emphasize estimation on a single daylight image in their suggested deep learning approach, which integrates ResNet with LSTM. The author's reference, in this case, was the equipment used at official air quality monitoring stations. Moreover, this test result does not change the value of PM2.5 and PM10 within a 500-meter radius. Additionally, 3024 images of the same structure that were shot both during the day and at night have been saved. The ResNet-LSTM model construction is then suggested by the researchers. After that, there is a discussion of Met-ResNet-LSTM, which is based on ResNet-LSTM and has made an effort to consider the impact of weather on air pollution. They also created ResNet-LSTM-SP and Met-ResNet-LSTM-SP, each of which has a single short path. The image-based estimating method, as well as ResNet-LSTM and Met-capacity ResNet LSTMs, perform better when detecting any rapid changes in pollutant concentrations due to the short path. The results of all trials and verification show that the proposed techniques and the three sections perform exceptionally well with deep learning, ResNet, and all other convolution. Additionally, the suggested method has less error than convolution.

Wang et al. [32] present that in order to learn the combined feature derived from various areas of environmental photos, the research suggests the Double Channel Weighted Convolution Network (DCWCN) crops learning technique. The research mostly discusses learning about DCWCN and Self-learning weighted features. With this, they create a dataset of environmental images with time and position information to continue the technique. Thus, there are four main contributions to the paper's discussion. Here, an accuracy level of about 87% is found, and the method indicated shows that it is highly successful and performs admirably in terms of both mean absolute error and recognition accuracy, as was previously mentioned. Overall, the performance comes across well.

Saha et al. [39] work on the impact of the pandemic and the monitoring of Dhaka's AQI on the pandemic period in the study. The relationship between the Covid 19 Fatality Rate (CFR) and AQI from Bangladesh and India is discussed in this passage. According to the report, this is the first time the AQI of Dhaka has been monitored using the Internet of Things, or IoT.
They essentially track the AQI in Dhaka City using various gases, such as CO and NO_2, and to work with an Arduino-based Node-MCU and other sensors. In addition to the WSN approach that is developed for AQI assessment both indoors and outdoors, other processes were also listed. The ESP-32 WiFi module will upload the data set to the server in this case so that it may be accessed from anywhere. They contrast CO and NO_2 emissions during a pandemic in this passage. Additionally, the researchers develop a novel formula to lessen the bad effects of the pandemic. Internet of Things (IoT) is essentially a network that appears to be a sensor that is

connected. They also offer a framework for monitoring that depends on temperature, humidity, information systems with high light levels, dispersed sensors, and more. With all that, a dataset is constructed by keeping an eye out for carbon monoxide isolated regions; this strategy encompasses a 100 sq. km. urban area.

2.0.2 How PM2.5 Affects Optical Image

In the air, PM influences an image in various approaches[18]. However, they are emerged from light interactions alongwith the particles of air, primarily through light scattering. It includes Rayleigh scattering and Mie scattering [1]. Light scattering lowers light transmission in air, which is described by the Beer-Lambert equation.

$$t = e^{-\beta d} \tag{2.1}$$

Here, β is the medium extinction coefficient, which varies with particle size and concentration, and d is the distance of light propagation. This equation illustrates that PM concentration can be calculated through determining the extinction coefficients at various wavelengths. The extinction coefficient can be calculated from an obtained picture using the equation 2.2, which is mentioned in many researches [3], [6], [7], [12].

The preceding explanation does not explicitly account for color information, which may potentially serve as crucial characteristics for PM estimates based on light scattering considerations. When particles (especially air molecules) are much smaller than the wavelength of light, Rayleigh scattering reigns supreme. It fluctuates with wavelength (λ) in accordance with $\lambda^{-}4$, which is responsible for the blue hue of the sky. Mie scattering, on the other hand, happens when the size of the particles is equivalent to the wavelengths of light, resulting in a white glare surrounding the sun when particles are present in the air. The combo of Rayleigh and Mie scattering affects the image's luminance and color vibrancy. In contrast, the color and brightness information provides information on particle concentration and size and may be used to determine PM. In addition to light attenuation, the current approach provides color information as a crucial feature point for PM assessment.

2.0.3 How Image Features can Represent AQI?

We predict the AQI from user photos using the following features, following the approach of Liu et al. [18].

1. Transmission
2. Blue Colour of the Sky
3. Gradient of Sky Region
4. Image Contrast
5. Entropy
6. Humidity

Traditional image processing methods are used to extract these characteristics, which are then integrated with a linear model.

Transmission

This identifies image deformation and the amount of light that enters a camera after being impacted by particles in the air, which can be characterized by Equation 2.2 [3], [6], [7], [12].

$$I(x, y) = t(x, y)J(x, y) + (1 - t(x, y))A \tag{2.2}$$

Here, the input hazy image is characterised by I. In addition, t is the scene-to-camera transmission, J is the scene radiance, and A is the airlight color vector. As seen in Fig. 2.1, the first component of Equation 2.2 is the instantaneous transmission of scene radiance into the lens, which consists of reflected light by scene object surfaces and filtered by air before entering the camera. The second component $(1 - t(x, y))A$ is referred to as airlight, which is the reflected light dispersed into the camera by air molecules and PM [6], [7], [12]. Wang et al. [15] estimate attenuating of light using the abovementioned calculation equation. By examining ROIs at various distances, the link between transmission value and PM density was studied in this study [18]. Equation 2.2 assumes a given fixed atmospheric and lighting conditions, in actuality which can change depending on the time of day, location, and season of the sun. In addition to the climate and the location of the Sun, both J and A rely on the distribution and concentration of PM. Using the principle of dark channel, which implies the presence of some pixels with zero or extremely low intensity for at least one-color channel in all outdoor photographs, the transmission for a particular hazy image is determined. Using Equation 2.3, the dark channel is computed for a haze-free image J.

$$J^{\text{dark}}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} (J^c(y)) \right) \tag{2.3}$$

Here, J^c is one of the color channels of J, and the localized patch which is positioned at x is characterized by (x). The sky or the brightest area can be utilized to calculate the airlight, therefore the transmission can be obtained by:

$$\sim t(x) = 1 - \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} \frac{I^c(y)}{A} \right) \tag{2.4}$$

Here, $I^c(y)/A$ is mainly the hazy image, which is normalized by the air-light A, and the second term on the right represents the hazy image's dark channel.

According to Liu et al. [18], a crucial assumption of the current model is that the transmission decreases rapidly with increasing scene object distance from the camera.

Blue color of the Sky

This characteristic is comparable to how we experience a polluted day in real life. When the sky is gloomy, we consider the day to be polluted. The color blue was calculated via RGB reduction.

The sky is blue on a clear day and gray or white on foggy or overcast days due to light dispersion. The hue of the sky is determined by the average value of the blue RGB channel component in the sky area.

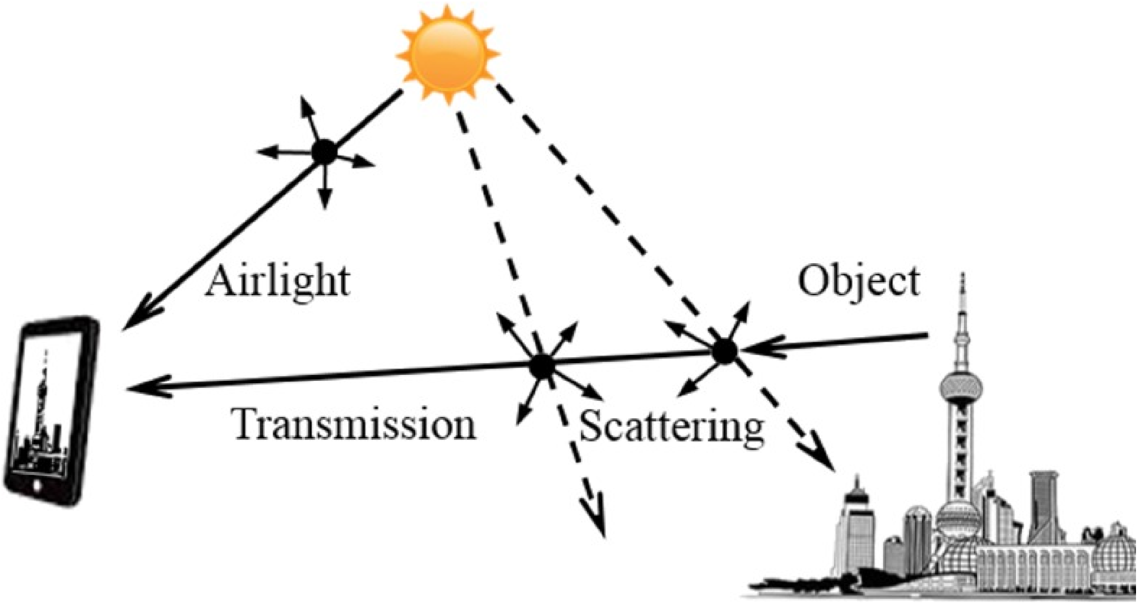


Figure 2.1: Radiance entering the smartphone camera is the total of light transmitted from the subject and airlight from the sun after spreading by air, moisture, and atmospheric particles [18]

Gradient of Sky

Due to cloud cover, the sky can look grayer in color and this characteristic is included to account for the chance to add the feature. Now, making a mask of the sky region and then calculating the region's Laplacian helps to determine the gradient. Here, the average gradient amplitude in the sky region determines how smooth the sky is. The gradient would be high if the pollution level were high.

Entropy, RMS Contrast

To get PM concentration in air, another feature is Image Contrast. As a matter of fact, picture contrast or visibility is related to how well a human can see the air quality [2], [9]. Based on Equation 2.2 the effect of PM on image contrast can be determined. As PM concentration increases, the airlight term (second term of Equation 2.2) arises from light scattering by PM increasing rate. The Air light lacks scene information, which is why PM causes the contrast in the image to drop. Due to transmission decrease with the distance between an object and camera, the airlight term contribution rises. Also, the distance got increased here. So, in short, the more PM there is in the air, the less contrast there will be in the image.
These characteristics also provide information about the specifics of an image Now, if the day is a polluted day then the details of the image will be missing. The contrast of an image can be measured in a variety of ways. Using the RMS of a picture to explain image contrast is the most used one. This method has been found to match with human perception of image contrast [5]. The standard deviation of the image pixel intensities is used to define RMS contrast, by the following Equation:

$$RMS = \sqrt{\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathrm{Iij} - \mathrm{avg}(I))^2} \tag{2.5}$$

Here, I_{ij} is the intensity at (i,j) pixel of the image with size $M \times N$, and $avg(I)$ is the average intensity of all pixels in the image. So, we can say that contrast contains an inverse relation with PM 2.5. To estimate entropy, we use the following Equation:

$$\text{Entropy } = -\sum_{i=1}^{M} pi \log 2pi \tag{2.6}$$

Here, pi is the probability where the pixel intensity is equal to i, and M is the maximum intensity of the image. We can notice that, the image increasingly loses its details whenever the PM concentration increases, and the image entropy decreases. So we can conclude by saying, it has an inverse relationship to PM 2.5.

Humidity

Humidity is a component of meteorology. Study [48] shows that pollution levels increase on humid days because PM 2.5 absorbs moisture and reduces visibility.

Since the majority of these characteristics have a linear relationship with PM 2.5, they are input into a neural network with a linear activation function, as outlined in Section 5.

Chapter 3

Dataset

We create a custom dataset consisting of images and PM2.5 readings from about that period that are collected inside Dhaka city, which is the capital of Bangladesh. The images are from some point in time between the years 2020 and 2021. The PM2.5 Concentration of that exact time or approximate time range near the time the picture is taken is included in the dataset, which primarily consists of images that are taken by people, the date and time that the picture is taken, the location it is taken (within Dhaka City), and the date and time that the picture is taken. The information on PM2.5 is obtained from the United States Consulate, which is the location where the AQI is published and updated on an hourly basis[1].

Due to the fact that this is an academic study exclusively, we have sent a Google Form to the residents of Dhaka City, requesting them to submit only photographs that they have taken themselves and in which at least half of the sky is visible, along with their permission to do so. In addition to that, we gather data from our end utilizing mobile phones that we have purchased ourselves.

After we have finished collecting the data, we then gather the PM2.5 levels at that moment using a predetermined method. We separate the time range for photographs that are taken in between a specified hour since the data that comes from the US embassy is provided on an hourly basis. For example, we have PM2.5 data from both 9 and 10 in the morning. Now, on that basis,

- If it is taken within 9:30 AM, we consider the PM2.5 of 9 AM for that picture.
- If it is taken after 9:30 AM and within 10 AM, we consider the PM2.5 of 10 AM for that picture.

Primarily, the dataset is comprised of 1,818 images. The images include a variety of time periods, geographic locations, and PM2.5 concentrations.

The dataset includes a much greater variety of photographs, each of which is distinguished from the others based on the PM2.5 concentration. There are six distinct PM2.5 criteria that are used to determine how severe the air quality index (AQI) really is. In order to demonstrate the accuracy of our findings, we generate PM2.5 level estimates and plot them on an Air Quality Index (AQI) scale that has a color gradient. This is the official standard scale that is used by the governments of all of the different countries, and it has been in use for a considerable amount of time. Each level has its own own distinct range from 0 to 50, and it always begins at 0.

[1] https://www.airnow.gov/international/us-embassies-and-consulates/

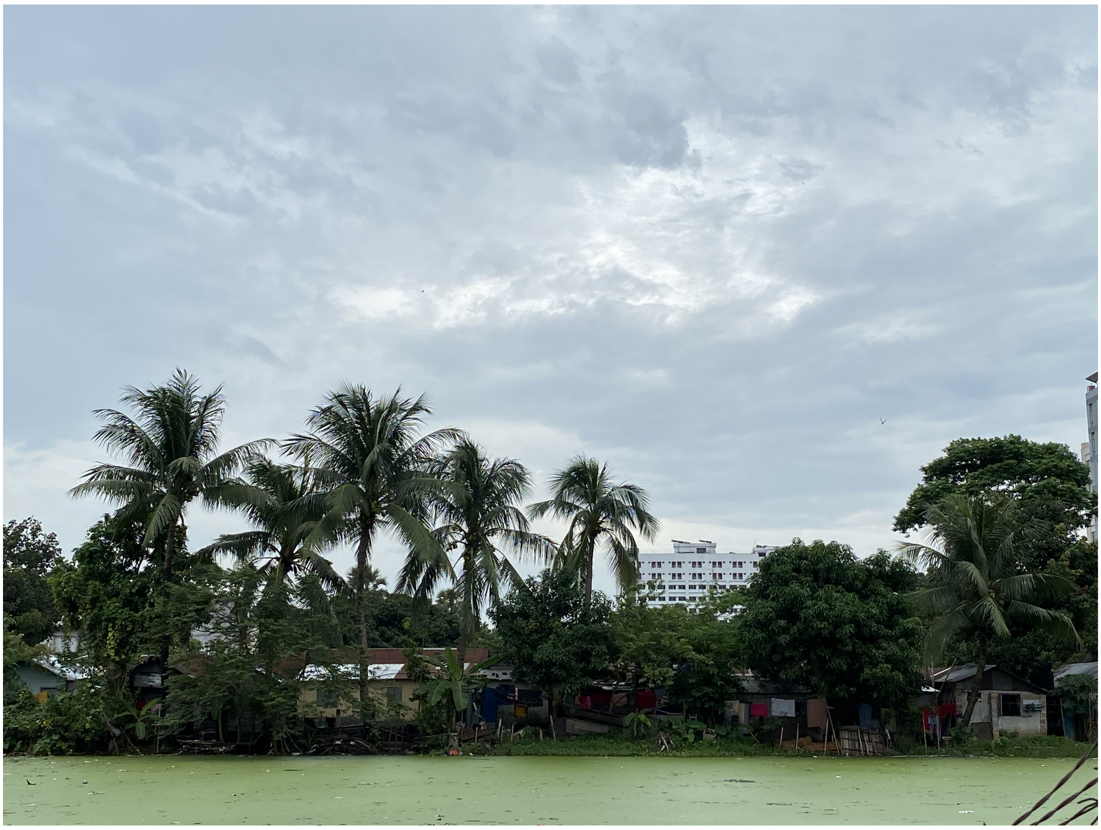


Figure 3.1: Sample Image from Dataset (AQI = 63)

Air Quality Index (AQI) Range	Category	Color
0-50	Good	Green
51-100	Moderate	Yellow Green
101-150	Caution	Yellow
151-200	Unhealthy	Orange
201-300	Very Unhealthy	Red
301-500	Extremely Unhealthy	Purple

Table 3.1: Air Quality Index Chart [11]

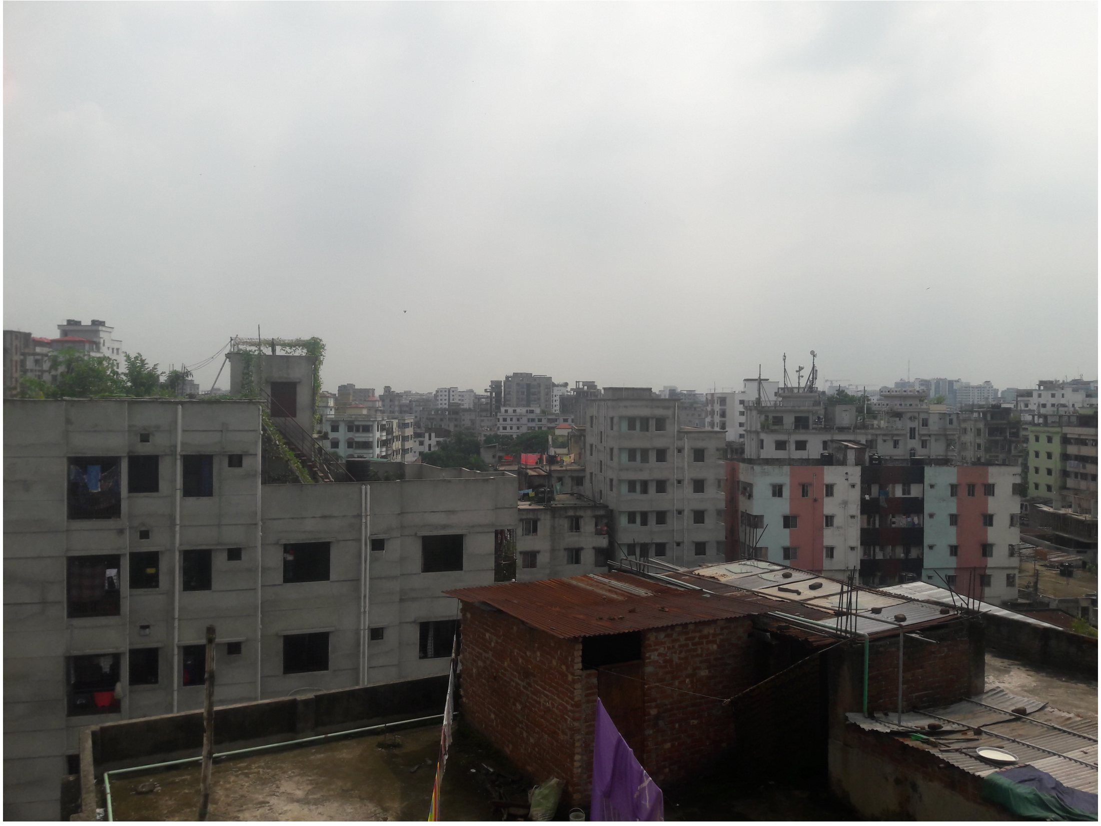


Figure 3.2: Sample Image from Dataset (AQI = 125)

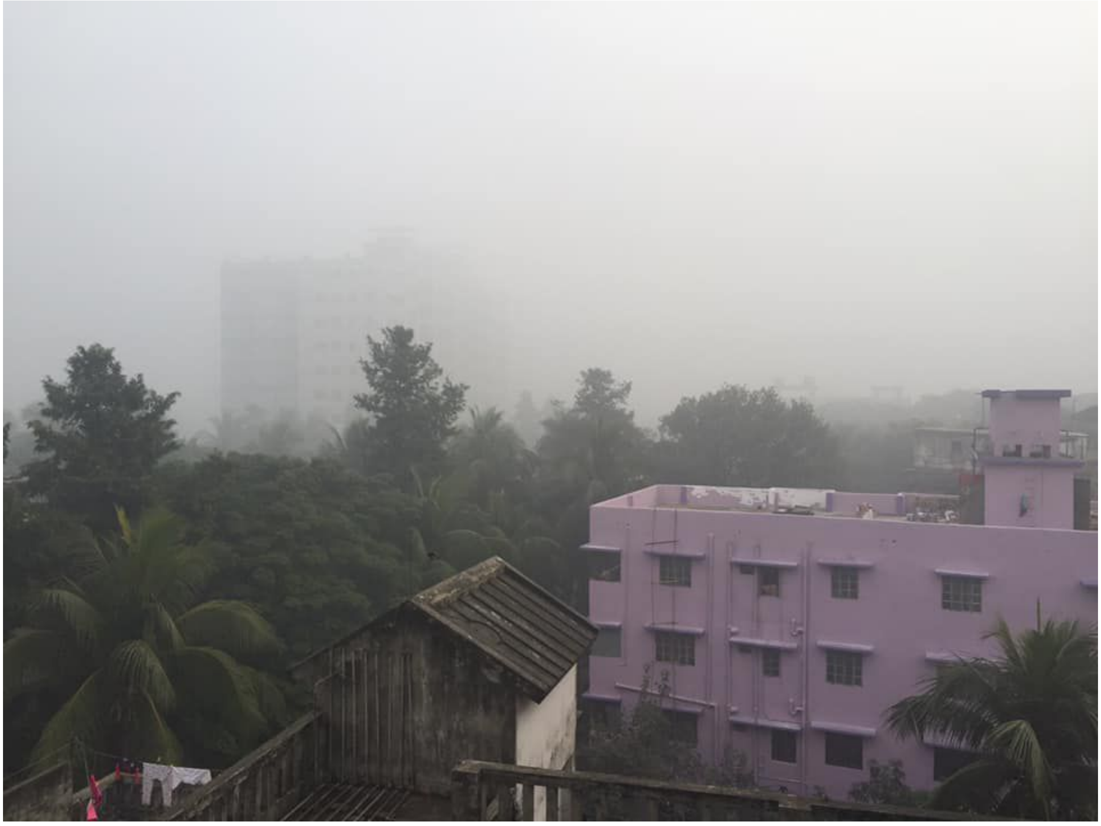


Figure 3.3: Sample Image from Dataset (AQI = 293)

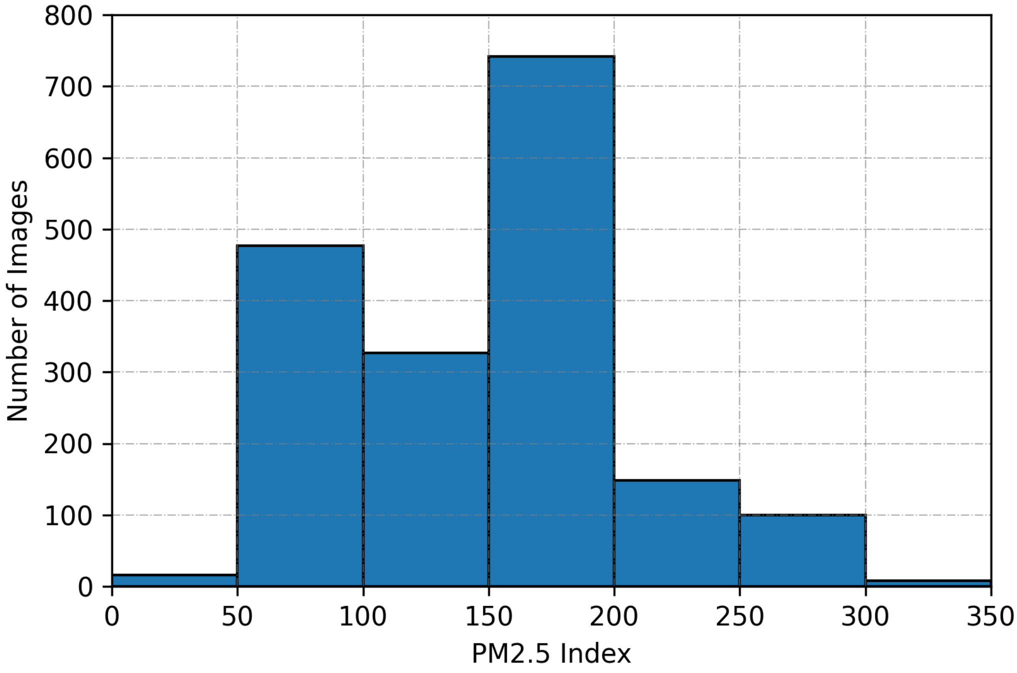
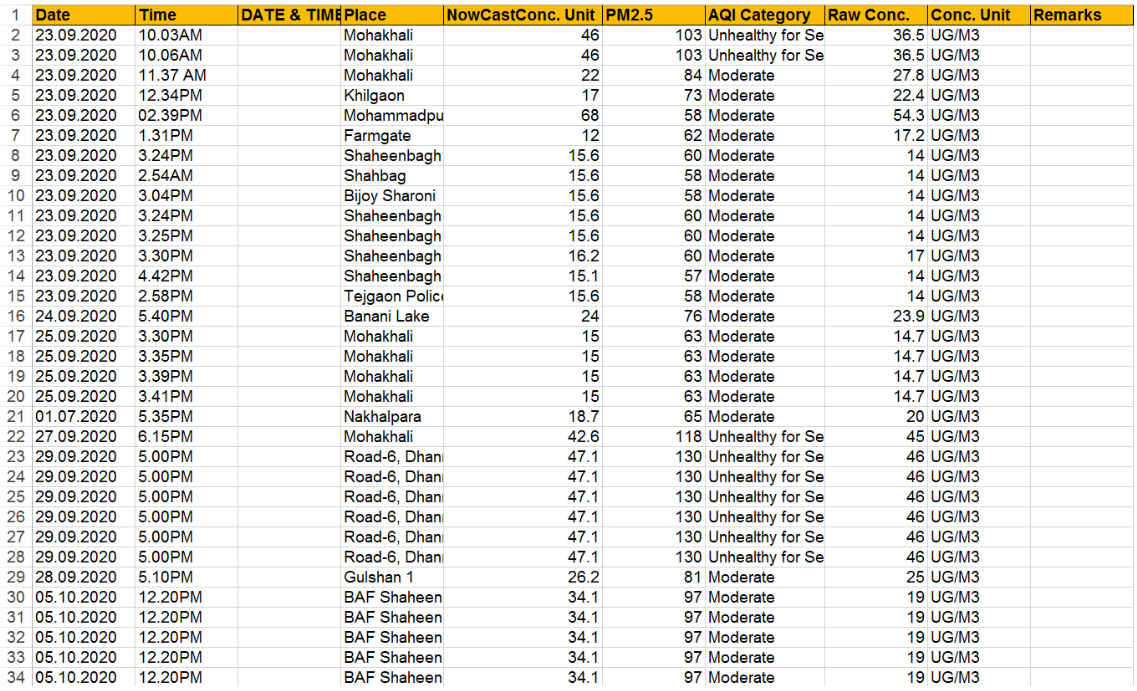


Figure 3.4: Histogram of PM2.5 of Our Dataset



1	Date	Time	DATE & TIME	Place	NowCastConc. Unit	PM2.5	AQI Category	Raw Conc.	Conc. Unit	Remarks
2	23.09.2020	10.03AM		Mohakhali	46	103	Unhealthy for Se	36.5	UG/M3	
3	23.09.2020	10.06AM		Mohakhali	46	103	Unhealthy for Se	36.5	UG/M3	
4	23.09.2020	11.37 AM		Mohakhali	22	84	Moderate	27.8	UG/M3	
5	23.09.2020	12.34PM		Khilgaon	17	73	Moderate	22.4	UG/M3	
6	23.09.2020	02.39PM		Mohammadpu	68	58	Moderate	54.3	UG/M3	
7	23.09.2020	1.31PM		Farmgate	12	62	Moderate	17.2	UG/M3	
8	23.09.2020	3.24PM		Shaheenbagh	15.6	60	Moderate	14	UG/M3	
9	23.09.2020	2.54AM		Shahbag	15.6	58	Moderate	14	UG/M3	
10	23.09.2020	3.04PM		Bijoy Sharoni	15.6	58	Moderate	14	UG/M3	
11	23.09.2020	3.24PM		Shaheenbagh	15.6	60	Moderate	14	UG/M3	
12	23.09.2020	3.25PM		Shaheenbagh	15.6	60	Moderate	14	UG/M3	
13	23.09.2020	3.30PM		Shaheenbagh	16.2	60	Moderate	17	UG/M3	
14	23.09.2020	4.42PM		Shaheenbagh	15.1	57	Moderate	14	UG/M3	
15	23.09.2020	2.58PM		Tejgaon Polic	15.6	58	Moderate	14	UG/M3	
16	24.09.2020	5.40PM		Banani Lake	24	76	Moderate	23.9	UG/M3	
17	25.09.2020	3.30PM		Mohakhali	15	63	Moderate	14.7	UG/M3	
18	25.09.2020	3.35PM		Mohakhali	15	63	Moderate	14.7	UG/M3	
19	25.09.2020	3.39PM		Mohakhali	15	63	Moderate	14.7	UG/M3	
20	25.09.2020	3.41PM		Mohakhali	15	63	Moderate	14.7	UG/M3	
21	01.07.2020	5.35PM		Nakhalpara	18.7	65	Moderate	20	UG/M3	
22	27.09.2020	6.15PM		Mohakhali	42.6	118	Unhealthy for Se	45	UG/M3	
23	29.09.2020	5.00PM		Road-6, Dhan	47.1	130	Unhealthy for Se	46	UG/M3	
24	29.09.2020	5.00PM		Road-6, Dhan	47.1	130	Unhealthy for Se	46	UG/M3	
25	29.09.2020	5.00PM		Road-6, Dhan	47.1	130	Unhealthy for Se	46	UG/M3	
26	29.09.2020	5.00PM		Road-6, Dhan	47.1	130	Unhealthy for Se	46	UG/M3	
27	29.09.2020	5.00PM		Road-6, Dhan	47.1	130	Unhealthy for Se	46	UG/M3	
28	29.09.2020	5.00PM		Road-6, Dhan	47.1	130	Unhealthy for Se	46	UG/M3	
29	28.09.2020	5.10PM		Gulshan 1	26.2	81	Moderate	25	UG/M3	
30	05.10.2020	12.20PM		BAF Shaheen	34.1	97	Moderate	19	UG/M3	
31	05.10.2020	12.20PM		BAF Shaheen	34.1	97	Moderate	19	UG/M3	
32	05.10.2020	12.20PM		BAF Shaheen	34.1	97	Moderate	19	UG/M3	
33	05.10.2020	12.20PM		BAF Shaheen	34.1	97	Moderate	19	UG/M3	
34	05.10.2020	12.20PM		BAF Shaheen	34.1	97	Moderate	19	UG/M3	

Figure 3.5: A snippet of Our Dataset

Chapter 4

Model Background

4.1 Convolutional Neural Network (CNN)

In order to accomplish our job, we will be analyze the pixels out of the pictures using convolution neural networks. CNN is used for image identification, regression prediction and classification because to the high level of accuracy it has in various applied areas. It is created to have compactness, efficiency, and a greatly reduced network architecture for the partitioning of fine structures in volumetric pictures. As a direct consequence of this, fundamental and adaptable components of existing convolutional networks, such as enhanced convolution and surviving association, need to be taken into consideration. In the majority of modern network topologies, a completely convolutional technique is used.

In the field of deep learning, a convolutional neural network is a kind of artificial neural network that analyzes data in the form of graphical representations. CNN has established itself as the standard to follow when tackling any problem involving images. They are far and wide above the competition in terms of accuracy. A number of other applications, such as natural language processing and recommender systems, make use of it as well. The most significant benefit that CNN has over its earlier neural networks is that it can recognize essential qualities even without the involvement of a human researcher. If a large number of pictures of different commodities and animals are shown to it, it may figure out on its own the distinctive characteristics of each category.

One of the most challenging features of using CNNs is the fact that, in contrast to standard image applications, poor data preparation capabilities can result the model perform worse. This is one of the most difficult components of employing CNNs.

Additionally, CNN models are now capable of running on any device, making it possible for them to reach a larger audience. Pooling techniques, sophisticated convolution, and parameter sharing are all used in a CNN's architectural process.

The use of a tensor as an input by CNN enables the network to better grasp the spatial connections (the link between nearby pixels in an image) between pixels and to perform better on more complex pictures.

4.2 CNN Architecture

A typical CNN can consist of these three layers:
Initially, it will contain a convolutional layer. Then there will be a layer which will pool the data in a certain manner. And finally, there will be a fully connected layer.

When the input image is supplied, CNN executes a sequence of convolution and then pooling operations followed by a number of fully connected layers.

4.3 Convolutional Layer

Convolutional Layer basically implements a mathematical operation on input data using its filter. It is considered the most crucial part of the whole procedure. Firstly, the image is taken as input in the convolutional layer. Then it performs a mathematical operation using its 3x3 kernel.

Generally, the convolutional operation takes place in the receptive field. As we have used a 3x3 kernel, our receptive field was a 3x3 matrix as well. The kernel is slid over the input to achieve a convolution and at each position, the sum of the result of component-wise matrix multiplication is plotted on the feature map.

4.4 Activation Function

It is basically a type of function which helps to determine whether a neuron/node should be activated or not, under some specific mathematical circumstances. It usually decides whether the input from a previous node is important or not and whether it should be added to the network based on some mathematical operations.

4.4.1 ReLU Activation

Non-linearity must be present in any neural network in order for it to be effective.
The ReLU activation function receives the result of the convolution operation. So instead of the sums, the final feature maps are the ReLU functions applied.
To achieve non-linearity, ReLU activation is applied to the feature map. ReLU activation takes the feature map and replaces all negative values with 0; if the value is greater than 0, it remains unchanged, according to the equation below and represented in Figure 4.1.

$$f(x) = \max(0, x) \tag{4.1}$$

4.4.2 Linear Activation

This function is often called as identify function which is proportionate to the input. It just returns the value which is inputted, according to the equation below and represented in Figure 4.2. For a regression based problem, linear activation function is used in the final output layer.

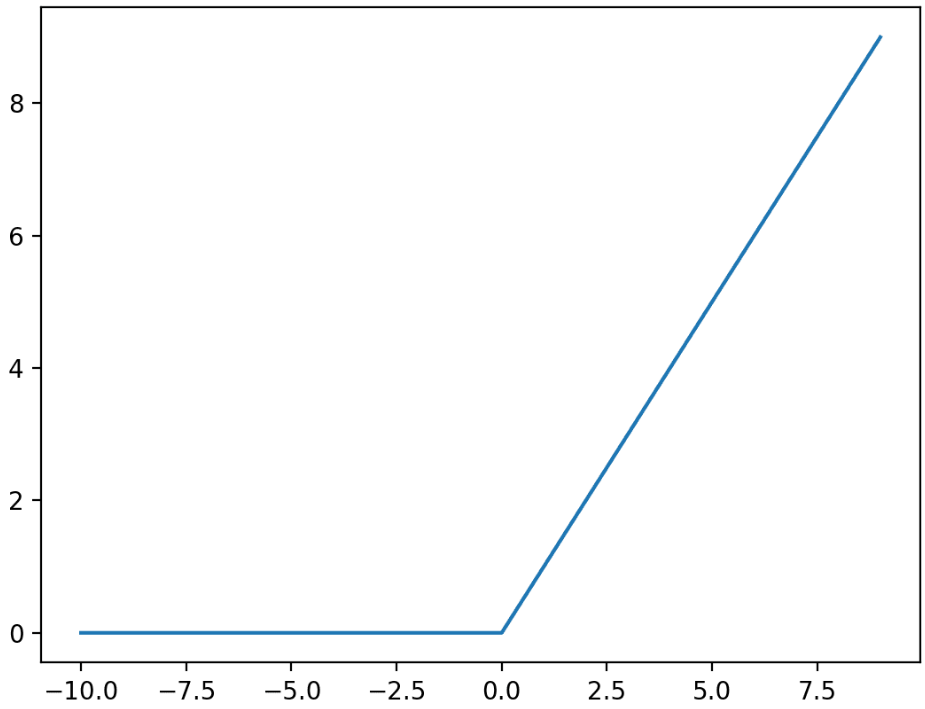


Figure 4.1: ReLU Activation Function Plot

$$f(x) = x \tag{4.2}$$

4.5 Pooling

In the pooling layer, the dimension of the feature map is reduced. Pooling is done on the feature map. In pooling at first, a spatial neighborhood which is a 2 pixel by 2-pixel window is taken. We can take any size of the window for pooling.
Here we have taken 2 × 2 windows for pooling. We take the largest value from the pooling window of the feature map and ignore the rest of the 3 elements. Through this process, we are preserving the important features. In pooling, the redundant pixels are canceled out and the important pixels are preserved. So through pooling, we get to reduce the size of the feature map which makes our program more efficient.

4.5.1 MaxPooling Layer

Using this layer, we can get the maximum values out of a specific region of the feature map of a given data. After taking the data, the output will contain the most significant characteristics of the previous feature map. It is shown to be significantly better to perform than average pooling for different computer vision tasks.

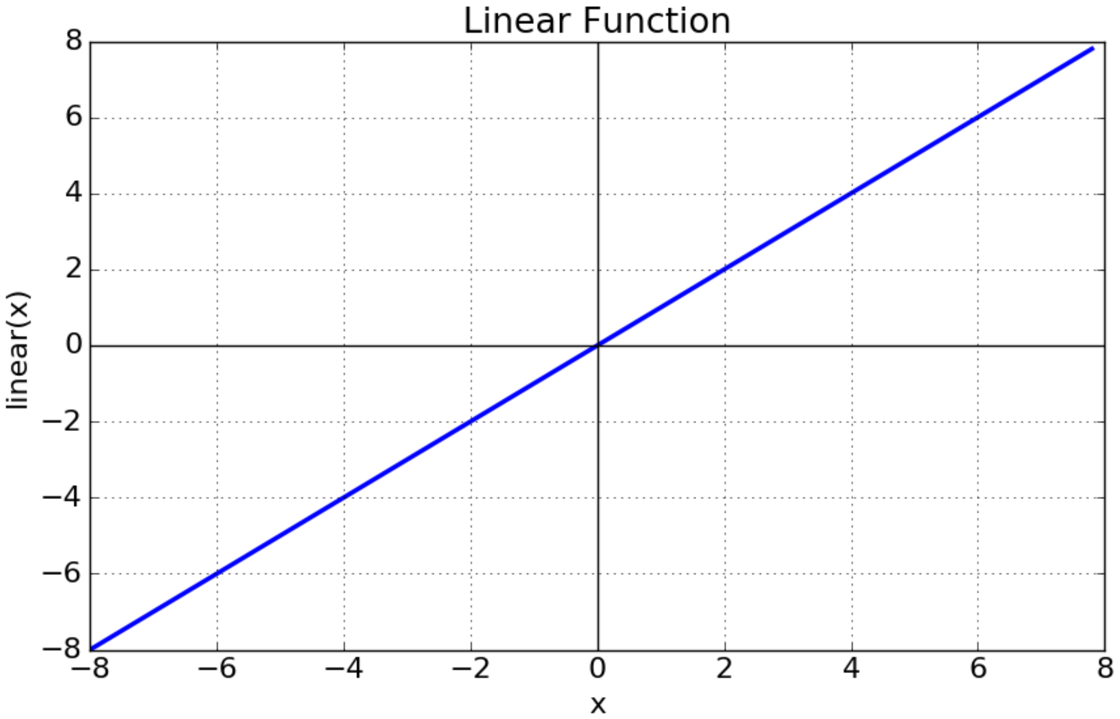


Figure 4.2: Linear Activation Function Plot

4.6 Hyperparameters

There are three crucial hyperparameters to choose from:

4.6.1 Filter

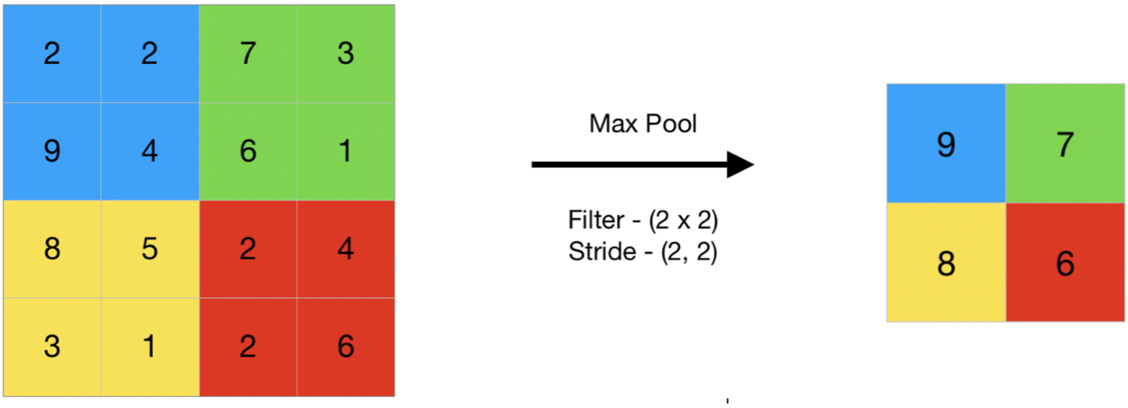
Filter is a window that scans the image. A filter is used to judge which feature do a pixel belong to. Does it belong to an arc, straight line, diagonal, etc.? One filter could be responsible for detecting arcs within an image, another could be responsible for diagonal lines, etc. Filters, also known as kernels, scan the image region after another. The region size is determined by the window or kernel size.

Filter size:

We commonly utilize 3×3 filters, however depending on the purpose, 5×5 or 7×7 filters are also used. There are other 1×1 filters, which will be discussed in a separate post. They may appear unusual at first glance, but they have interesting applications.

Filter count:

The filter count can be changed. It consists of two powers that can range from 32 to 1024. More filters show a much more powerful model, however, the risk rises of overfitting due to the count of higher parameters. Generally, at first, we start our proceedings with a definite number of filters in the first layer and eventually increase the number of filters as we approach deeper into the network.



2	2	7	3
9	4	6	1
8	5	2	4
3	1	2	6

9	7
8	6

Figure 4.3: How MaxPooling Works [49]

4.6.2 Stride

In convolution, stride indicates the number of pixels we can move in each convolution. When we apply convolution, there is a filter window which we can move by 1 pixel (by default) in either the x or y-direction. This is called a stride of 1 by 1. We can apply a stride of 2 by 2 which means the filter window will move in either x or y direction by 2 pixels. The striding is 1 by 1 by default.
We apply a filter matrix to the input image and we get an output feature map having a reduced size. This process of reducing the size of the input image is called padding. Thus through the padding, we reduce the size of the input image and get a featured map having a reduced size.

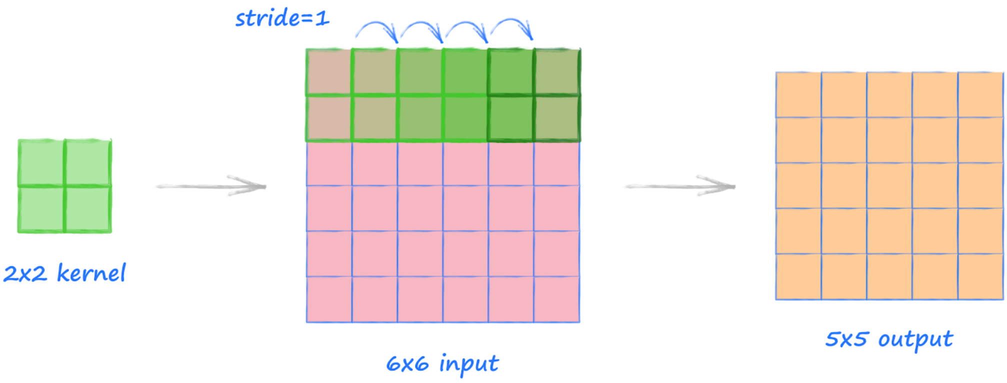


Figure 4.4: How Stride Works

4.6.3 Padding

The word "padding" is important in the context of convolutional neural networks (CNNs) since it describes the number of additional pixels that are added to a picture while the image is being processed by the CNN kernel. If, for instance, the padding in a CNN is configured to have a value of zero, then each and every pixel value that is subsequently added will also have a value of zero. We make extensive use of the padding that is available to us.

4.6.4 Zero Padding Layer

This layer basically adds a border of pixels with zeros. The zeros will be on the edges of a given image. Figure 4.5 illustrates if we pad our input with a border of zero valued pixels, what the resulting output size may get after convolving our input. The method of zero padding is one technique that enables us to maintain the size of the data that was first entered.

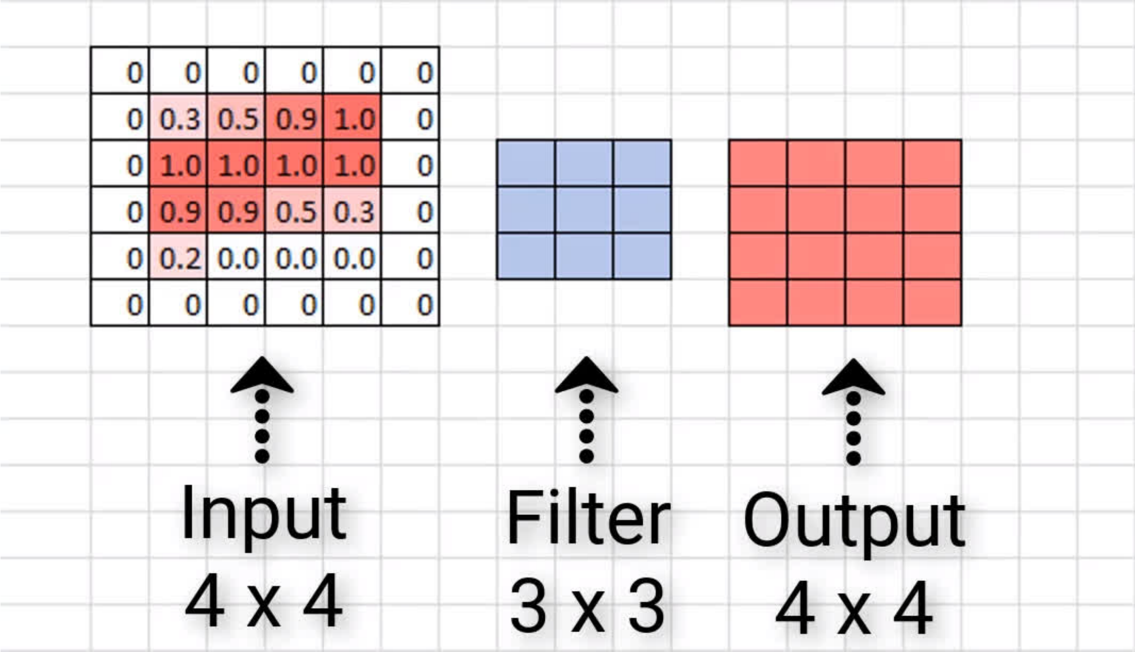


Figure 4.5: How Zero Padding Works [50]

4.7 Fully Connected Layer

Fully Connected Layer is the core of feed forward neural networks. The next few tiers of the network are made up of layers that are completely linked to one another. The output of the final pooling or convolutional layer is the input to the fully connected layer. This layer is the last step in the learning process. This output is first made flattened, and then it is sent into the layer that is fully connected.

4.8 Dropout

Dropout serves as a safeguard against the overfitting of the white training model. While training, dropout will occasionally remove neurons from the population. When this is done, it contributes to the overall resilience of the network.

4.9 ResNet-50

Residual Networks also known as Res-Nets learn residual functions depending on layer inputs instead of unreferenced functions. Residual nets permit these stacked layers to fit a residual mapping. They create a network by piling convolutions on top

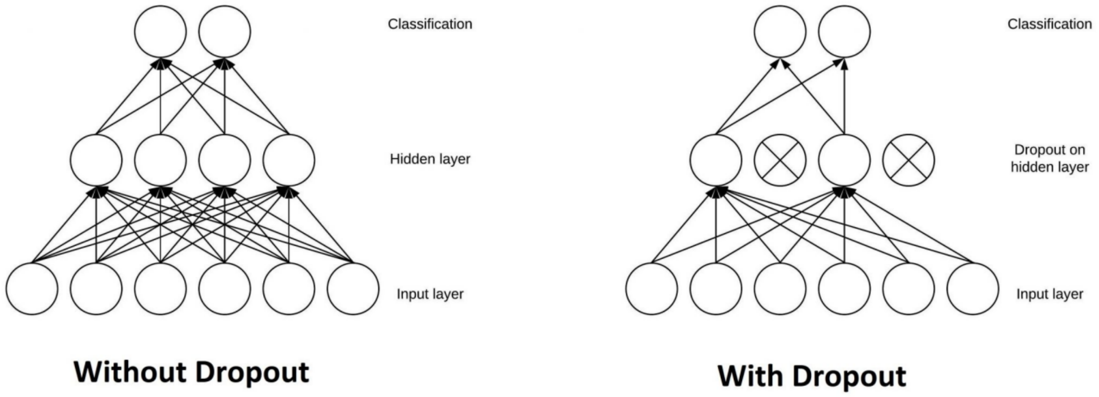


Figure 4.6: Dropout

of each other; for example, a ResNet-50 has fifty layers composed of these blocks. Formally, we let the arrayed nonlinear levels fit a further mapping of by denoting the desired sequence as. The original mapping is reshaped into. There is empirical evidence that these connections are easier to optimize and can benefit from higher depth.

4.10 VGG-19

The VGG network architecture was introduced first at 2014 [13]. VGG19 is a VGG model type that has total of 19 layers, 16 of them are convolution layers, 5 are MaxPool layers, 3 are fully-connected layers, and finally remaining 1 is the SoftMax layer. It carries over and improves mostly on ideologies out of its previous versions, and it uses the deep Convolutional neural layers in order to increase and improve accuracy. It could load a pre-trained version of the network from Image datasets which trained on more than a million pictures. It has been pre-trained to divide the pictures into 1000 segmented categories which include different components, stationery, and a variety of animals. As an outcome, the network has learned a rich classification model for a diverse set of images. The network accepts images with a resolution of 224 by 224.It are an example of traditional Convolutional Neural Network architecture. It was founded on a study of how to strengthen such networks. The network employs small 3 x 3 filters, and is characterized by its simplicity, with just pooling layers and convolutional layers as additional components. It can be used in a variety of ways, as suitable classification architecture for a large amount of datasets, for instance. Weights are easily available in other frameworks like Keras, so they may be explored with or utilized about any purpose the user wants.

4.11 InceptionV3

InceptionV3 is a Convolutional Neural Network architecture [14] that incorporates Smoothing of Labels that is factorized in 7 x 7 convolutions. It also uses an auxiliary classifier to produce Information label lower down the network onward with batch normalization for layers in the side head. If Inception is compared and contrasted

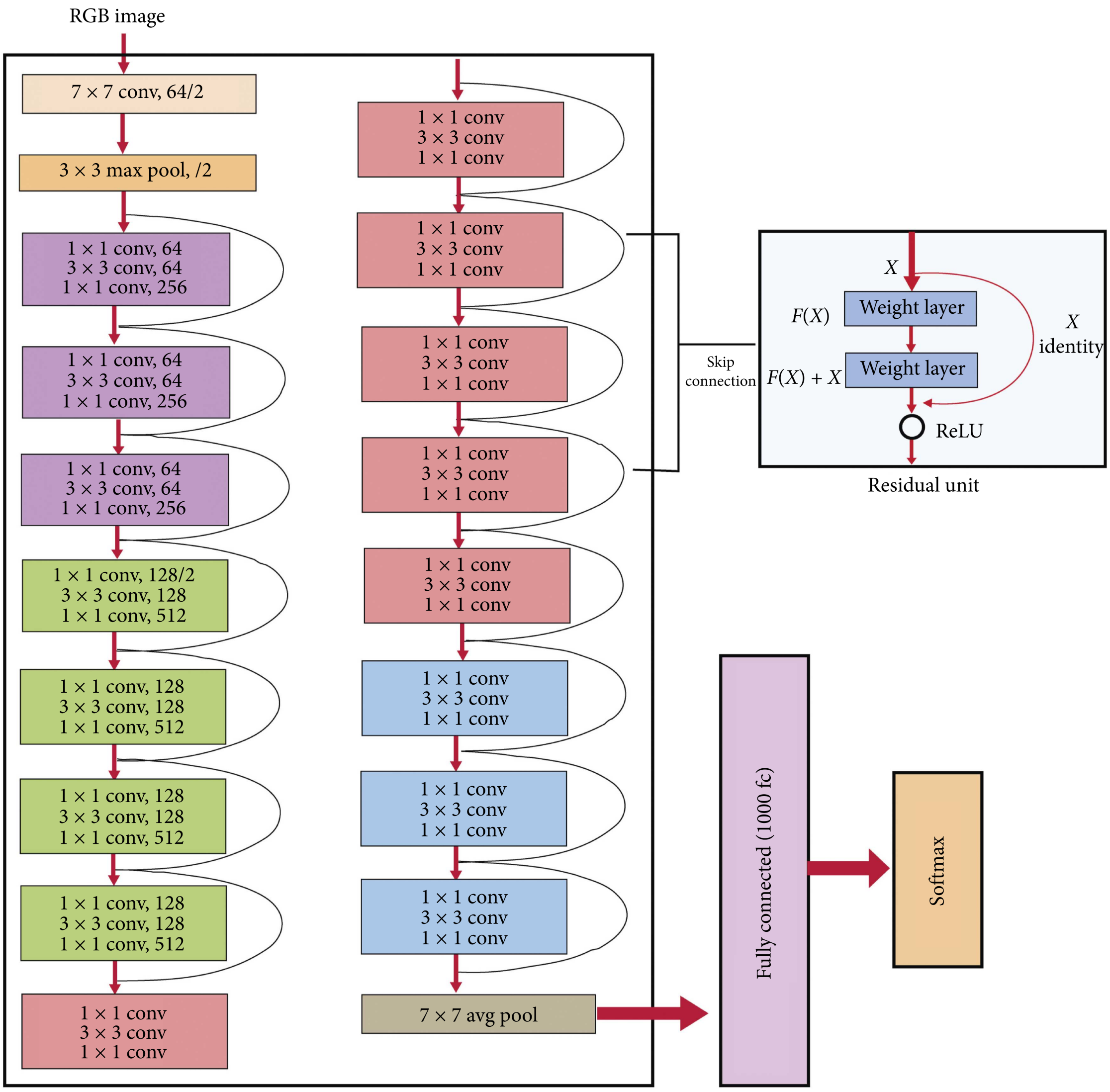


Figure 4.7: ResNet50

with VGG-Net, the networks of Inception have verified to be more computationally well- planned and significantly better structured, both in terms of the number of parameters propagated by the network and the economic cost provoked for both memory and other resources. In order to make alterations to an Inception network, it should be done extremely cautiously to make sure that the computational benefits are not lost. If an InceptionV3 model is considered, various systematic approaches have been suggested to slacken the constraints for a handier model adjustment. Factorized convolutions, regularization, dimension reduction and parallelized computations are some of the approaches used. An InceptionV3 architecture is progressively constructed one step followed by the other. At first, it is the use of Factorized Convolutional which aids in decreasing the computational efficiency by decreasing the number of parameters included in a network. Moreover, it keeps a check on the productivity of a network. The second step is a smaller convolution which operates by replacing bigger convolutions with smaller ones leading to more rapid training. If Asymmetric convolutions are considered, a 3 x 3 convolution can be substituted by a 1 x 3 convolution preceded by a 3 x 1 convolution. The number of parameters will rise by a tiny amount if a convolution of 3 x 3 is substituted with a convolution

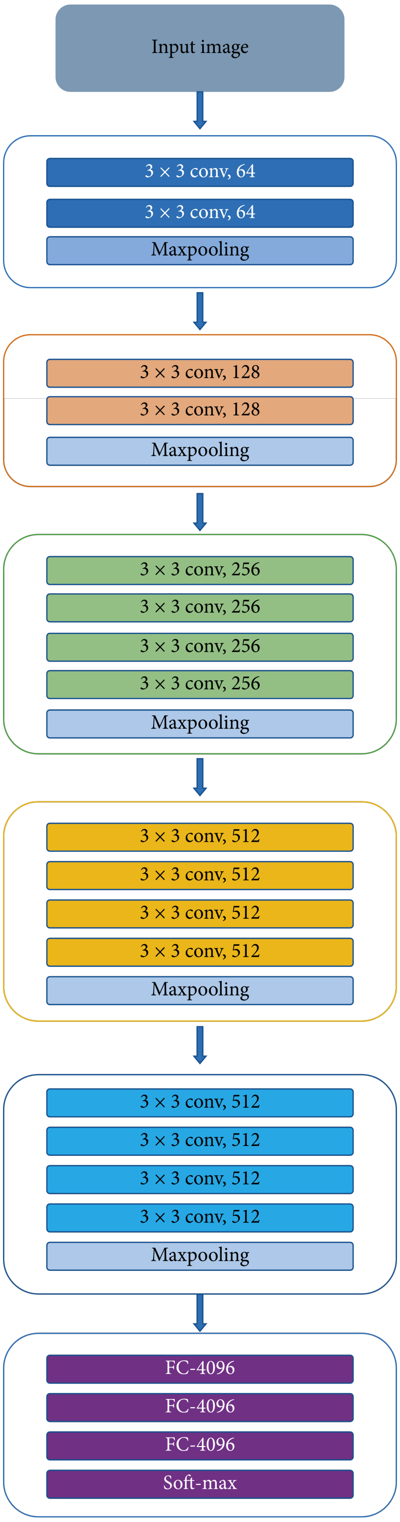


Figure 4.8: VGG19

of 2 x 2. A smaller CNN is deployed in the auxiliary classifier between the layers and adds losses to the main network during the training. Finally, to cut the volume of the Grid, pooling operations are employed.

4.12 EfficientNetB7

EfficientNet is the name given to the architecture [31] of a convolutional neural network in which the resolution, breadth, and depth are scaled consistently using a compound coefficient and a scaling algorithm. In the context of EfficientNet, unique model scaling strategies are implemented. Despite its simplicity, this strategy is highly effective. EfficientNet has a relatively basic architecture in which a baseline network is utilized to do the neural architecture search, an automated method for building neural networks. Due to its structure, both accuracy and efficiency are enhanced and improved to a large amount. EfficientNet tends to provide superior results compared to other models, and the overall number of FLOPs is drastically reduced. There are a variety of EfficientNet frameworks, with varying degrees of effectiveness and precision. EfficientNet is much smaller than other models with comparable ImageNet accuracy. The ResNet50 model, as shown in the Keras application, has a total of 23,534,592 parameters, yet it still surpasses the smallest EfficientNet model, which has just 5,330,564 parameters. Mobile inverted bottleneck MBConv, which was initially introduced in MobileNetV2, is the fundamental building block of EfficientNet. By combining direct shortcuts between bottleneck layers with depthwise separable convolution, which reduces computation by almost a fold, we may link a substantially less number of channels than expansion layers.

4.13 MobileNetV2

MobileNetV2 is a convolutional neural network approach [26] that is improved for cell phones. It is the second kind of technology that enables the picture-handling functionality of various well-known mobile apps. The design has likewise been installed towards systems like Tensor-Flow Lite. Versatile organizations should cautiously adjust signs of progress in PC vision and with the constraints of portable conditions, profound learning overall. Web indexes, for example, Google have been delivering updates to the Mobile Nets engineering consistently, consolidating the absolute most original thoughts in the profound learning space. It depends on an altered remaining design, with lingering associations between bottleneck layers. As a channel, the moderate upgrade layer utilizes lightweight profundity astute convolutions. The engineering of MobileNetV2 by and large contains the underlying completely functional regarding a convolution layer with 32 channels, there are 19 leftover bottleneck layers. The essential guideline behind MobileNetV2 is that the model's moderate information sources and yields, while the internal layer exemplifies the model's ability to change from lower-level ideas, for example, pixels to more prominent descriptors like picture classes. At last, easy routes, as customary remaining associations, take into account quicker preparing and more prominent precision, and exactness.

4.14 Vision Transformer (ViT)

The Transformer is mainly introduced by Vaswani et al. [22] and Dosovitskiy et al. introduce the concept of Vision Transformer [33]. They are image classification models built on the principles of Transformers. When an image is sent as input, the architecture divides that image into patches of specified size. Here, each patch is linearly embedded, positions are re-embedded, and the final vector sequence is sent to a conventional Transformer encoder. Adding an additional “classification token” to the sequence, which is learnable is the conventional method for doing classification.

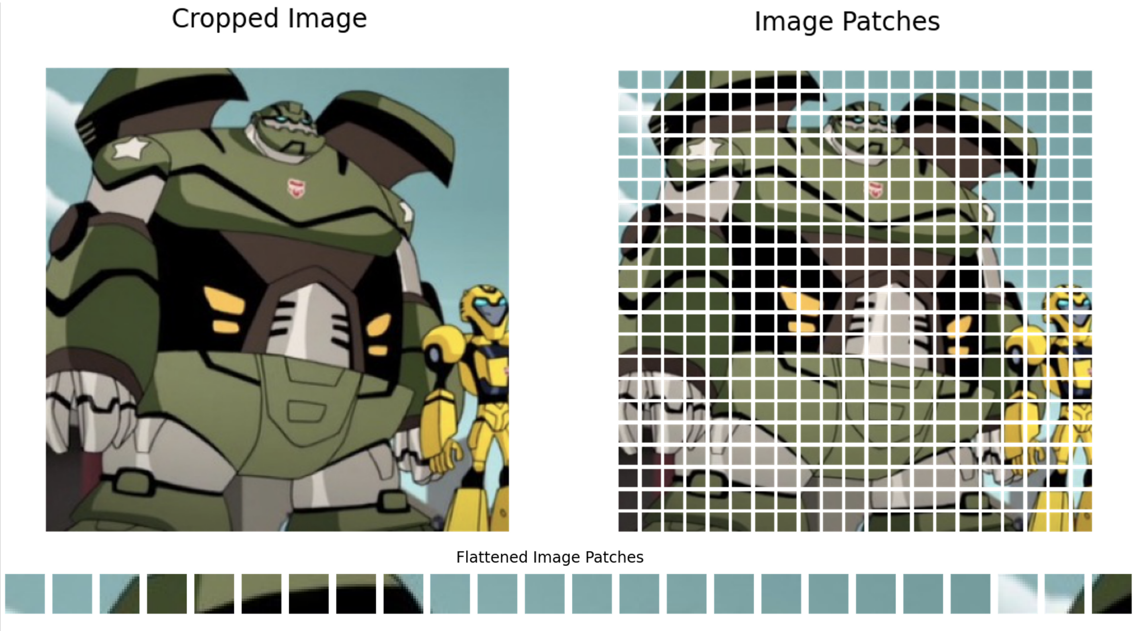


Figure 4.9: Vision Transformer Patching Approach

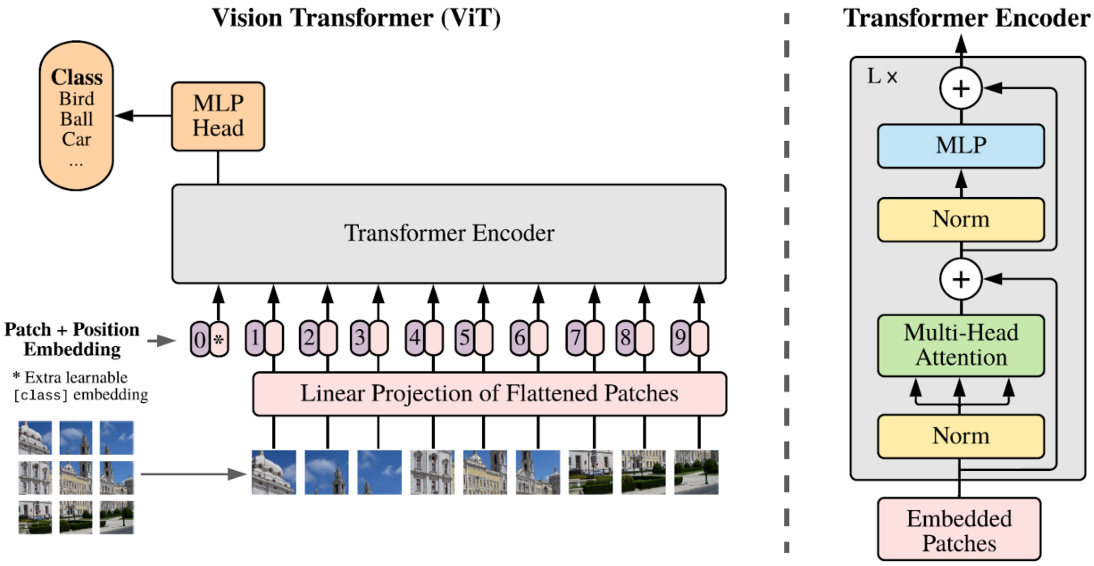


Figure 4.10: Vision Transformer Architecture

4.15 Involutional Neural Network (INN)

Convolution is the foundation of the majority of contemporary neural networks for computer vision. A convolution kernel is channel-specific and spatially neutral. This prevents it from adapting to diverse visual patterns in relation to distinct spatial places. In addition to location-related issues, the receptive area of convolution makes it difficult to capture protracted spatial connections.

Li et al. [37] reconsider the features of convolution in order to overcome the aforementioned problems. The authors propose the location-specific and channel-independent "Involution kernel". Due to the operation's location-specific character, the authors assert that self-attention comes under the Involution design paradigm.

To infer the concept of Involutions properly, we have to look at the process of convolution. Consider a tensor X with dimensions H, W, and C_{in} as an input. We take a collection of C_{out} convolution kernels with K, K, C_{in} shapes. With the multiply-add operation between the input tensor and the kernels, the output tensor Y possesses the dimensions H, W, and C_{out}.

In Figure 4.11, $C_{out} = 3$ which produces an output tensor with the shapes H, W, and 3. The convolution kernel is independent of the spatial position of the input tensor, rendering it location-independent. Alternatively, each channel in the output tensor is based on a distinct convolution filter, making it channel-specific.

The goal is to establish an operation that is both channel-agnostic and location-specific. It is difficult to implement these precise qualities. With a fixed number of Involution kernels (for each spatial point), variable-resolution input tensors cannot be processed.

The authors have proposed constructing each kernel based on specified spatial positions in order to resolve this issue. This technique should facilitate the processing of variable-resolution input tensors. Figure 4.12 illustrates this method of kernel creation. Here, $K \times K \times C$ filters are generated, where C is the number of channel groups. Instead of employing a single filter and broadcasting it overall C input channels, we generate C filters and broadcast them into each of the C input channels.

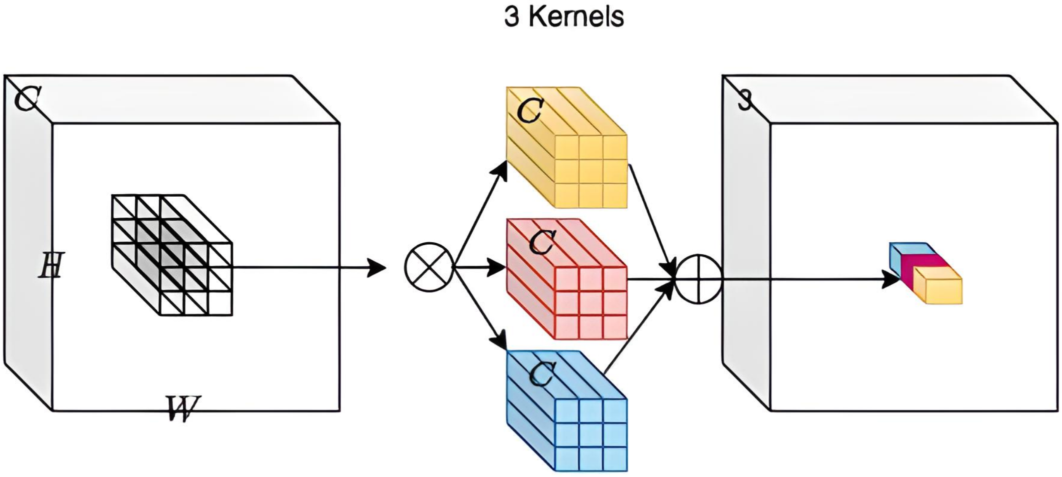


Figure 4.11: Convolution Process

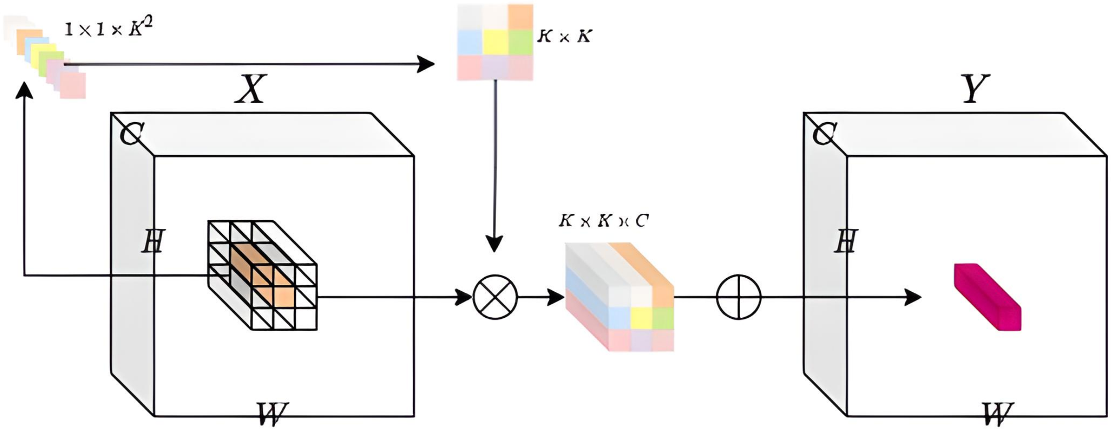


Figure 4.12: Involution Process

Chapter 5

Research Methodology

In-depth discussions of our approach to do data pre-processing, our proposed algorithm for PPPC, and the suggested model's structure are provided in this section.

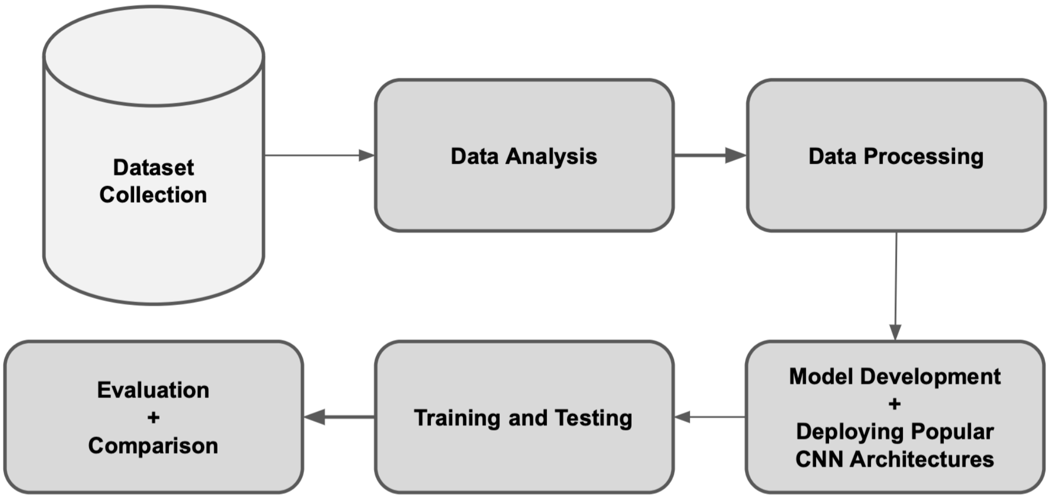


Figure 5.1: Workflow of Our Proposed Approach

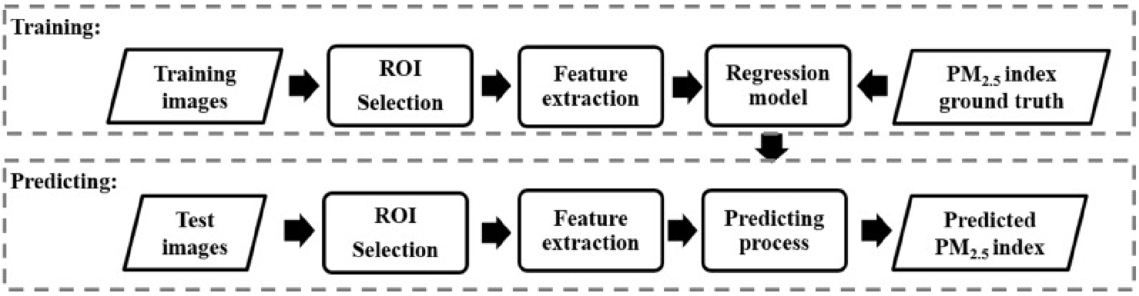


Figure 5.2: Outdoor Image Analysis and Estimation Approach to Predict PM [18]

To solve the problem, first of all, we take pictures of the people of Dhaka and ourselves and the AQI of that time and date, which is elaborately described in Section 3. After that, we do some data analysis based on that.

5.1 Data Preprocessing

Our dataset is made up of photos with varying resolutions, however, our proposed model demands a fixed input dimensionality. As a result, as said in the earlier section, we downsample the photos to a fixed resolution of 200 × 200. Given a rectangular image, we first rescale it so that any unwanted pixels that take up around 50% of the image are omitted, and the new resolution of the image is 120 × 200. We do not conduct any extra pre-processing on the photos. As a result, we train our network using the pixels' raw RGB values. We part our dataset into two parts initially. One part containing 90% percent of our data is used for training then the rest 10% part are equally distributed to be used for testing. From the 90%, we take 10% for validation.

5.2 Proposed Model Architecture

The architecture of our network is summarized in Figure 5.3. It contains 9 learned layers — one zero-padding, five convolutional, and two fully-connected.
The proposed model begins with inputting an image which consists of total three-channels. They are taken from the dataset, and scaled to a resolution of 120 × 200 pixels. We initially use ZeroPadding2D layer with padding of (3,3). Then five Conv2D layers are used which consists a kernel of (3,3) size. Afterward, MaxPooling layers are employed at the end of the Conv2D layers which has a pool size of (2,2) in order to minimize the layers' computational cost. Also, a default stride has been applied for the convolutional layers.
For the activation function, we choose Rectified Linear Unit (ReLU) as an activation function as the gradient is unsaturated, resulting significantly faster calculation of stochastic gradient descent (SGD) comparing other activation functions, including widely used functions like Sigmoid / Tanh function. Next, we transform the values into a 1D array and begin adding fully connected (FC) layers to the CNN with 128 nodes. To avoid overfitting in the model, we utilize a 10% dropout after the Dense layer to the final output layer in our experiment. The linear activation function is used as a network classifier at the output layer. It is a simple straight-line activation function in which our function is proportionate to the weighted sum of neurons or input. Linear activation functions can have a wider range of activations as options, and a positively sloped line may enhance the activation level as the input rate increases.
Table 5.1 shows the optimum parameters applied in different datasets for the proposed Deep CNN model.

Hyperparameter					
Image Input Size	Epoch	Batch Size	Learning Rate	Dropout Rate	Parameters
120 x 200	350	8	0.000001	10%	4,849,601

Table 5.1: Hyperparameter of Our Proposed Model

After some experimentation, this is the most optimum architecture we have found to solve this problem. The framework we propose has the most minimized number of parameters as well as performs the best. Hence, it uses comparatively fewer computing resources.

Layer	**Shape of Output**	**of Parameters**
Zero Padding 2D Layer	(None, 126, 206, 3)	0
Conv2D Layer	(None, 124, 204, 32)	896
Batch Normalization	(None, 124, 204, 32)	128
Conv2D Layer	(None, 122, 202, 64)	18,496
Batch Normalization	(None, 122, 202, 64)	256
2D Max Pooling	(None, 61, 101, 64)	0
Conv2D Layer	(None, 59, 99, 128)	73,856
Batch Normalization	(None, 59, 99, 128)	512
2D Max Pooling	(None, 29, 49, 128)	0
Conv2D Layer	(None, 27, 47, 256)	2,95,168
Batch Normalization	(None, 27, 47, 256)	1,024
2D Max Pooling	(None, 13, 23, 256)	0
Conv2D Layer	(None, 11, 21, 512)	11,80,160
Batch Normalization	(None, 11, 21, 512)	2,048
2D Max Pooling	(None, 5, 10, 512)	0
Flatten	(None, 25600)	0
Dense	(None, 128)	32,896
Dropout	(None, 128)	0
Dense	(None, 1)	129

Table 5.2: Number of Parameters and Shape of Output from Different Layers of Proposed Architecture [Here, Conv2D = 2D Convolutional]

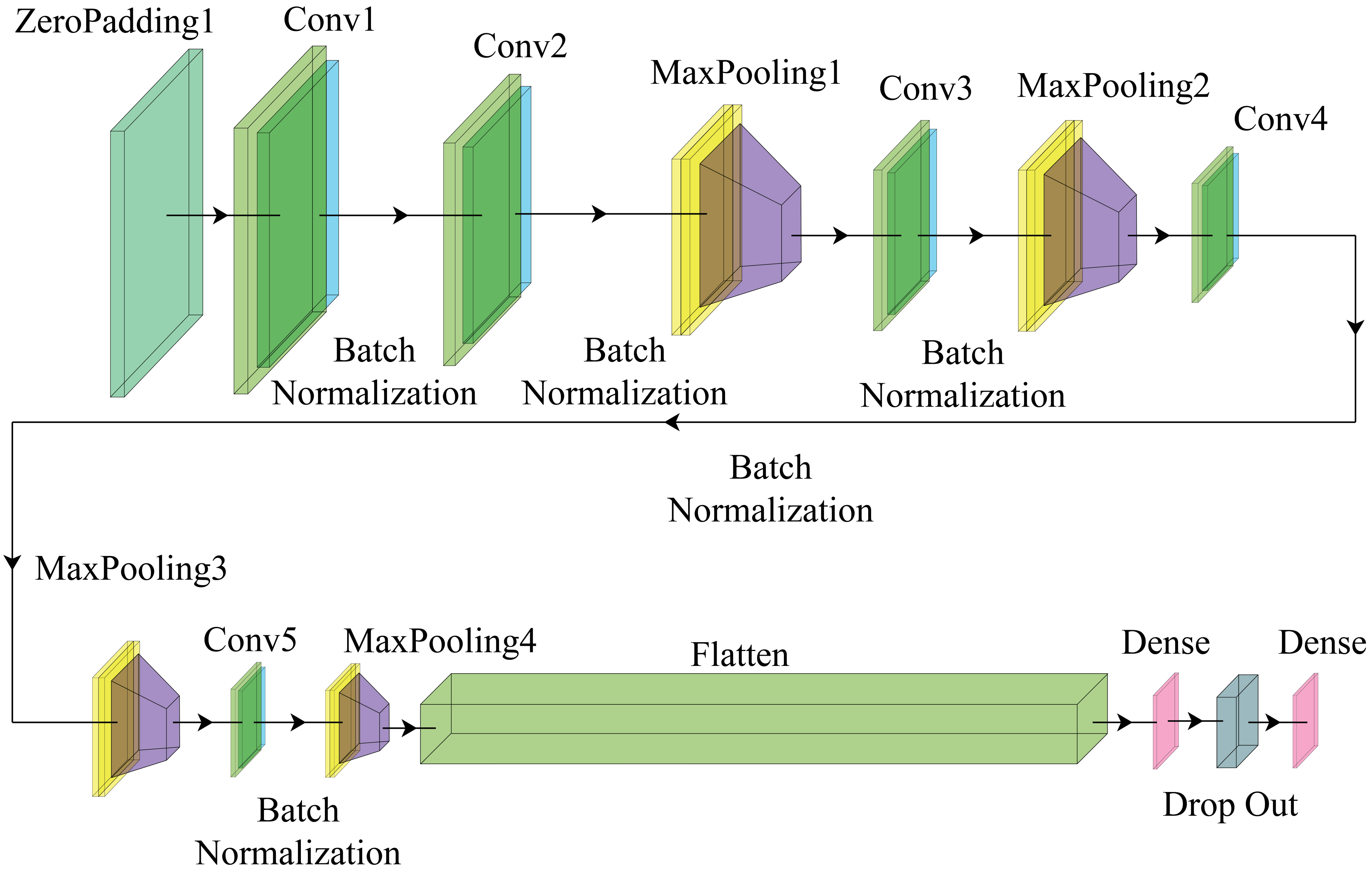


Figure 5.3: Illustration of the Proposed Deep CNN Architecture

Chapter 6

Experimental Evaluation

Following the design of our proposed architecture, we test our model using different standard metrics for regression-based tasks and compare it to many current deep learning architectures. Our major goal is to arrive at a paradigm that yields the optimal outcome, irrespective of the computing resources required.

6.0.1 Experimental Setup

All other popular CNN models, including our proposed Deep CNN model, are trained and tested using Tensorflow, Keras, Pillow, and OpenCV Python libraries. For the study, Python 3.8 is used alongside Tensorflow 2.9, Keras 2.9, and OpenCV 4.5. The models are trained and evaluated on two different devices; one with an NVIDIA RTX 2070 with 7.5 TeraFLOPs of performance and another with an NVIDIA RTX 3080TI GPU which has 34.1 TeraFLOPs of performance.

6.0.2 Experimental Findings

Commonly, three performance measures are kept in consideration while analyzing the prediction results to evaluate the performance of the proposed model: MAE, MSE, RMSE, and R^2. The metrics are represented as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |F_p - F_t| \tag{6.1}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (F_p - F_t)^2 \tag{6.2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (F_p - F_t)^2} \tag{6.3}$$

where F_p is the predicted value and F_t is the actual value.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - y_i')^2}{\sum_{i=1}^{N} (y_i - \text{avg}(y))^2} \tag{6.4}$$

where $avg(y)$ is the average forecast value, $y_i^{'}$ is the h forecast value, and y_i is the i^{th} observed value, $i = 1, 2....N$. R-squared has a maximum value of one, which shows

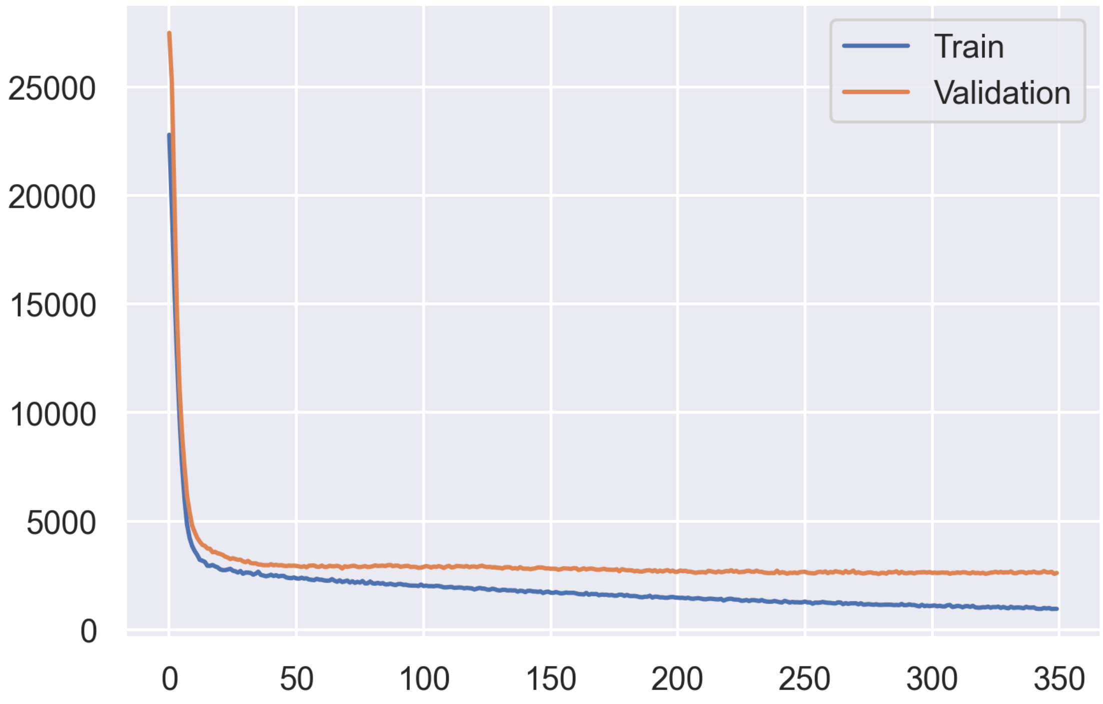


Figure 6.1: MSE Loss

MSE	RMSE	MAE	R^2
1782.08	42.22	29.68	0.387

Table 6.1: MSE, RMSE, MAE, and R^2 Value for Proposed Model

the best possible match, and its value rises as the degree of agreement between the observed value and the projected value grows.

After doing research on our problem, we discover that our model shows an MSE of 1782.08, an RMSE of 42.22, an MAE of 29.68, and a R^2 of 38.7 percent when it is put through its tests, as shown in Table 6.1. The value of R^2 that we are able to get indicates that 39% of the data matched our regression model. If the difference between the anticipated PM2.5 data and the observed PM2.5 data is less than fifty, then the forecast may be regarded as satisfactory. This is because a PM2.5 value difference of fifty or less is regarded as a close prediction. In addition, we find that the correlation coefficient of the model is 0.6304, which, when considered in light of the data from the actual world, is an acceptable correlation. We present a scatter plot of predicted vs observed PM2.5 values based on the 182 test samples in Figure 6.4. The plot proves how similar the measurements are here. Our achieved value of correlation coefficient is more than 0.5 which means when one variable changes, the other variables change in the same direction. The coefficient of determination, abbreviated as R^2, is most often interpreted as a measure of how well a regression model fits the data that was observed. Also, the model takes a runtime of 2140.25 seconds.

To verify our model, we also run k-Fold Validation using our dataset where k=10. After running our model through the folds, we observe that the loss difference is less than 10%.

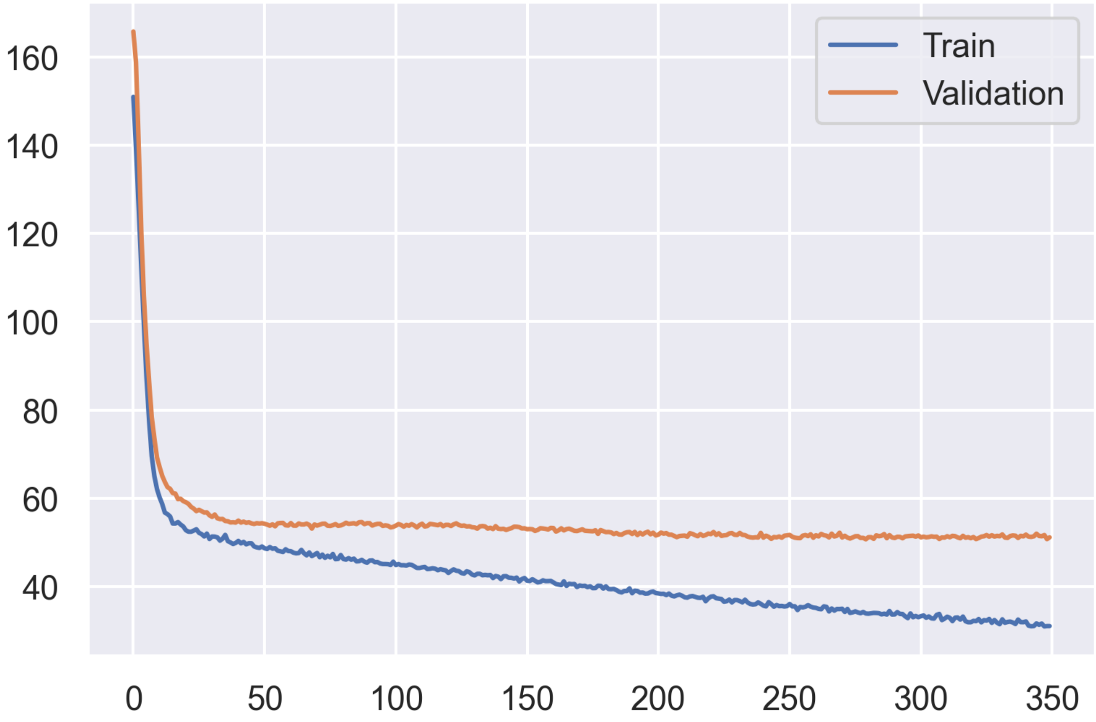


Figure 6.2: RMSE Loss

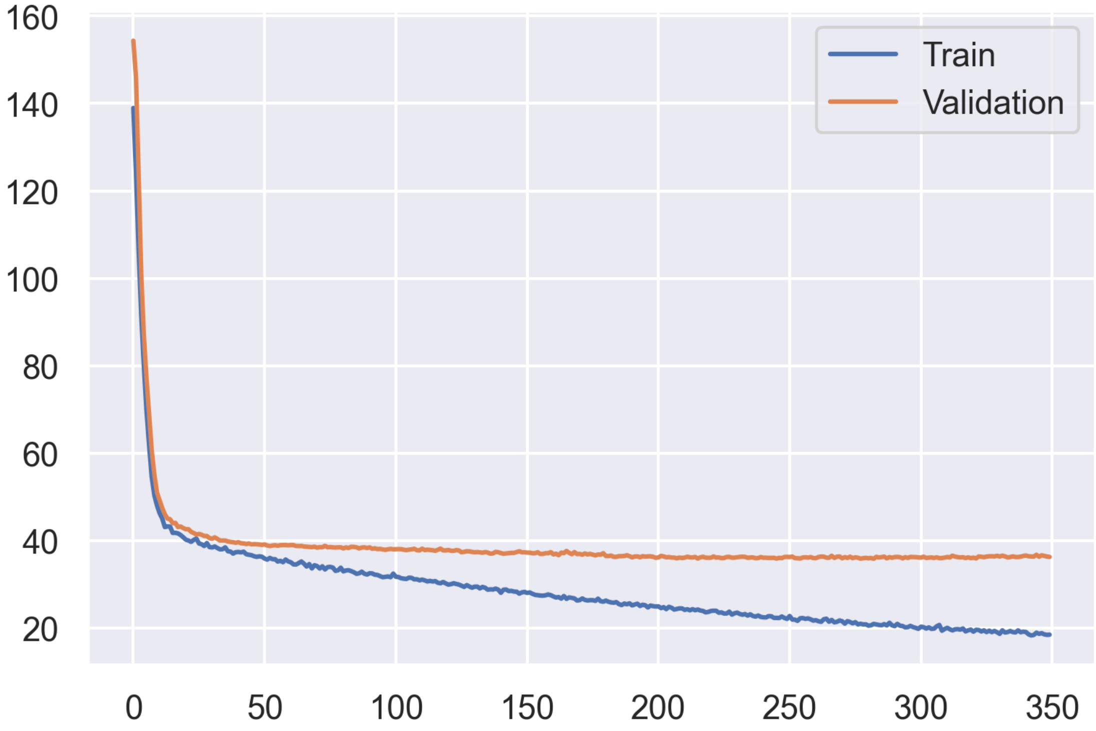


Figure 6.3: MAE Loss

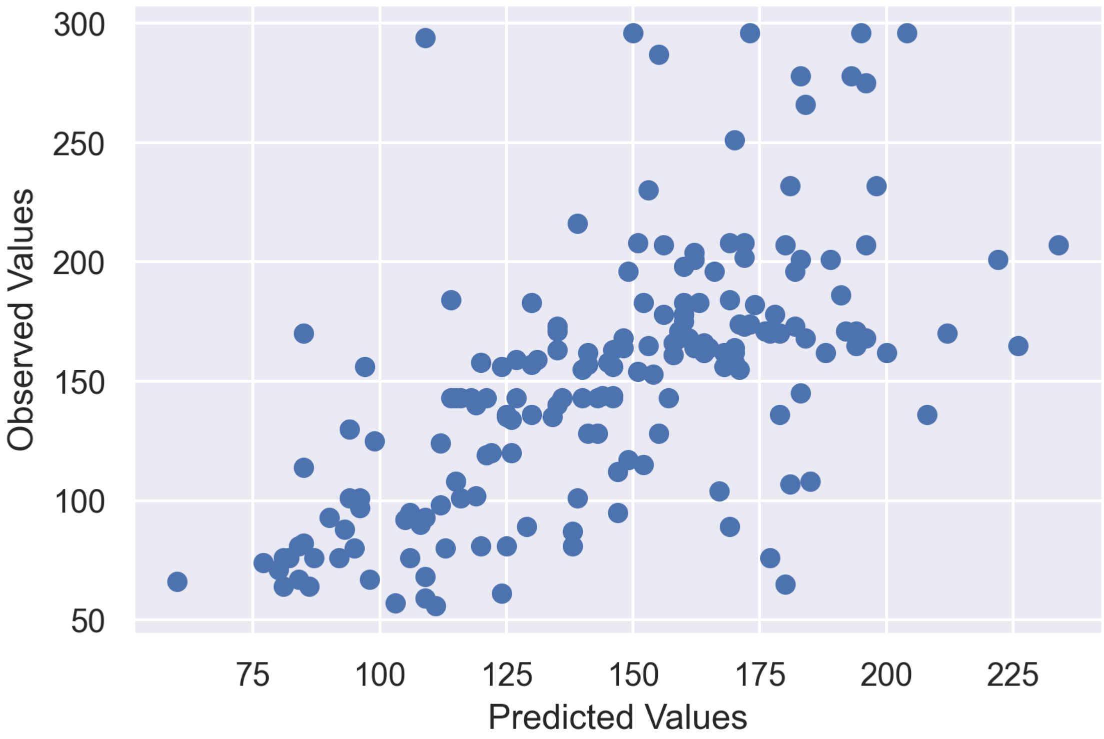


Figure 6.4: Correlation of PM2.5 Index (Estimated VS Observed)

After evaluating our model, as no method in the literature can be used directly for performance comparison with their reliance on different datasets, we compare it with some deep learning architectures available, including ResNet-50, VGG-19 , InceptionV3, EfficientNetB7, and MobileNetV2 that use a pre-trained weight of the ImageNet dataset. From Table 6.2, we observe most of these architectures use larger amount of parameters in total whereas the proposed architecture employs far fewer parameters. Again, if we compare with MobileNetV2 which has a lesser amount of parameters than our approach, our approach shows significantly better results than MobileNetV2. Here, we can see negative R^2 values of other models and it is usually negative when the chosen model does not follow the trend of the data.

When we observe some output images, we see that, when the noise amount is the least, it gives the closest prediction value, in comparison to the ground/observed value. Again, if we look into some pictures, for instance, Figure 6.8, it gives a wrong prediction value due to the noise amount. For many images, it gives to give a close value when the noise amount is huge. Again, in Figure 6.10, it gives an incorrect prediction since there are some colorful pixels which hamper the prediction.

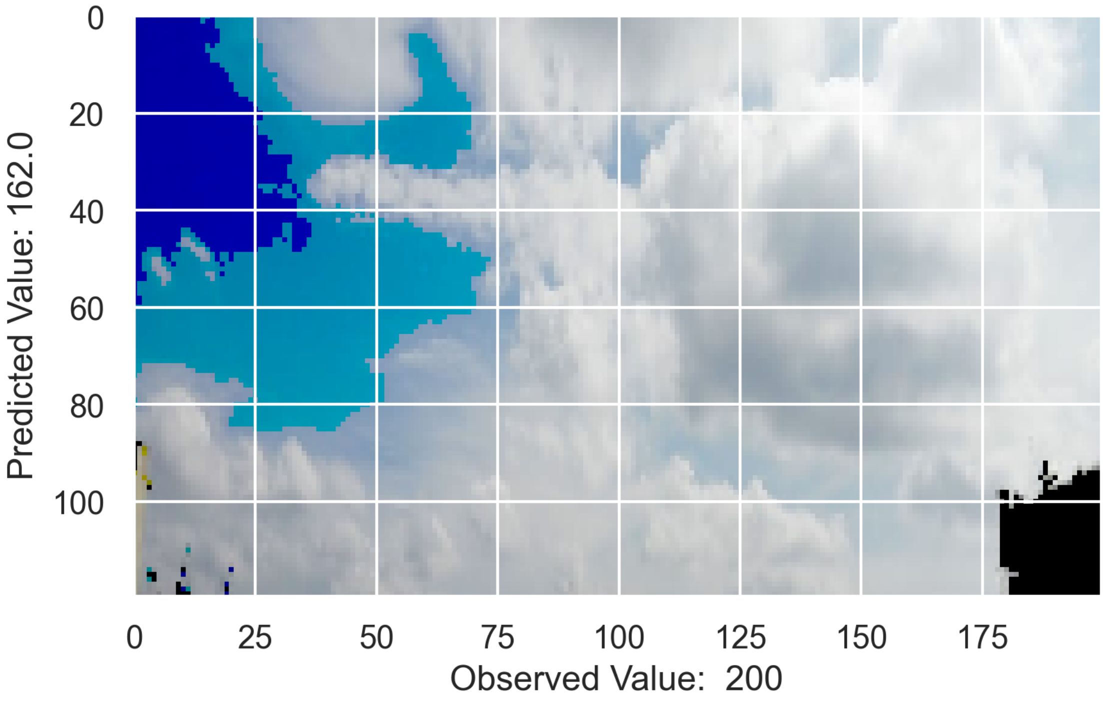


Figure 6.5: Output 1 [Medium Error]

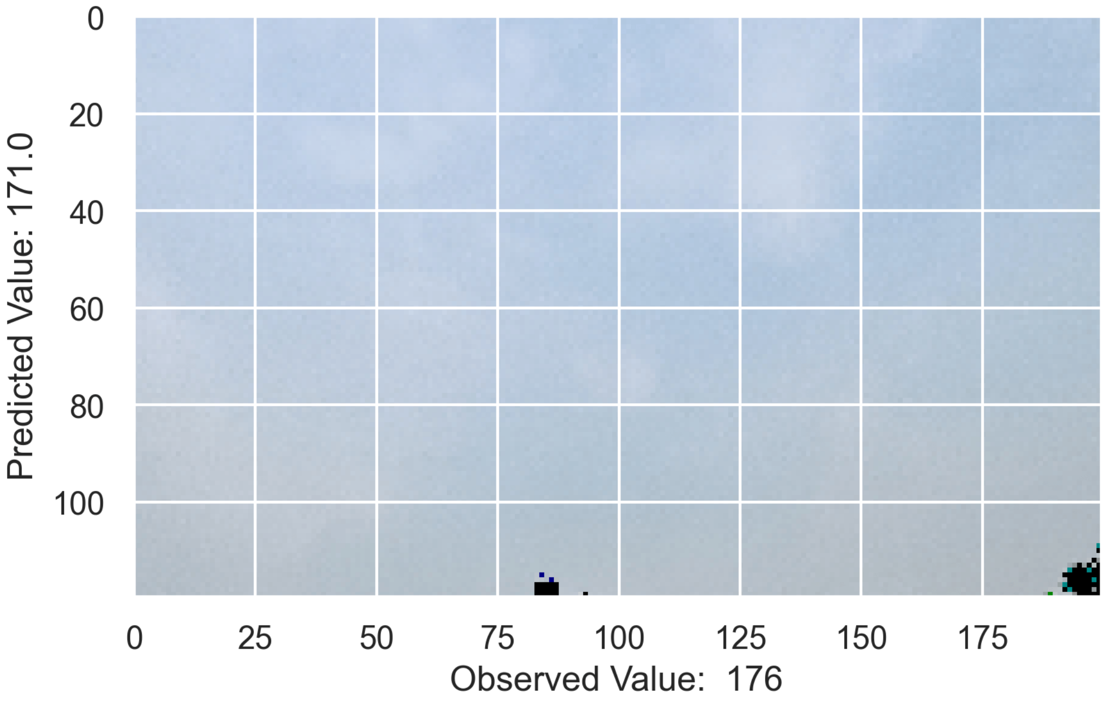


Figure 6.6: Output 2 [Low Error]

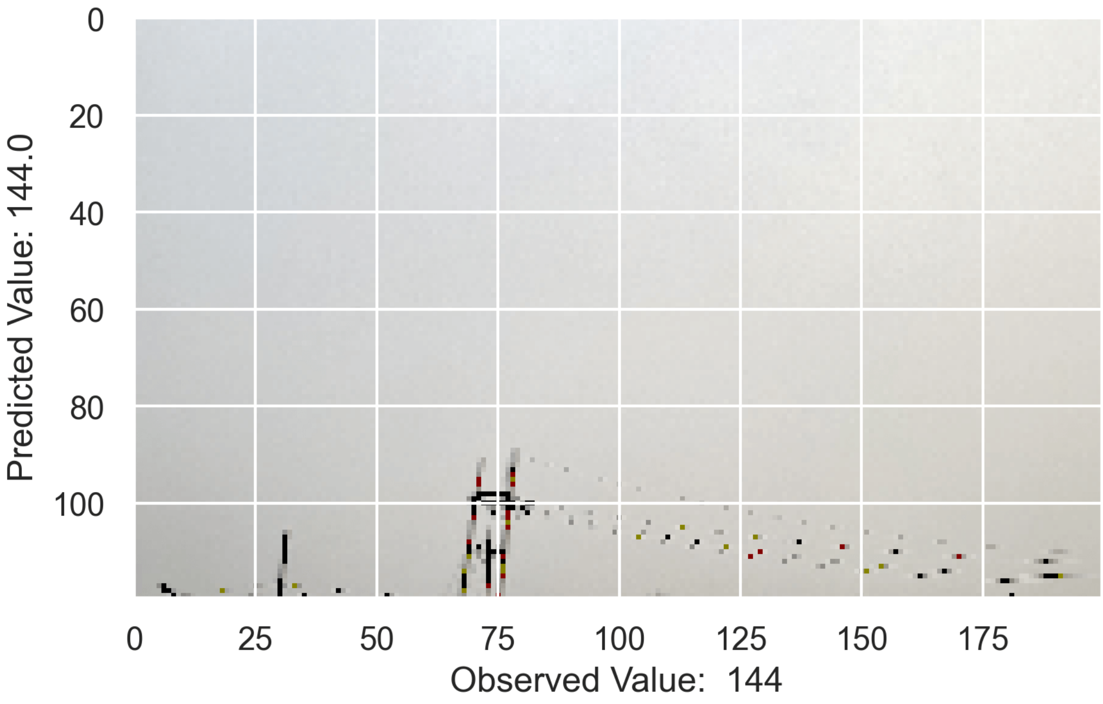


Figure 6.7: Output 3 [Low Error]

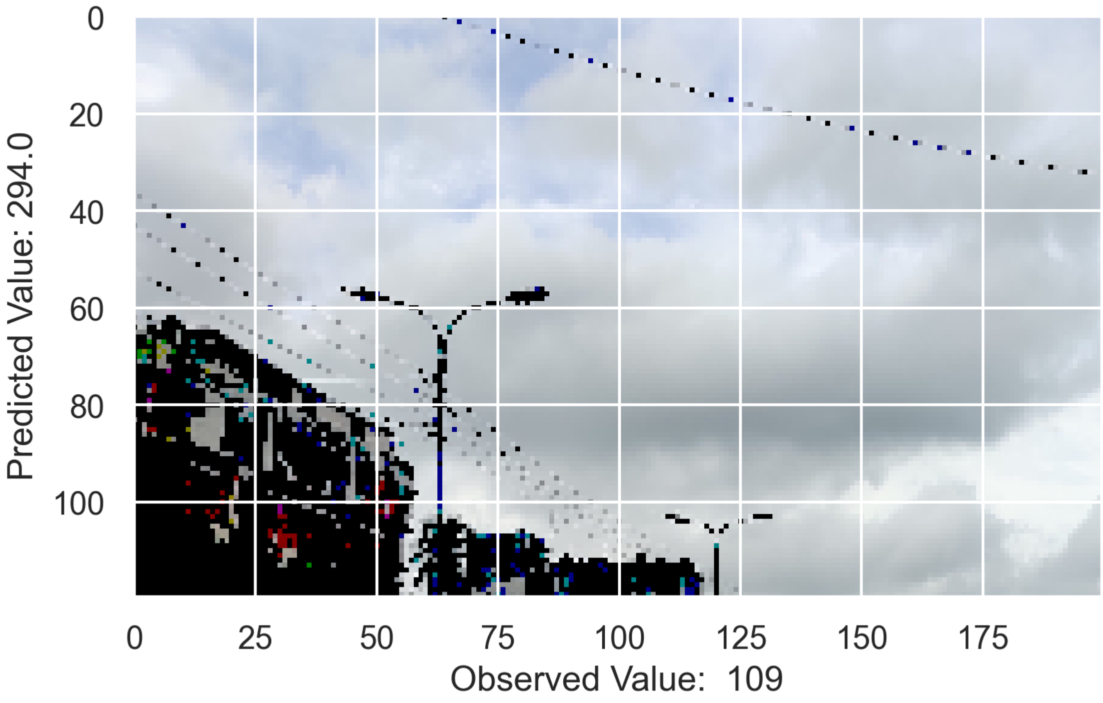


Figure 6.8: Output 4 [High Error]

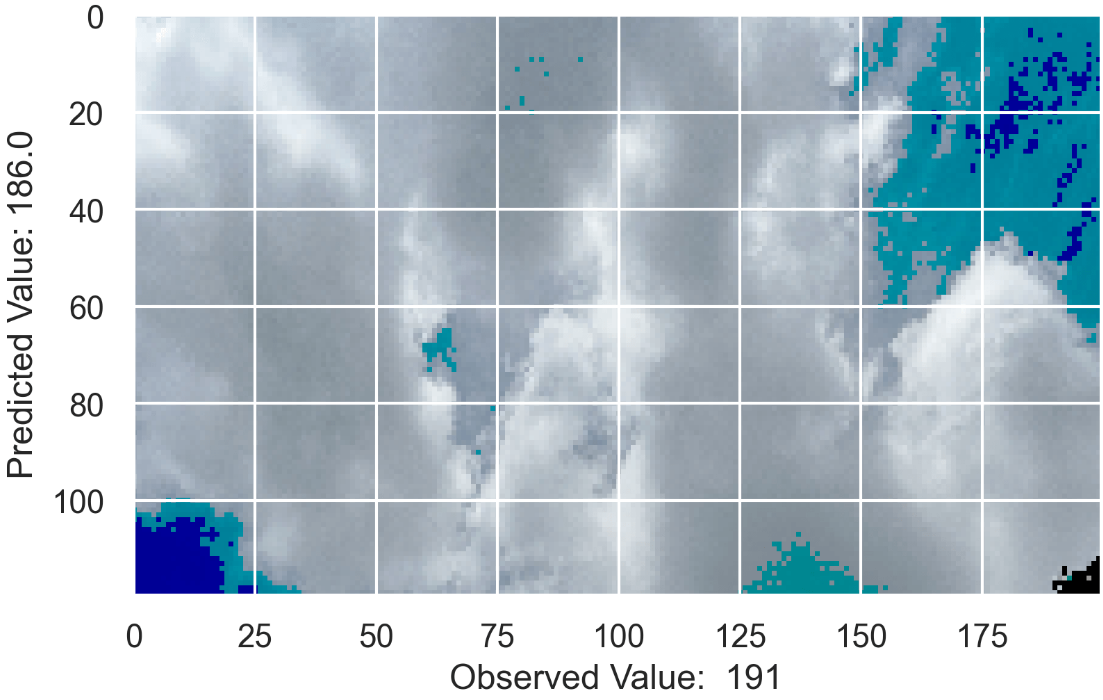


Figure 6.9: Output 5 [Low Error]

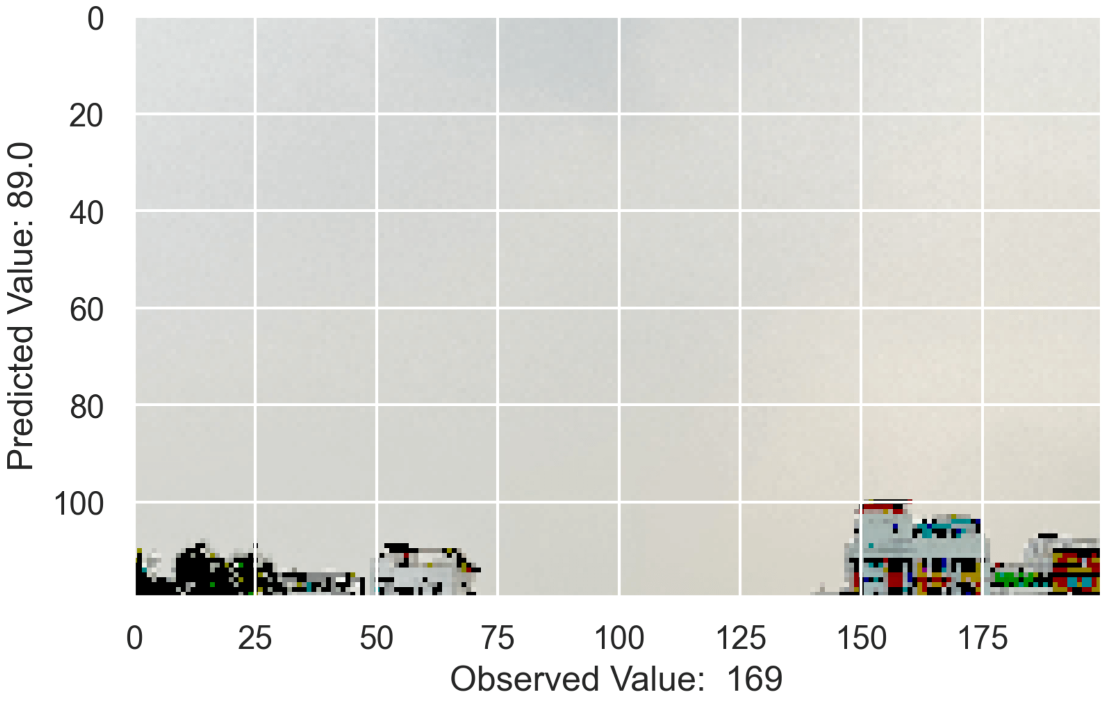


Figure 6.10: Output 6 [High Error]

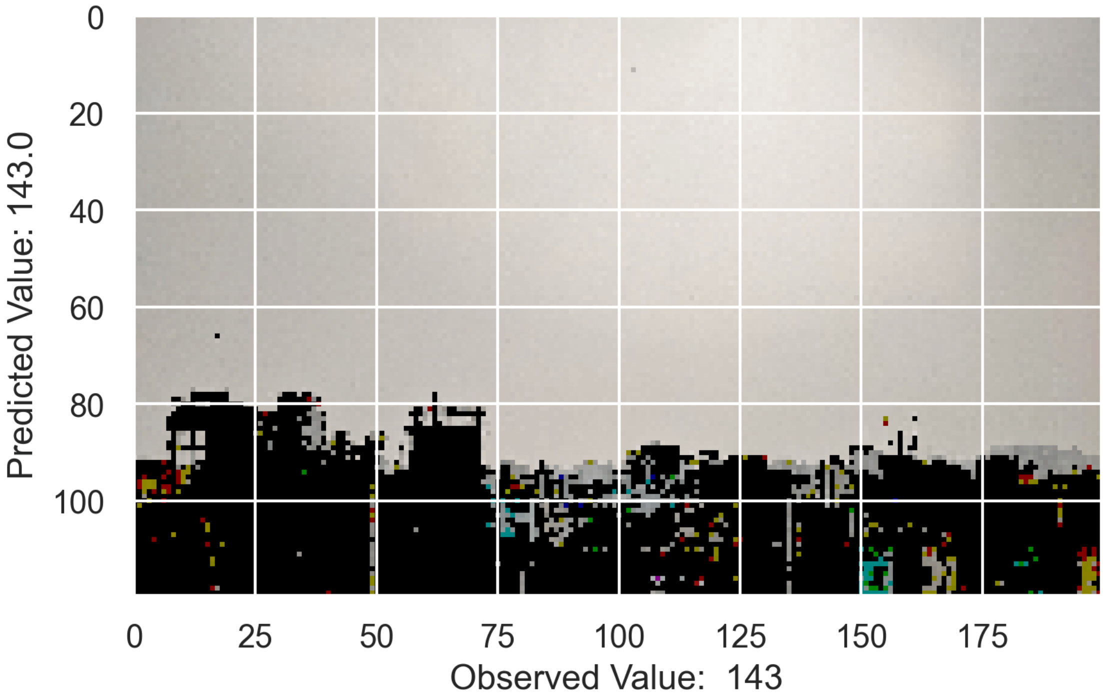


Figure 6.11: Output 7 [Low Error]

Deep Learning Architectures	Parameters (in Millions)	MSE	RMSE	MAE	R^2
ResNet-50	23.6	2840.57	53.30	40.23	0.03
VGG-19	20	8034.04	89.63	74.15	-1.78
InceptionV3	21.8	5243.73	72.41	55.16	-0.80
EfficientNetB7	64	2945.95	54.28	41.35	-0.01
MobileNetV2	2.2	2746.39	52.41	40.09	0.05
ViT	21.7	3304.33	57.48	44.66	-0.14
INN	**0.032**	25397.24	159.37	149.93	-7.79
Ours	4.8	**1782.08**	**42.22**	**29.68**	**0.387**

Table 6.2: Comparison after Training Different Deep Learning Architectures in Our Dataset (Here, Epoch=350 for Each Model)

6.1 Baseline Testing

To get the best model possible, we tweak our model with a few approaches, bringing changes in:

- Different Numbers of Nodes in Final Dense Layer before Output Dense Layer
- Different Learning Rates
- Different Batch Sizes

6.1.1 Different Numbers of Nodes in Final Layer before Output Dense Layer

We test our model in nine different types of pre-output dense layers.

- 128
- 2
- 4
- 16
- 32
- 64
- 96
- 256
- 512

After testing the model and tweaking this, we can see a better test result in 256 nodes, considering MAE, MSE, RMSE, Correlation Coefficient, Average Difference from True Value, Near Percentage, and R^2. But in terms of runtime, 32 nodes show the best result.
In Figure 6.19, we get the percentage of images can be considered as truly close to actual value since a difference of 50 is considered to be in the same type of class.

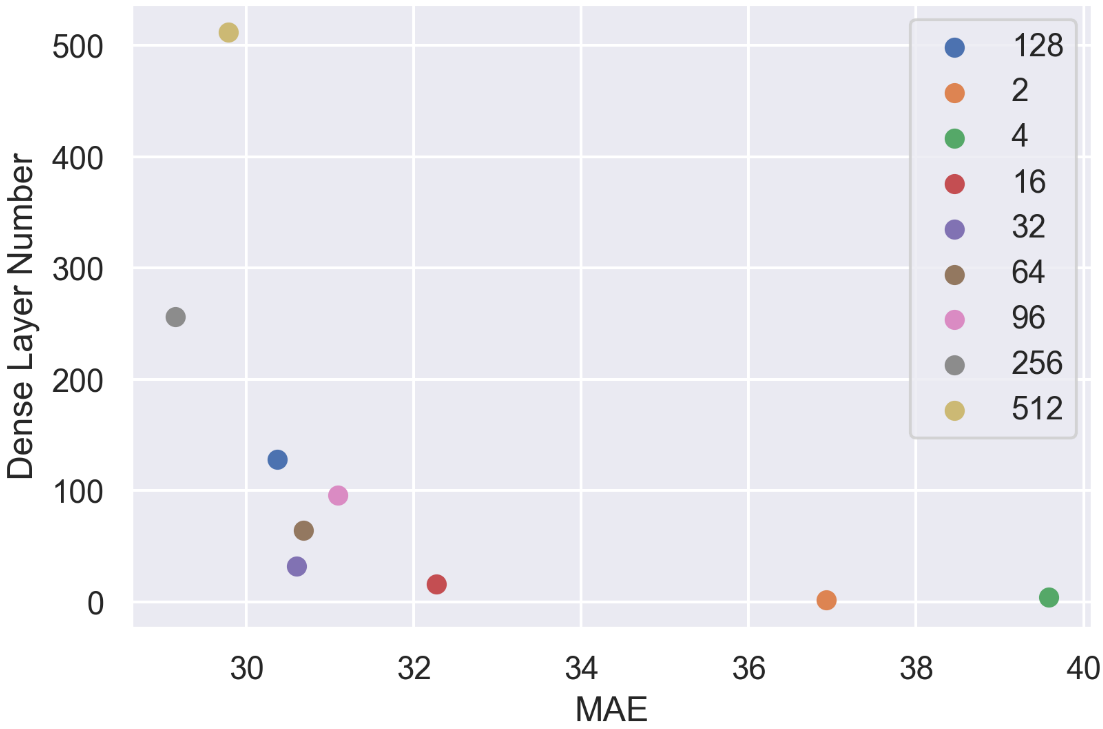


Figure 6.12: MAE in Different Numbers of Nodes (Lower, The Better)

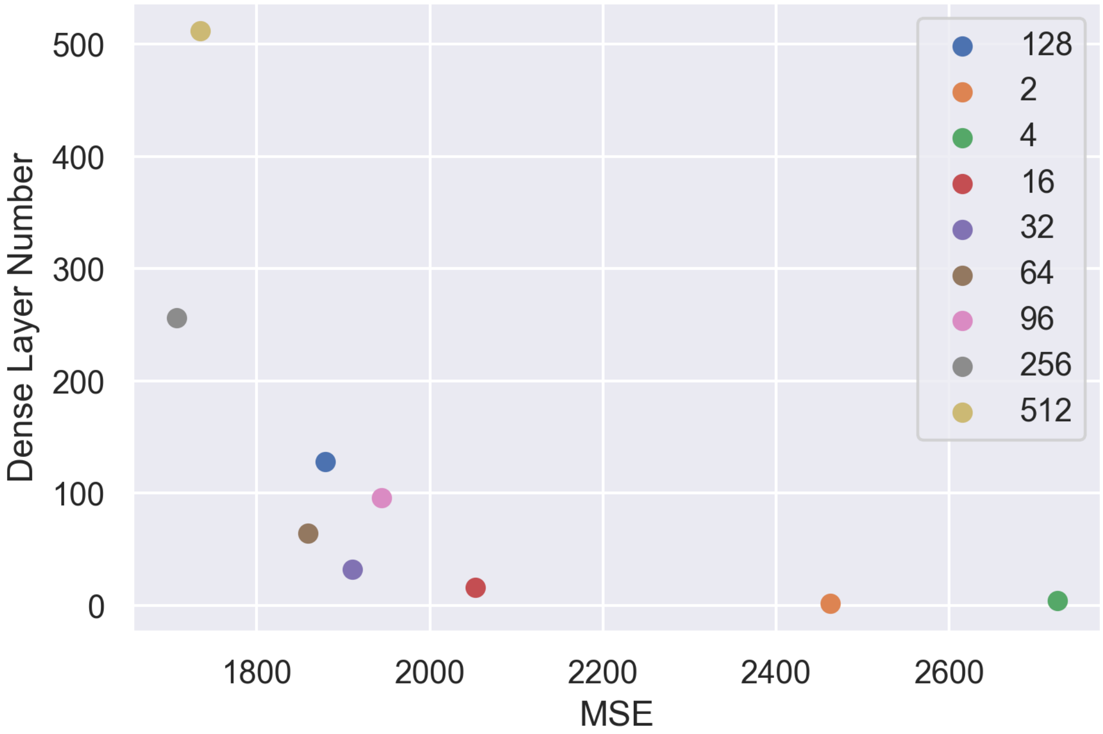


Figure 6.13: MSE in Different Numbers of Nodes (Lower, The Better)

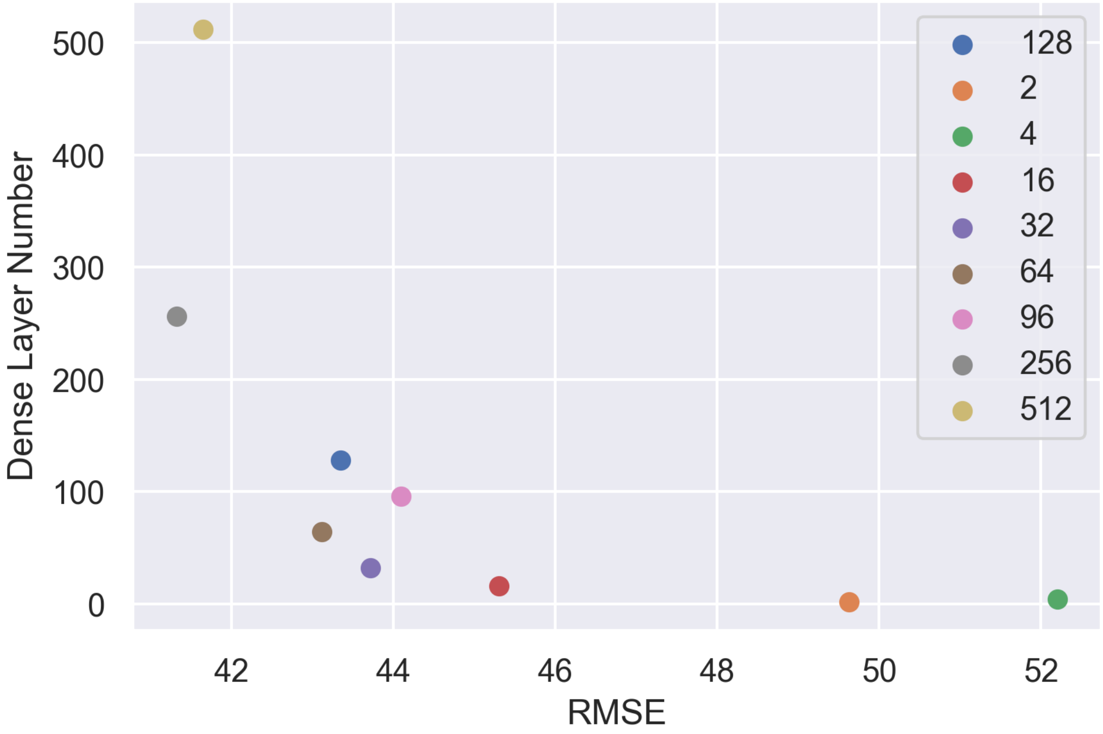


Figure 6.14: RMSE in Different Numbers of Nodes (Lower, The Better)

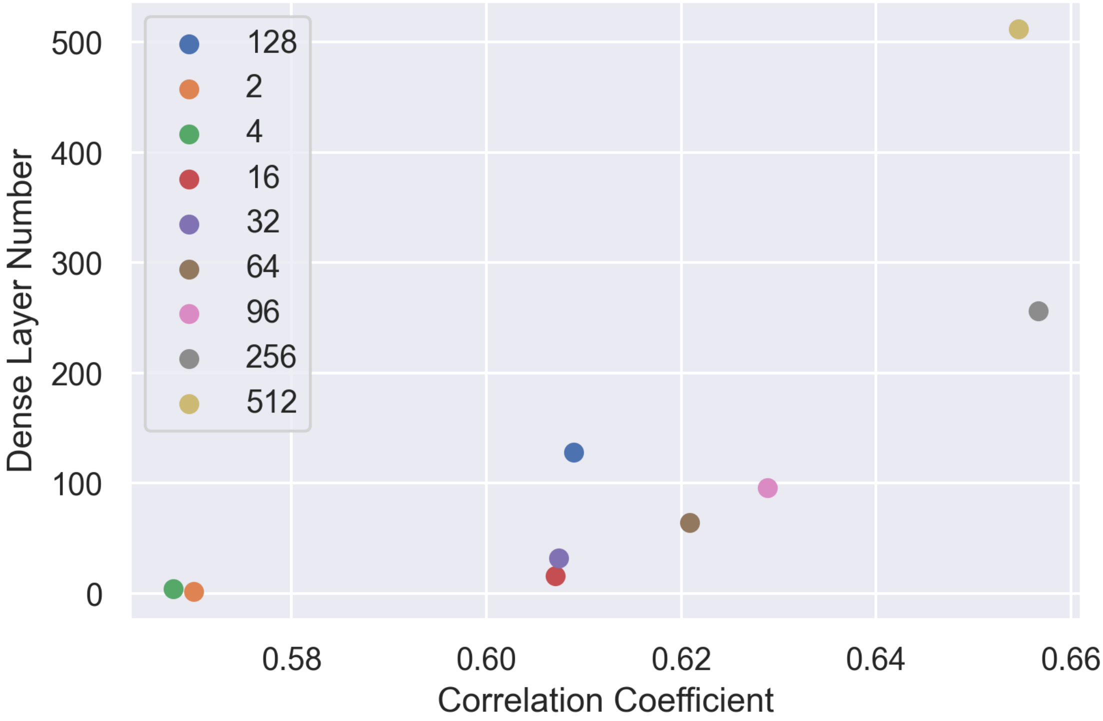


Figure 6.15: Correlation Coefficient in Different Numbers of Nodes (Higher, The Better)

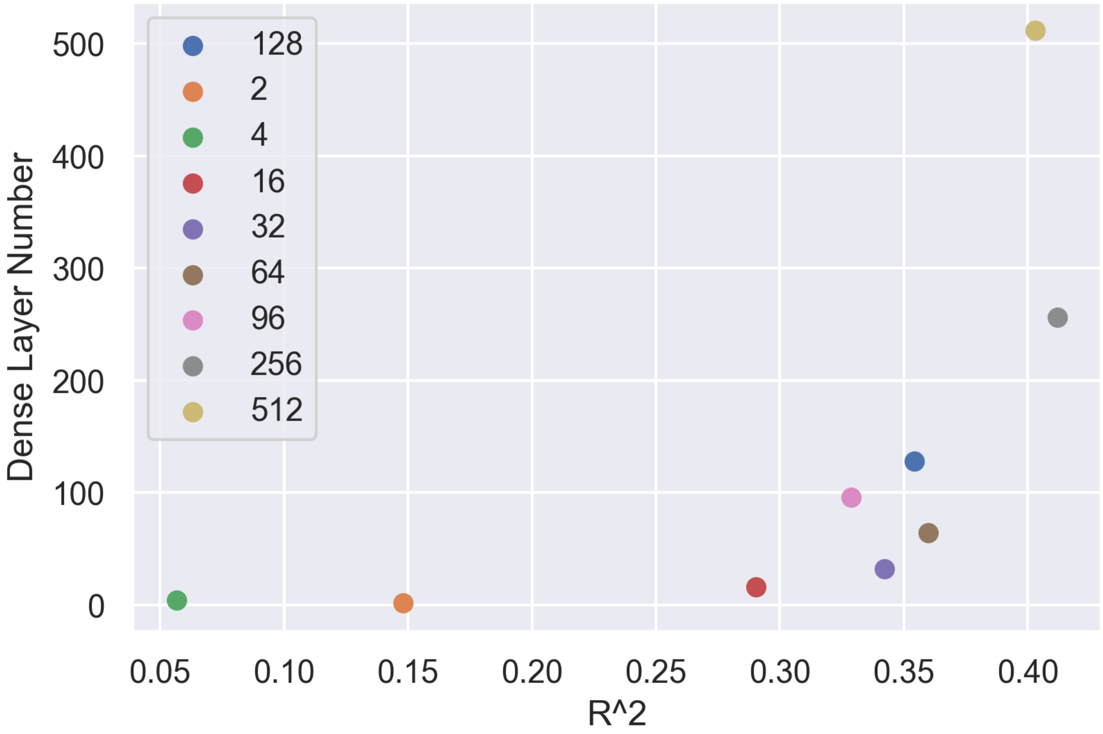


Figure 6.16: R^2 in Different Numbers of Nodes (Higher, The Better)

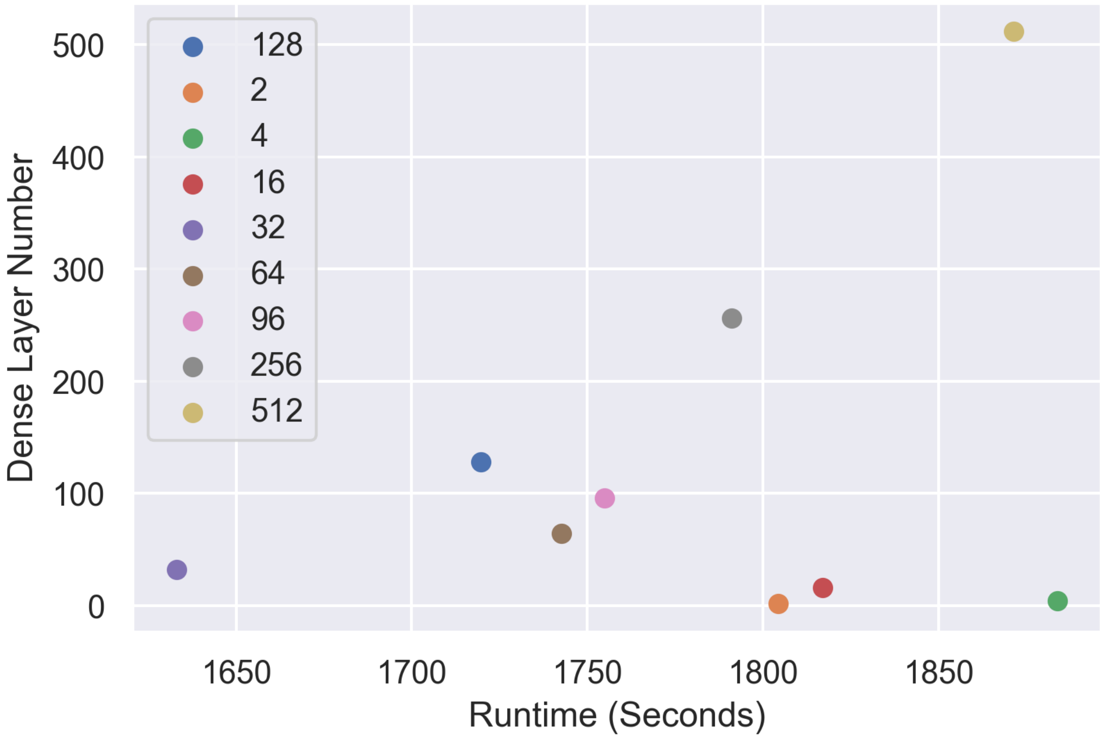


Figure 6.17: Rumtine in Different Numbers of Nodes (Lower, The Better)

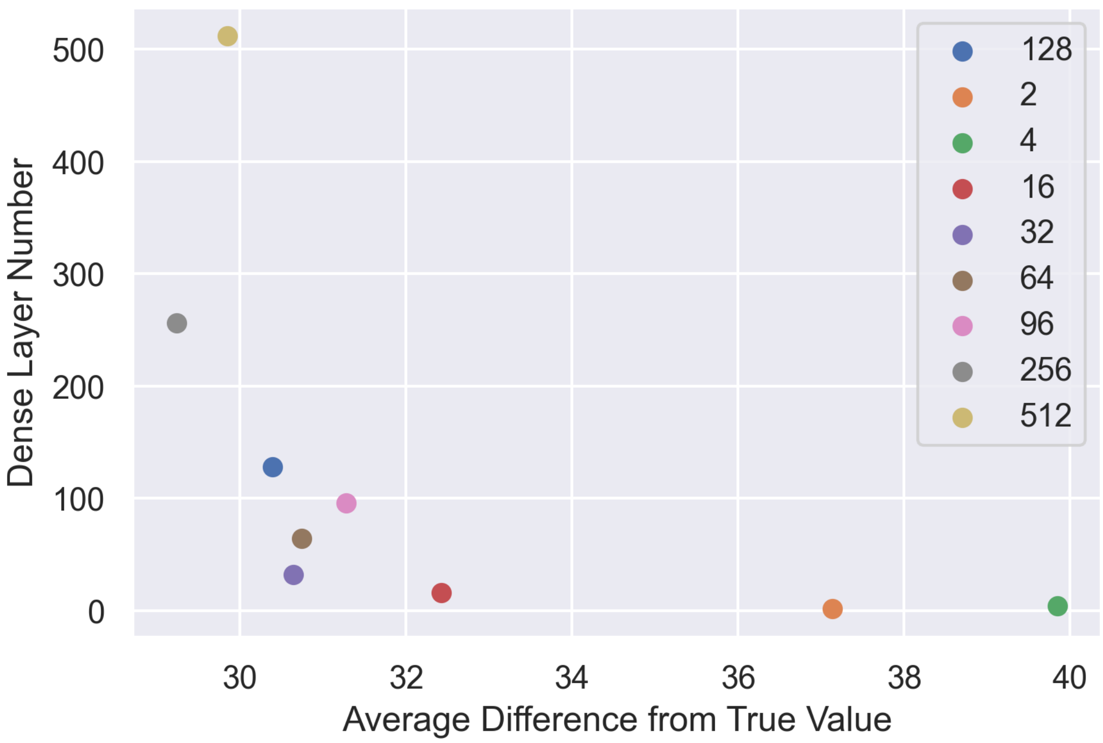


Figure 6.18: Average Difference from True Value in Different Numbers of Nodes (Lower, The Better)

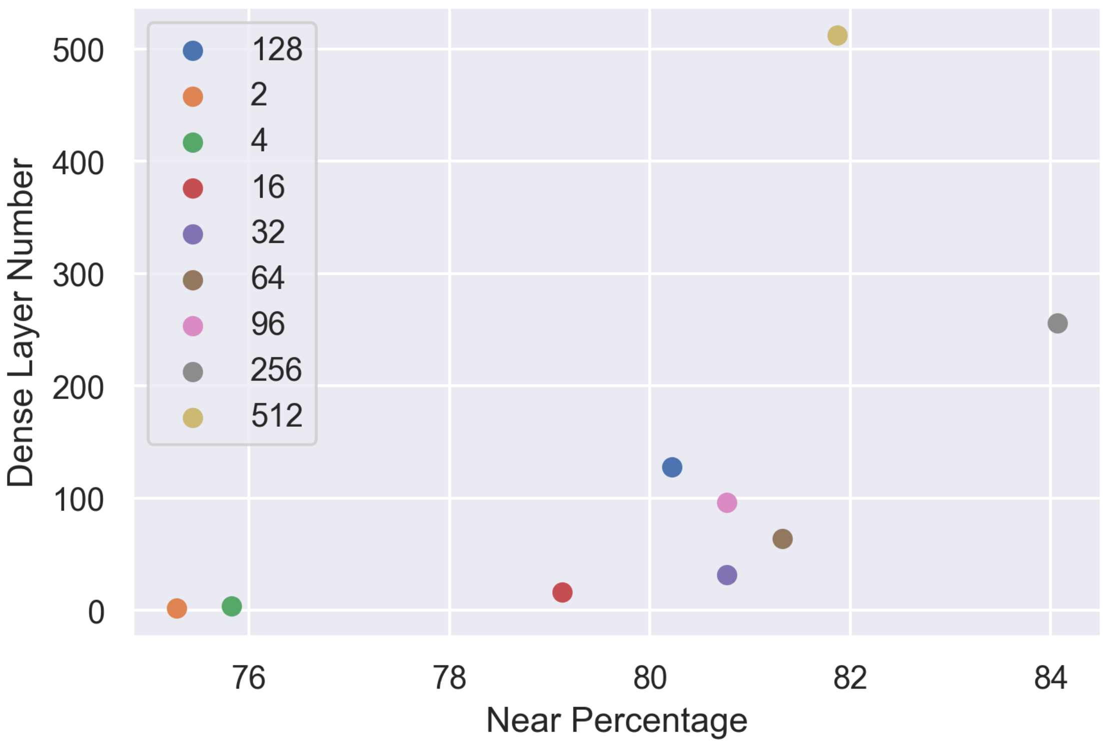


Figure 6.19: Near Percentage in Different Numbers of Nodes (Higher, The Better)

6.1.2 Different Learning Rates

We test our model in six different types of learning rates.

- $1e^{-06}$
- $1e^{-05}$
- 0.0001
- 0.001
- $1e^{-07}$
- $1e^{-08}$

After testing the model and tweaking this, we can see a better test result in $1e^{-05}$, considering MAE, MSE, RMSE, Correlation Coefficient, Average Difference from True Value, Runtime, and R^2. But in terms of Near Percentage, $1e^{-05}$ show the best result.

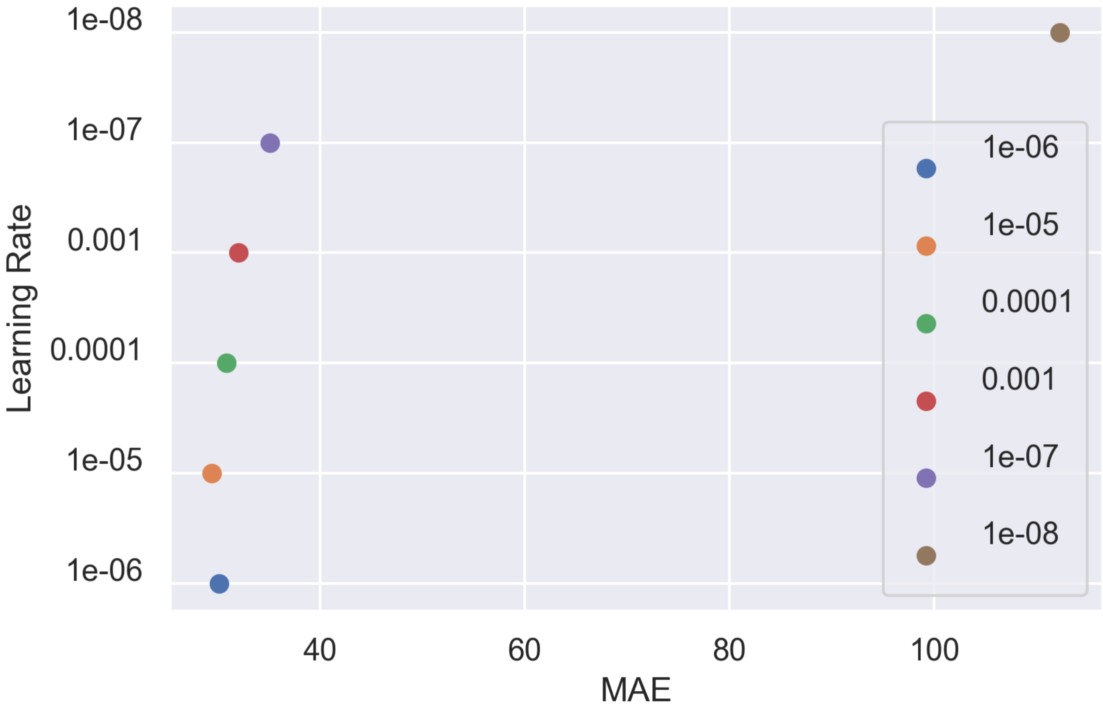


Figure 6.20: MAE in Different Learning Rates (Lower, The Better)

In Figure 6.27, we get the percentage of images can be considered as truly close to actual value since a difference of 50 is considered to be in the same type of class.

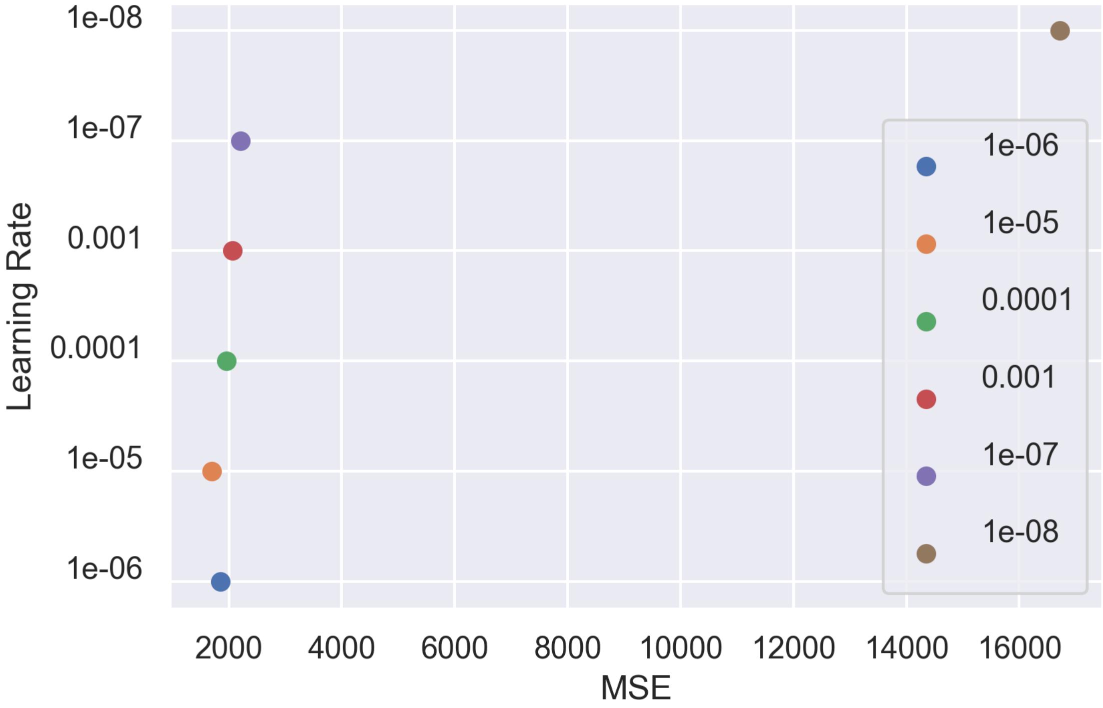


Figure 6.21: MSE in Different Learning Rates (Lower, The Better)

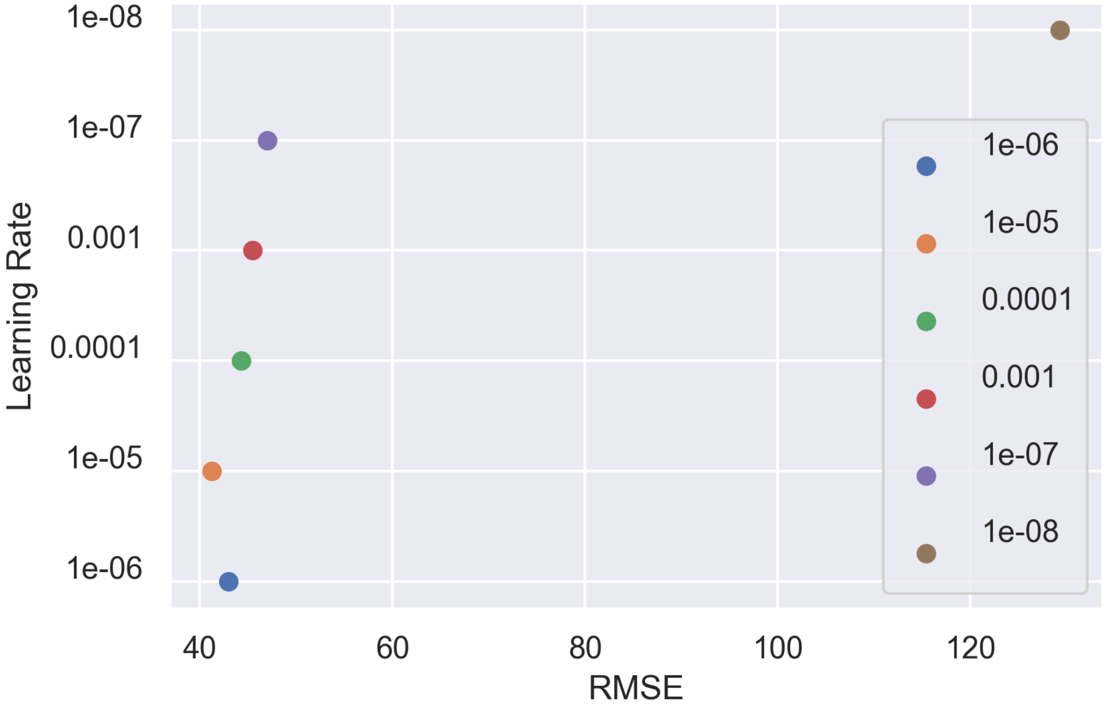


Figure 6.22: RMSE in Different Learning Rates (Lower, The Better)

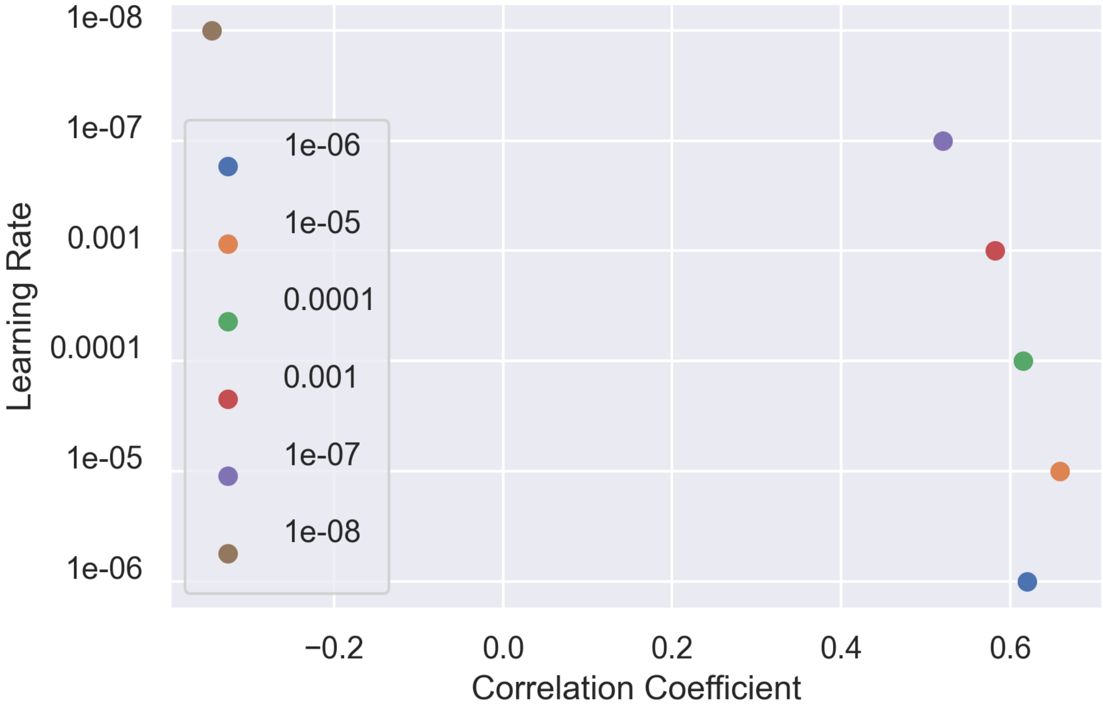


Figure 6.23: Correlation Coefficient in Different Learning Rates (Higher, The Better)

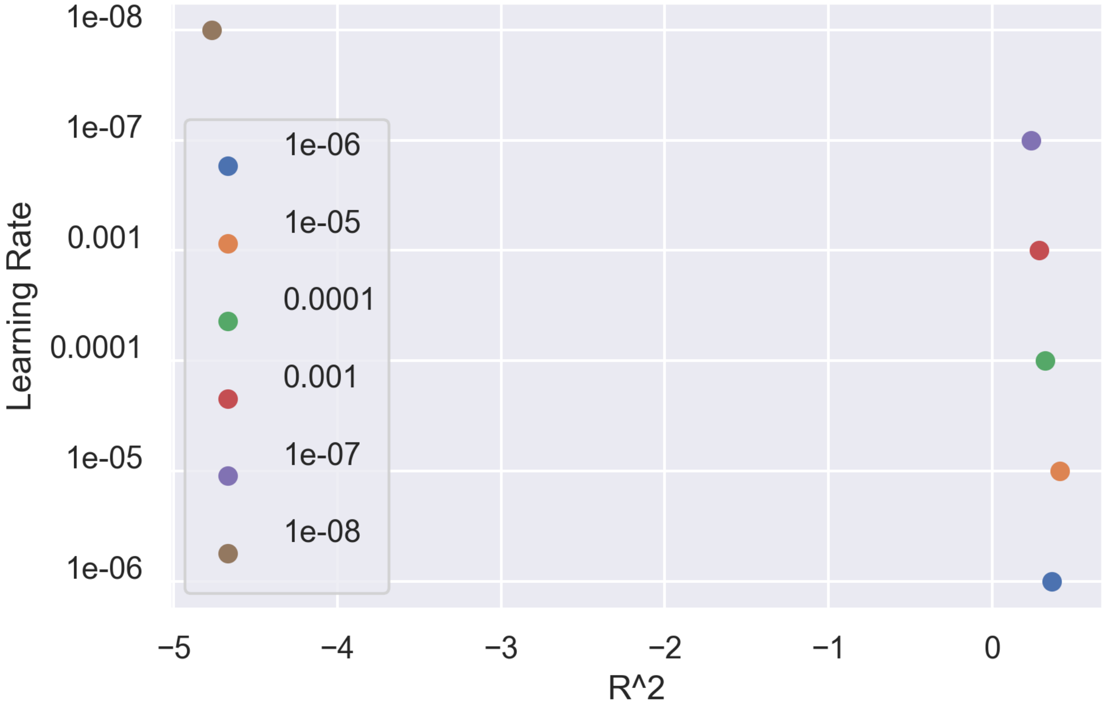


Figure 6.24: R^2 in Different Learning Rates (Higher, The Better)

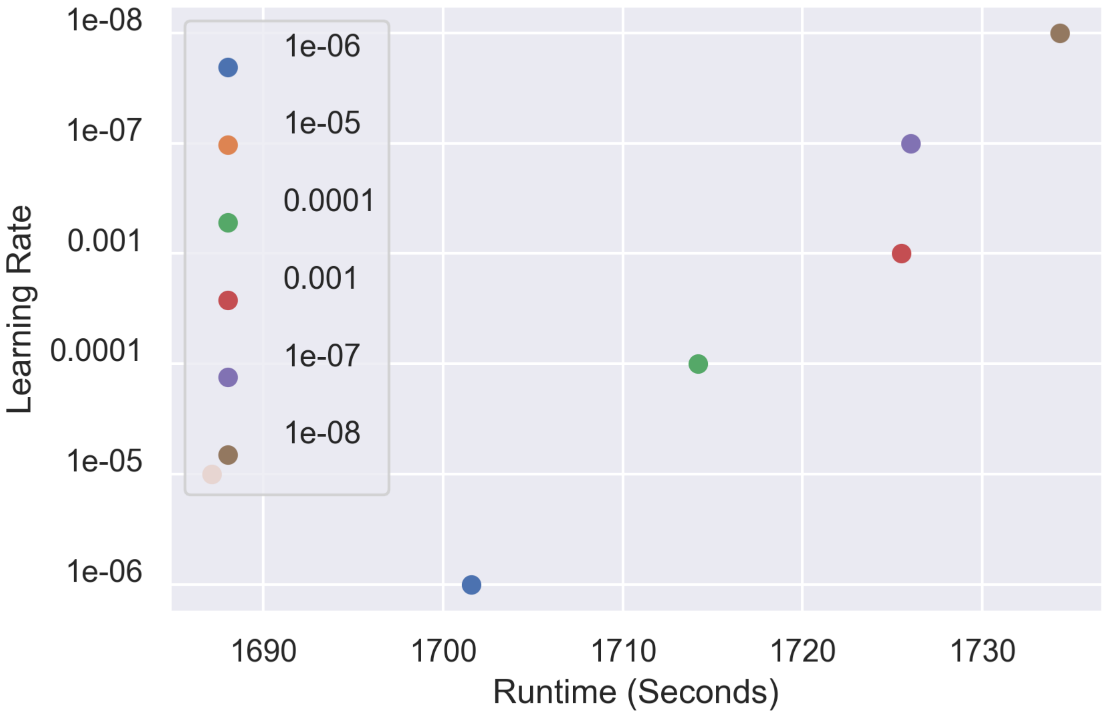


Figure 6.25: Rumtine in Different Learning Rates (Lower, The Better)

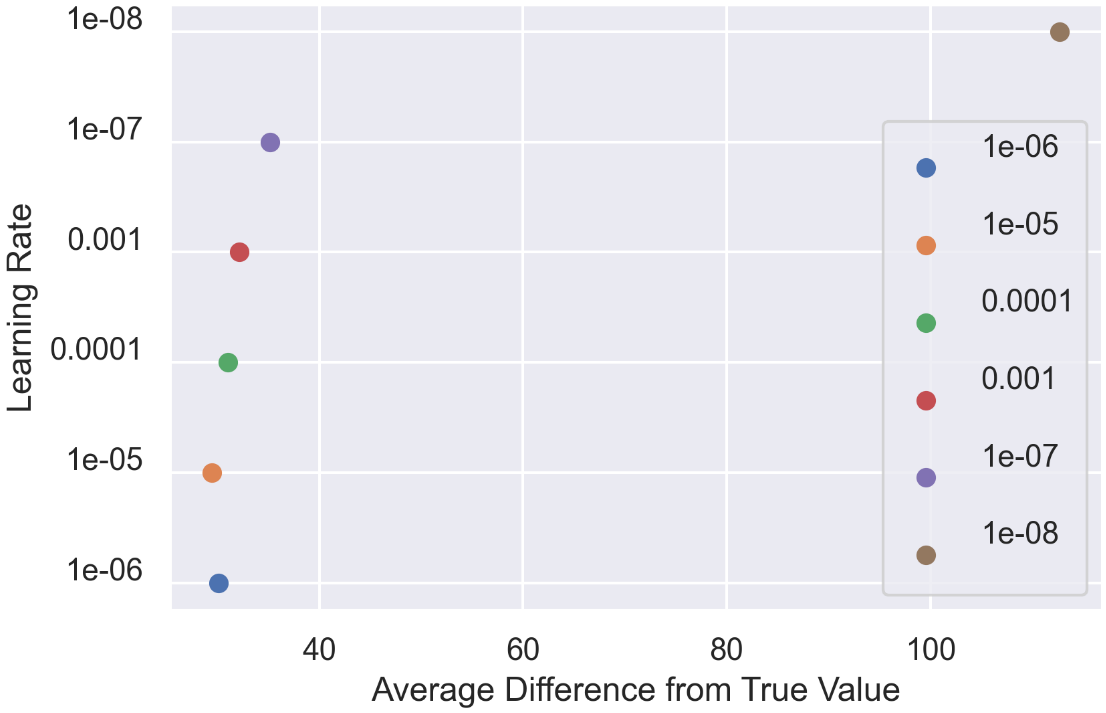


Figure 6.26: Average Difference from True Value in Different Learning Rates (Lower, The Better)

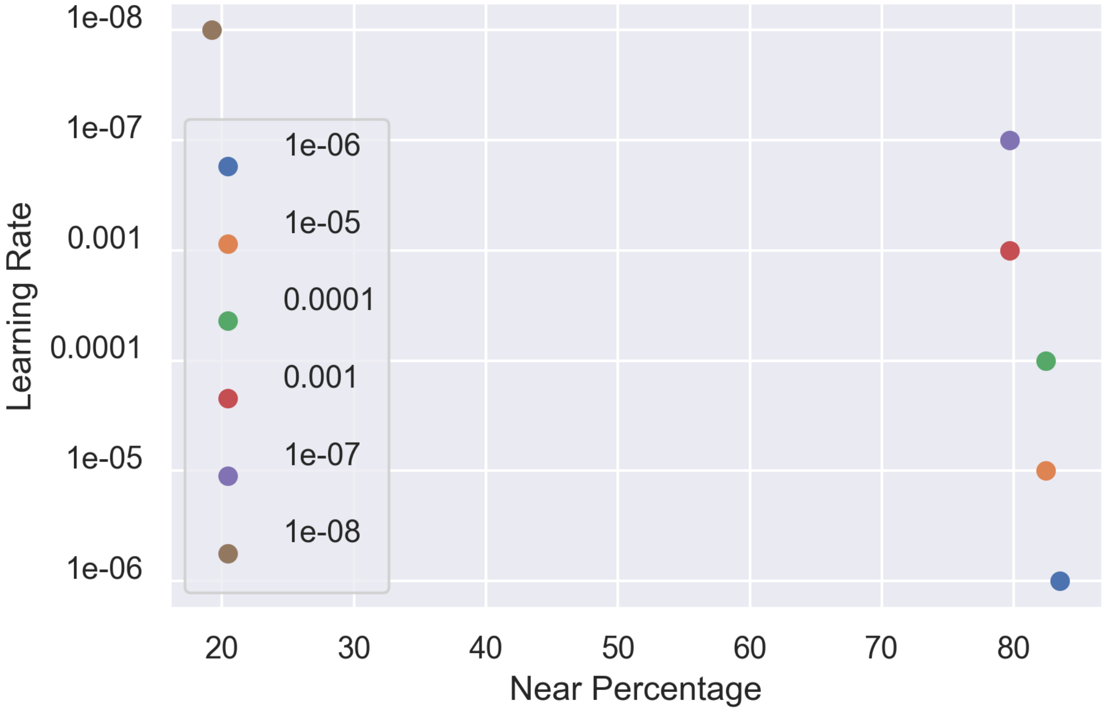


Figure 6.27: Near Percentage in Different Learning Rates (Higher, The Better)

6.1.3 Different Batch Sizes

We test our model in seven different batch sizes.

- 8
- 2
- 4
- 16
- 32
- 48
- 64

After testing the model and tweaking this, we can see a better test result in 16, considering MAE, MSE, RMSE, Correlation Coefficient, Average Difference from True Value, and R^2. But in terms of runtime, 8 shows the best result and when it comes to near percentage, both 4 and 16 perform really close.

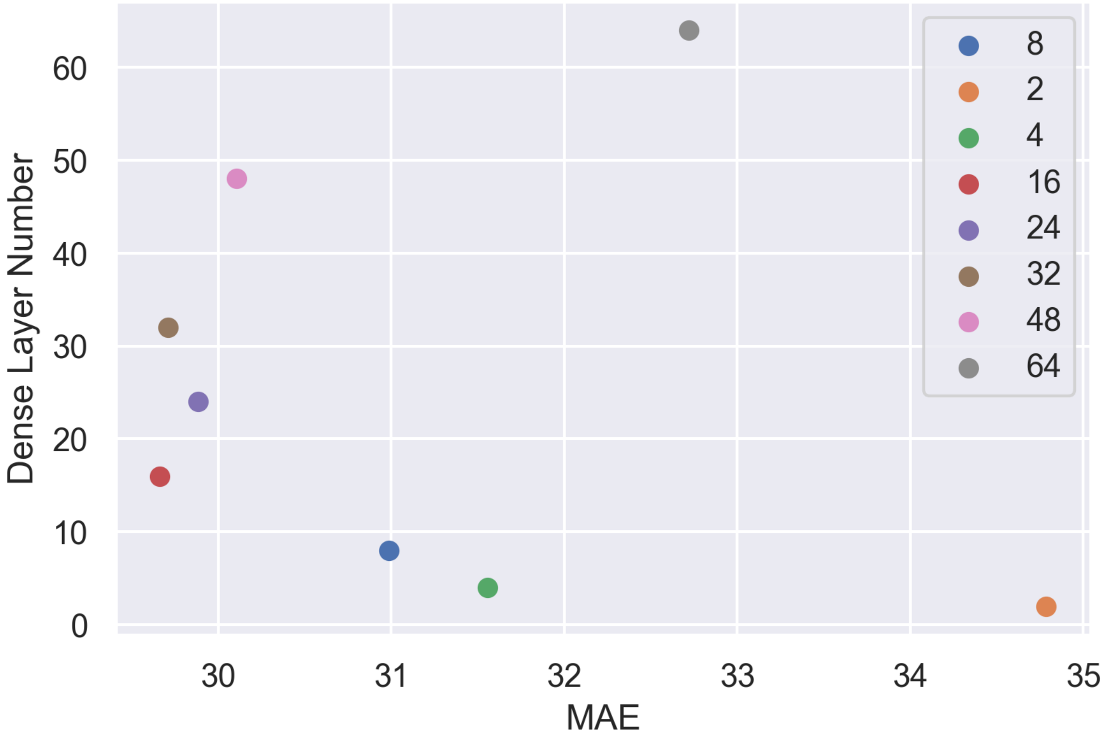


Figure 6.28: MAE in Different Batch Sizes (Lower, The Better)

In Figure 6.35, we get the percentage of images can be considered as truly close to actual value since a difference of 50 is considered to be in the same type of class.

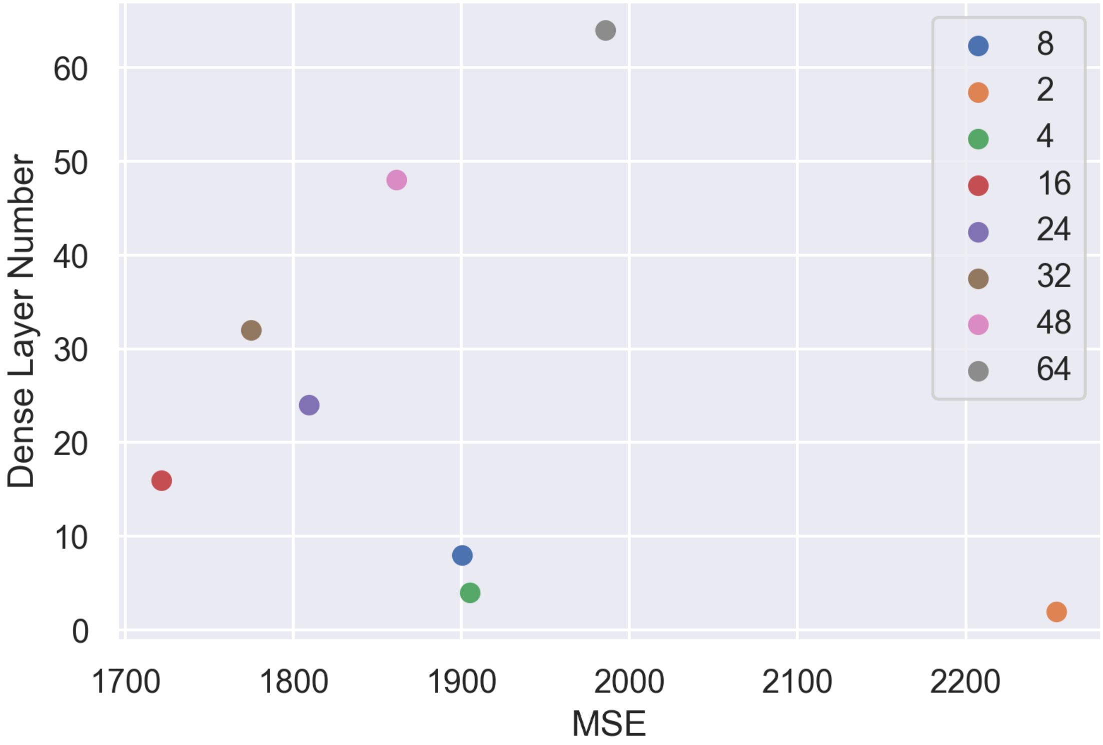


Figure 6.29: MSE in Different Batch Sizes (Lower, The Better)

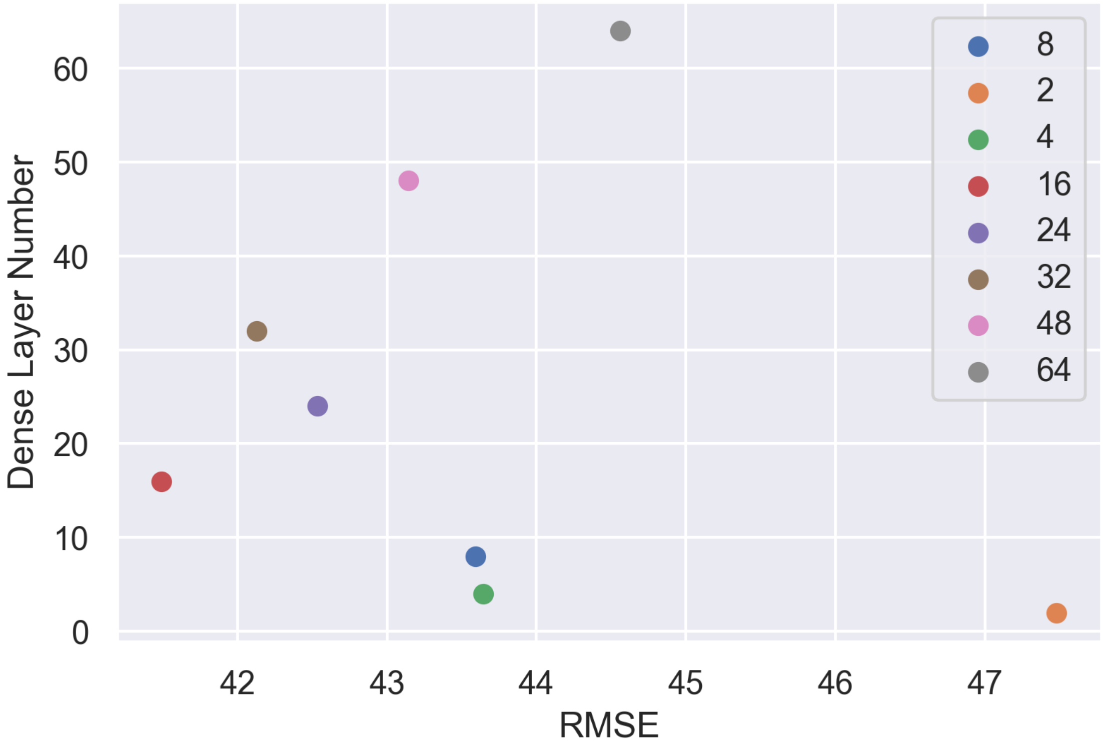


Figure 6.30: RMSE in Different Batch Sizes (Lower, The Better)

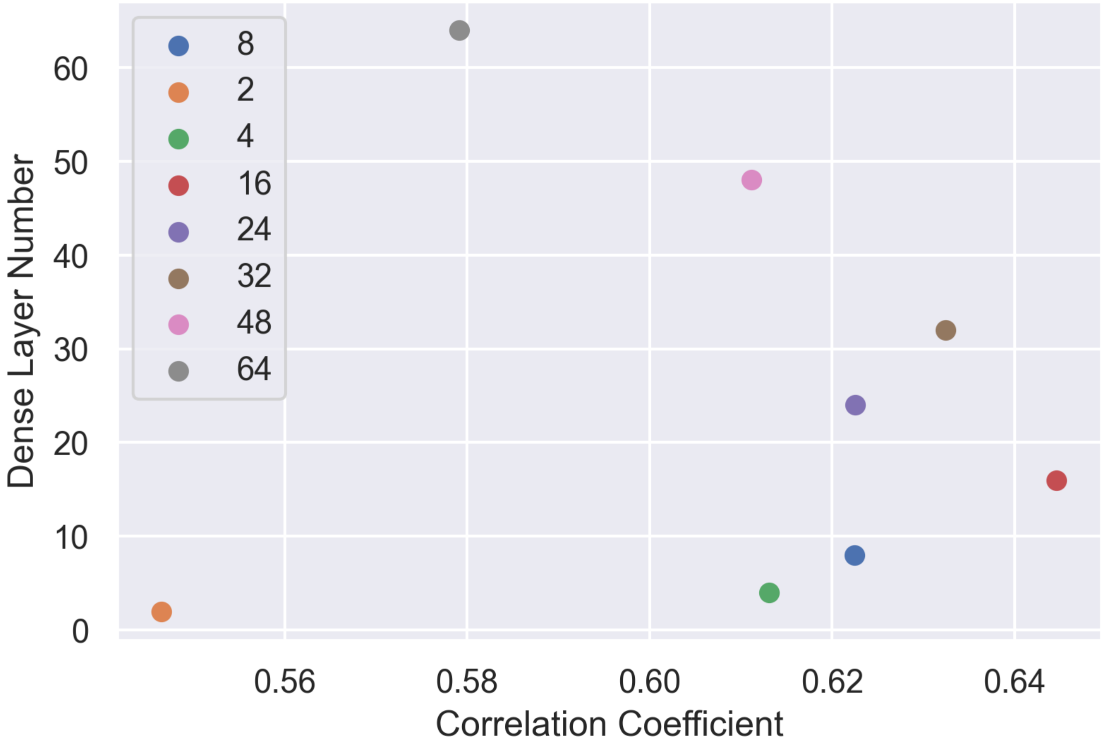


Figure 6.31: Correlation Coefficient in Different Batch Sizes (Higher, The Better)

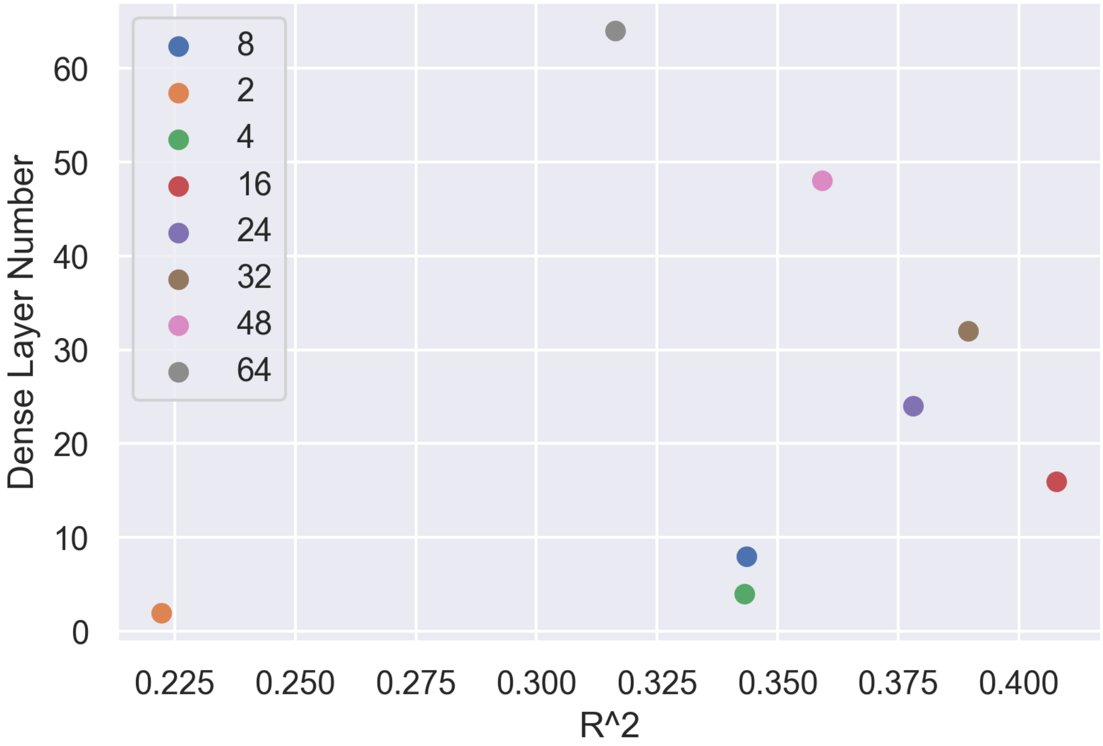


Figure 6.32: R^2 in Different Batch Sizes (Higher, The Better)

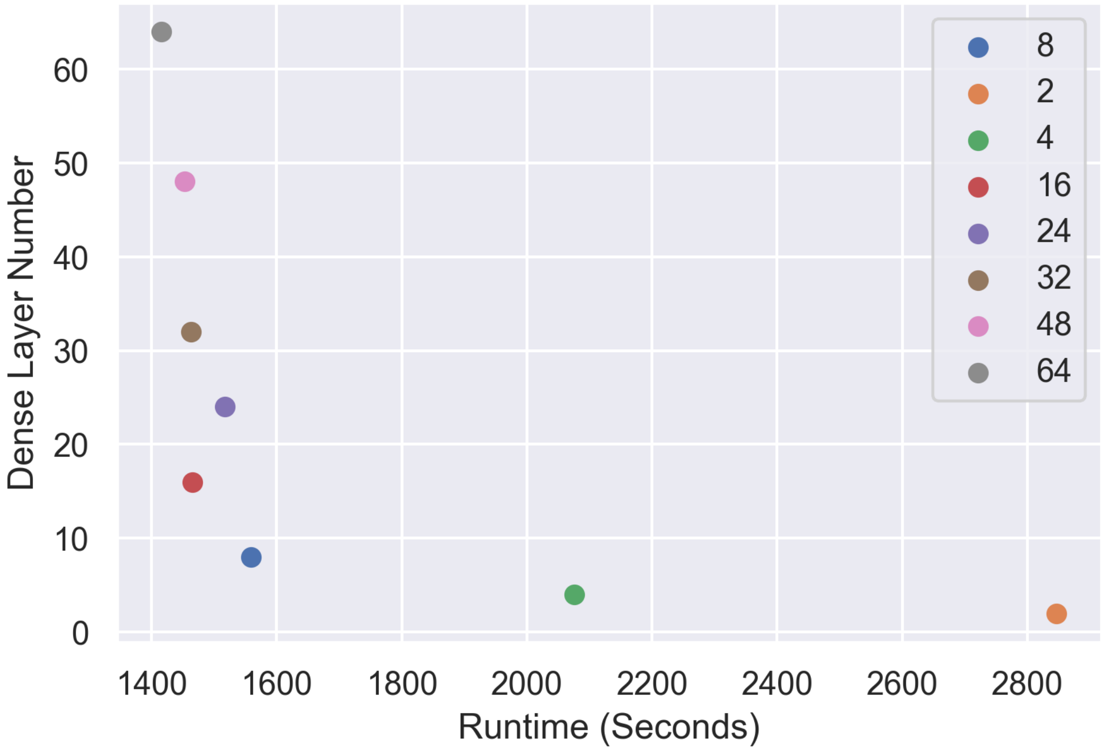


Figure 6.33: Rumtine in Different Batch Sizes (Lower, The Better)

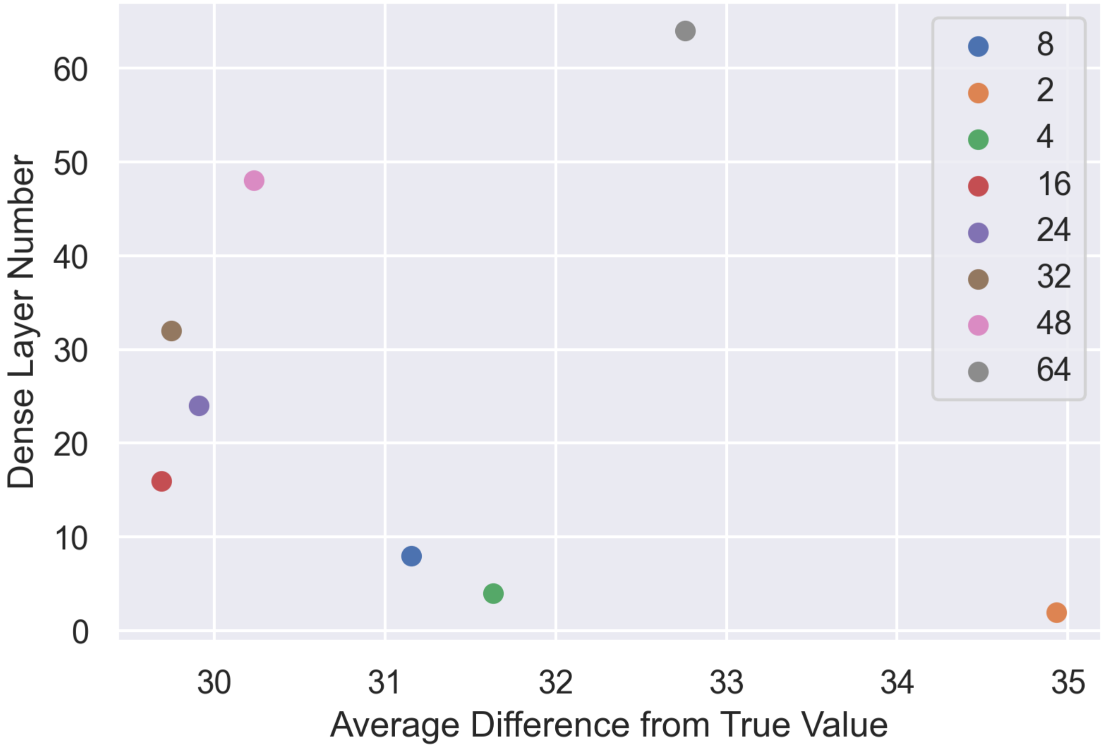


Figure 6.34: Average Difference from True Value in Different Batch Sizes (Lower, The Better)

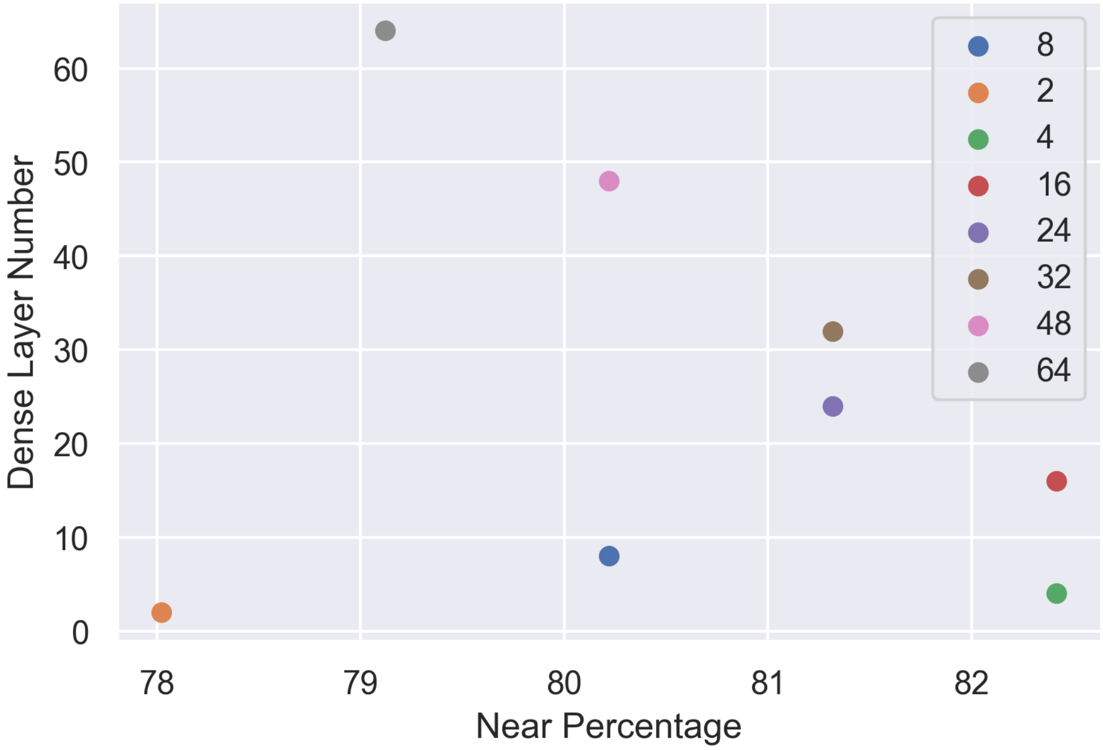


Figure 6.35: Near Percentage in Different Batch Sizes (Higher, The Better)

MSE	RMSE	MAE	R^2
1659.48	40.74	29.63	0.43

Table 6.3: MSE, RMSE, MAE, and R^2 Value for The Changed Model

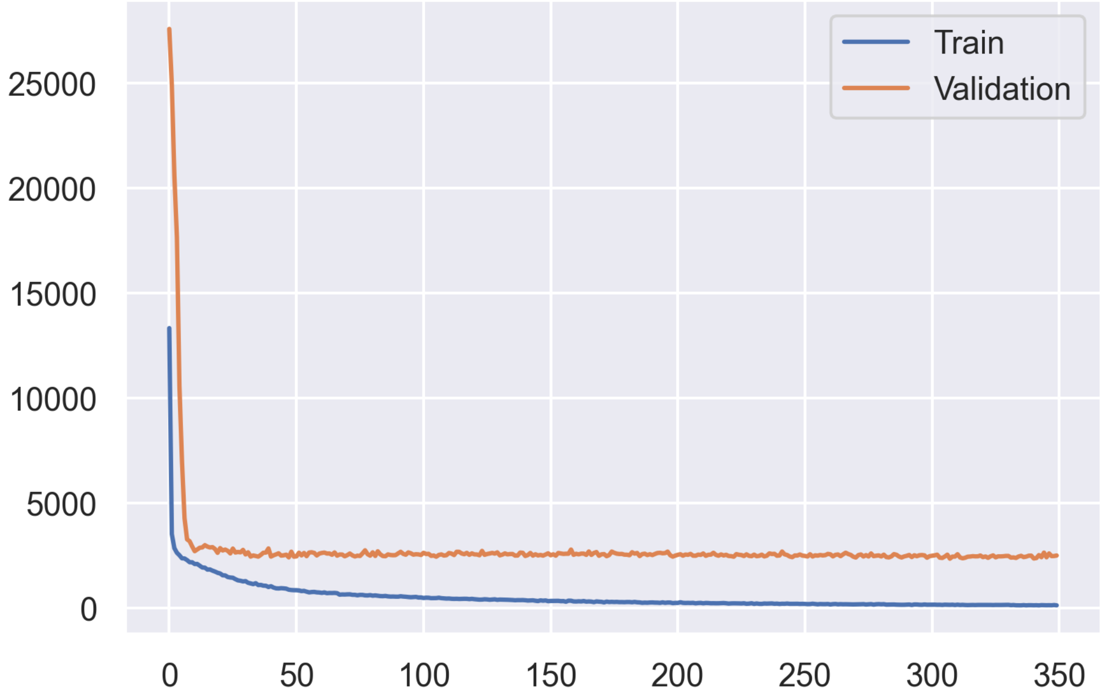


Figure 6.36: MSE Loss Curve of The Tweaked Model

6.2 Post-Implementation of Best Parameters from Baseline Test

Here, we keep the architecture same. But we change the best parameters of learning rates, batch sizes, and number of nodes of the dense layer before the Output Dense Layer.
Learning Rate = 0.00001
Batch Size = 16
Number of Nodes = 32
After running for same epochs, we see that it performs faster and better in terms of numbers but the difference between training loss and validation loss is higher than the proposed model which is a concerning issue and comparing with the proposed model, it is less stable than proposed one.
When observing the resultant data, we see 84% of the predicted values have 50 or fewer PM2.5 value differences than the actual value. we find that the correlation coefficient of the model is 0.6582.

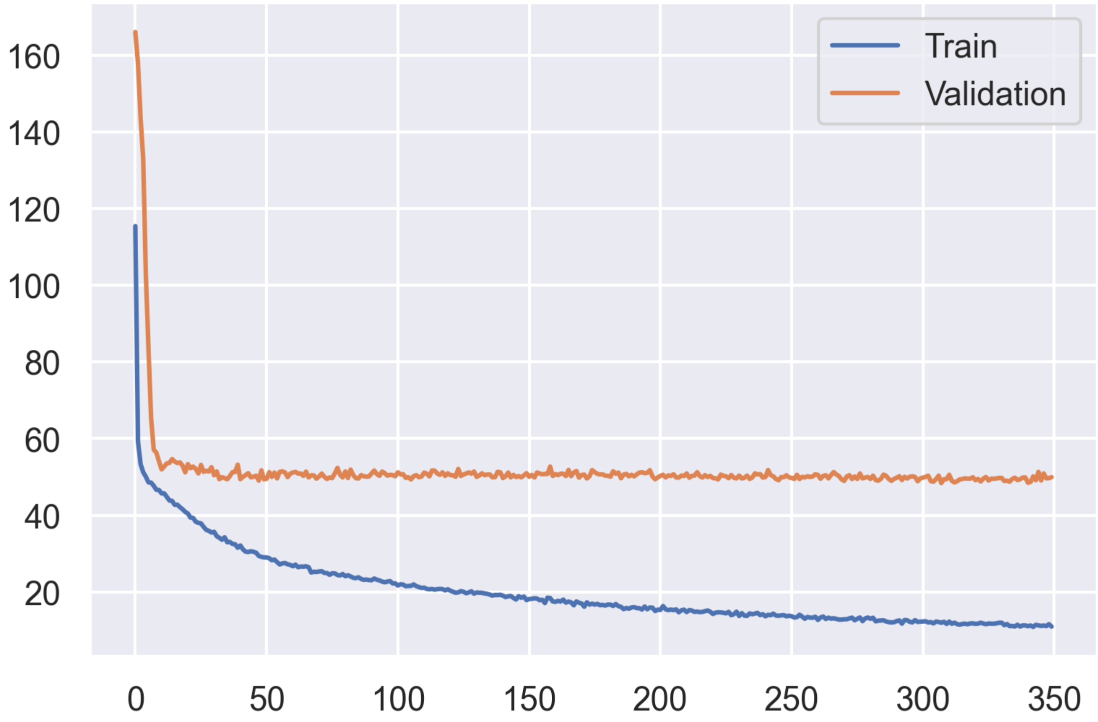


Figure 6.37: RMSE Loss Curve of The Tweaked Model

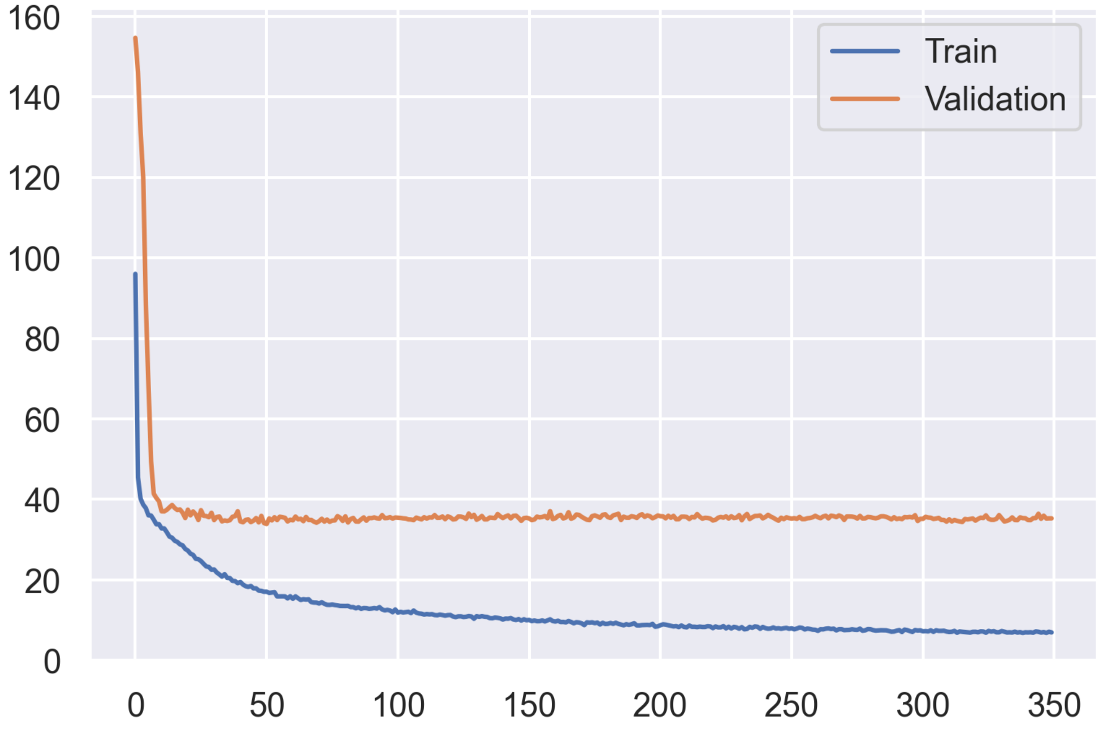


Figure 6.38: MAE Loss Curve of The Tweaked Model

Chapter 7

Discussion

In the preceding sections, we present our results on PPPC utilizing deep learning methods.

Our dataset is less skewed which is a crucial characteristic of a real-world dataset. Our achieved R^2 value is greater than 0.3 which is representative of real-world data. Also, our model, unfortunately, does not work well for noisy pictures. Moreover, it may not provide the expected result if the test image is taken outside Dhaka City since our training data is generated exclusively from the city.

Since we have a lower amount of test data while running the correlation coefficient test, our correlation coefficient value also results lower than expected.

When observing the resultant data, we see 29.69% of average error in between predicted value and ground value. Here, we follow the equation below, where n = number of the test samples of the dataset = 182.

$$\text{Error } = (\sum_{1}^{n} \text{ (Actual-Predicted) })/n \tag{7.1}$$

We also see 81% of the predicted values have 50 or fewer PM2.5 value differences than the actual value. We keep 50 as a threshold for this because if the difference is less than 50 AQI, the impact on the environment can be considered similar.

As we work on other models, we first run them without any kind of pre-trained weights. After we run the models, we discover an intriguing feature of the situation. It has come to our attention that, apart from ResNet-50, every model that we evaluate is unable to even train correctly despite being given the same amount of epochs (350), which leads to inaccurate forecasts. Every model that is put through its test without weight pre-training predicts a value of 0 for every single outcome based on the test dataset.

Unlike larger models, our smaller model with five Conv2D blocks, Batch Normalization and MaxPooling adequately extracts the features for solving our problem. In order to simplify the model and avoid overfitting, we add another dropout layer at the end with a 10% dropout rate. Zero-padding is used to correlate to the data frame's time-limited assumption and more zero-padding results in the denser interpolation of the frequency samples around the unit circle. There is a common misconception that zero padding in the time domain results in greater spectral resolution in the frequency domain. In fact, we can construct networks more easily if we keep the height and width constant and pay less attention to the tensor dimensions when moving from one layer to another. Without padding, the size of the volume would shrink too rapidly. By keeping information at the edges, padding ends up enhancing performance. MaxPooling, in addition to preventing overfitting, reduces the computational cost by reducing the number of parameters to learn. Unlike other examined models that employ Average Pooling and choose smooth features, our model's usage of MaxPooling enables it to predict outcomes with more accuracy. We restrict the number of nodes in the Dense layer because employing a very deep fully-connected layer would be redundant since it would deliver the same performance but with more parameters.

While testing it on attention-based models, DCNN outperforms both ViT and INN. ViT and INN both have different learning approaches. ViT identifies patches and tries to find patterns from them, making it superior for tasks such as object detection or semantic segmentation. But in our problem, the whole image as a whole contains information, not a specific part of it, which makes ViTs [33] unsuitable. Moreover, even though ViTs are being used in regression based tasks recently [20], [38], it does not perform as expected for this specific task. On the other hand, INNs are location-specific and channel-agnostic [37]. While our task at hand is not specific to any color channels, the task of AQI recognition from images is spatially agnostic, since AQI is not computed from any certain location of the image, or should not be offset by the presence of any objects. There is no specific information in any particular segment of the images from which both ViT and INN look for information from. This is why DCNN works better than both ViT and INN. Also, even though this is yet quite unexplored if INN can perform well in regression-based tasks, we take the approach to work on that as an initial work and we can even see that, INN performs the worst out of every models we ran for this task.

Even though ViTs are reported to be relatively more resilient than DCNNs, DCNNs outperform ViT in this problem, as ViTs struggle when they have to extract any indistinct feature from an image. They are also more susceptible to contrast corruption as per Filipuk et al. [41], which is a common source of noise in detecting AQI from images as human-made constructions appear in contrast to the surrounding ambiance. Moreover, Chen et al. [36] report that ViTs are more sensitive to noise and outliers.

Chapter 8

Limitation and Future Work

8.1 Limitations

When we work with the data, we notice that there are certain circumstances where the processing does not work well. For example, when the image has a direct reflection of the sun or when a bright sun is present in the picture, it distorts the picture with excessive brightness, making it harder for the model to predict. In addition, due to the poor picture quality at night, visibility is often inadequate. This experimental model focuses only on daytime air quality monitoring (until evening) and cannot be declared acceptable for nighttime usage. Also, the data of each pixel carries significance toward the result. Therefore, cameras which clicks lower-resolution images may provide some unexpected results since the pictures may contain noise. Noisy data in this case can heavily impact a negative outcome. Again, since every photo has been clicked using smartphones and many smartphones nowadays focus on computing the approximate image rather than taking the absolute photons, it may change some pixels which can impact toward wrong results [45]. Furthermore, images taken through smartphones are mostly focused to be "social-media ready", to be able to be uploaded to social media as soon as the image gets clicked; without any sort of editing, which focuses on changing the brightness and contrast of the actual scene of that area which also can be negatively impactful.

8.2 Future Work

Our future work plan includes converting the model to be used as a transfer learning model and implementing the model as a mobile app since our work totally focuses to be implemented in real life. Moreover, we will continue collecting more photos along with the PM2.5 of that exact location using AQI detection sensors from various contexts (e.g., cities, rural regions) from different time ranges and different state-of-the-art models will be tested to get improved results. Also, some comparisons depending on different weathers can be done. Furthermore, we will focus on how images can be inputted better into Deep Learning Model to solve this task.

Chapter 9

Conclusion

It is a complex field of research due to the fact that it requires knowledge of the changes in the visibility of various objects of varying depths. Assessing the PM2.5 index based on an image requires such a degree of expertise. This, in turn, calls for a significant level of knowledge on the subject matter of the picture. By using a Deep Convolutional Neural Network (CNN) model, we develop a method that is capable of forecasting PM2.5 concentrations from pictures as part of the scope of this project. The images are shot in such a way that the sky can be seen in each and every one of them, constituting at least fifty percent of the whole image. After that, the images are used as the input, and the bottom fifty percent of the picture is cropped in order to remove the component of the picture that does not feature the sky. After that, the improved image is loaded into the representation that was advised. We conclude that the saturation of a picture can have a substantial influence on the accuracy with which one may detect the PM2.5 concentration from an image as a consequence of our observations and analysis. It is one of the main reasons why we come to this conclusion. Therefore, on the basis of this information, we run CNN architecture in supervised mode, with the model obtaining a picture as well as the AQI over the time period during which the image is taken. This allows us to better understand the relationship between the two. We examine the recommended method using a dataset we have built expressly for this purpose. The findings reveal that CNN can be used to predict PM2.5 from photos and enhance estimate accuracy. This is accomplished by increasing the number of observations included in the model. In addition to the process that is being discussed, we are also building an image collection of PM2.5 concentrations. There are now 1,818 images in this collection, and each of those photographs has a PM2.5 value corresponding to it. Also, after running our dataset in different architectures, including state-of-the-art approaches, we see that Deep CNN outperforms all of them.

Bibliography

[1] E. J. McCartney, "Optics of the atmosphere: Scattering by molecules and particles," *New York*, 1976.

[2] W. C. Malm, K. K. Leiker, and J. V. Molenar, "Human perception of visual air quality," *Journal of the Air Pollution Control Association*, vol. 30, no. 2, pp. 122–131, 1980.

[3] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International journal of computer vision*, vol. 48, no. 3, pp. 233–254, 2002.

[4] R. D. Brook, B. Franklin, W. Cascio, *et al.*, "Air pollution and cardiovascular disease: A statement for healthcare professionals from the expert panel on population and prevention science of the american heart association," en, *Circulation*, vol. 109, no. 21, pp. 2655–2671, 2004.

[5] C. A. Olman, K. Ugurbil, P. Schrater, and D. Kersten, "Bold fmri and psychophysical measurements of contrast response to broadband images," *Vision research*, vol. 44, no. 7, pp. 669–683, 2004.

[6] R. Fattal, "Single image dehazing," *ACM transactions on graphics (TOG)*, vol. 27, no. 3, pp. 1–9, 2008.

[7] P. Carr and R. Hartley, "Improved single image dehazing using geometry," in *2009 Digital Image Computing: Techniques and Applications*, 2009, pp. 103–110. DOI: 10.1109/DICTA.2009.25.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 1956–1963.

[9] W. Huang, J. Tan, H. Kan, *et al.*, "Visibility, air quality and daily mortality in shanghai, china," *Science of the Total Environment*, vol. 407, no. 10, pp. 3295–3300, 2009.

[10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intel. Syst. Technol*, vol. 2, no. 3, 2011.

[11] Ministry of Environment and Forests, Bangladesh, *About air quality index (AQI)*, en, http://case.doe.gov.bd/index.php?option=com_content&view=article&id=9&Itemid=31, Accessed: 2022-4-14, 2011.

[12] R. Fattal, "Dehazing using color-lines," *ACM transactions on graphics (TOG)*, vol. 34, no. 1, pp. 1–14, 2014.

[13] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2014. DOI: 10.48550/ARXIV.1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556.

[14] C. Szegedy, W. Liu, Y. Jia, *et al.*, *Going deeper with convolutions*, 2014. DOI: 10.48550/ARXIV.1409.4842. [Online]. Available: https://arxiv.org/abs/1409.4842.

[15] H. Wang, X. Yuan, X. Wang, Y. Zhang, and Q. Dai, "Real-time air quality estimation based on color image processing," in *2014 IEEE Visual Communications and Image Processing Conference*, 2014, pp. 326–329. DOI: 10.1109/VCIP.2014.7051572.

[16] Y. Li, J. Huang, and J. Luo, "Using user generated online photos to estimate and monitor air pollution in major cities," in *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service - ICIMCS '15*, ACM Press, 2015. DOI: 10.1145/2808492.2808564. [Online]. Available: https://doi.org/10.1145%2F2808492.2808564.

[17] M. Suga and A. Runa, *Cities - united nations sustainable development action 2015*, en, https://www.un.org/sustainabledevelopment/cities/, Accessed: 2023-1-7, Jan. 2015.

[18] C. Liu, F. Tsow, Y. Zou, and N. Tao, "Particle pollution estimation based on image analysis," en, *PLoS One*, vol. 11, no. 2, e0145955, 2016. DOI: \url{https://doi.org/10.1371/journal.pone.0145955}.

[19] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural fields," en, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2024–2039, 2016.

[20] J. Carreira and A. Zisserman, *Quo vadis, action recognition? a new model and the kinetics dataset*, 2017. DOI: 10.48550/ARXIV.1705.07750. [Online]. Available: https://arxiv.org/abs/1705.07750.

[21] A. Chakma, B. Vizena, T. Cao, J. Lin, and J. Zhang, "Image-based air quality analysis using deep convolutional neural network," in *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2017, pp. 3949–3952.

[22] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," *Advances in neural information processing systems (NeurIPS)*, vol. 30, 2017.

[23] Q. Bo, W. Yang, N. Rijal, Y. Xie, J. Feng, and J. Zhang, "Particle pollution estimation from images using convolutional neural network and weather features," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 3433–3437.

[24] J. Ma, K. Li, Y. Han, and J. Yang, "Image-based air pollution estimation using hybrid convolutional neural network," in *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, 2018, pp. 471–476.

[25] N. Rijal, R. T. Gutta, T. Cao, J. Lin, Q. Bo, and J. Zhang, "Ensemble of deep neural networks for estimating particulate matter from images," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, IEEE, 2018, pp. 733–738.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2018. DOI: 10.48550/ARXIV.1801.04381. [Online]. Available: https://arxiv.org/abs/1801.04381.

[27] M. S. B. Auvee, *Air pollution monitoring system based on air pollutant index(api) and geographic information system(gis)*, http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/12290/17341043_CSE.pdf, Accessed: 2022-3-29, 2019.

[28] K. Gu, J. Qiao, and X. Li, "Highly efficient picture-based prediction of PM2.5 concentration," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3176–3184, 2019.

[29] M. M. Hossain and Work for a Better Bangladesh Trust (WBBT), Bangladesh, "Study on ambient particulate matter (PM2.5) with different mode of transportation in dhaka city, bangladesh," *American Journal of Pure and Applied Biosciences*, vol. 1, no. 4, pp. 12–19, 2019.

[30] M. M. Samsami, N. Shojaee, S. Savar, and M. Yazdi, "Classification of the air quality level based on analysis of the sky images," in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, IEEE, 2019.

[31] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2019. DOI: 10.48550/ARXIV.1905.11946. [Online]. Available: https://arxiv.org/abs/1905.11946.

[32] Z. Wang, W. Zheng, C. Song, *et al.*, "Air quality measurement based on double-channel convolutional neural network ensemble learning," *IEEE Access*, vol. 7, pp. 145 067–145 081, 2019.

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2020. DOI: 10.48550/ARXIV.2010.11929. [Online]. Available: https://arxiv.org/abs/2010.11929.

[34] S. Song, J. C. K. Lam, Y. Han, and V. O. K. Li, "ResNet-LSTM for Real-Time PM2.5 and PM estimation using sequential smartphone images," *IEEE Access*, vol. 8, pp. 220 069–220 082, 2020.

[35] Q. Zhang, F. Fu, and R. Tian, "A deep learning and image-based model for air quality estimation," *Science of The Total Environment*, vol. 724, p. 138 178, 2020, ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2020.138178. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720316910.

[36] G. Chen, M. Wang, Y. Yue, Q. Zhang, and L. Yuan, *Full transformer framework for robust point cloud registration with deep information interaction*, 2021. DOI: 10.48550/ARXIV.2112.09385. [Online]. Available: https://arxiv.org/abs/2112.09385.

[37] D. Li, J. Hu, C. Wang, *et al.*, *Involution: Inverting the inherence of convolution for visual recognition*, 2021. DOI: 10.48550/ARXIV.2103.06255. [Online]. Available: https://arxiv.org/abs/2103.06255.

[38] W. Mao, Y. Ge, C. Shen, Z. Tian, X. Wang, and Z. Wang, *Tfpose: Direct human pose estimation with transformers*, 2021. DOI: 10.48550/ARXIV.2103.15320. [Online]. Available: https://arxiv.org/abs/2103.15320.

[39] R. Saha, S. N. M. A. Hoque, M. M. R. Manu, and A. Hoque, "Monitoring air quality of dhaka using IoT: Effects of COVID-19," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, IEEE, 2021.

[40] S. K. Sarkar and M. M. H. Khan, "Impact of COVID-19 on PM2.5 pollution in fastest-growing megacity dhaka, bangladesh," en, *Disaster Med. Public Health Prep.*, pp. 1–4, 2021.

[41] M. Filipiuk and V. Singh, "Comparing vision transformers and convolutional nets for safety critical systems.," in *SafeAI@ AAAI*, 2022.

[42] T. D. Star, *2021 world air quality report: Bangladesh remains most polluted*, https://www.thedailystar.net/environment/pollution/air-pollution/news/2021-world-air-quality-report-bangladesh-most-polluted-2987961, Accessed: 2022-3-29, Mar. 2022.

[43] T. D. Star, *Dhaka again ranks world's most polluted city*, https://www.thedailystar.net/environment/pollution/air-pollution/news/dhaka-again-ranks-worlds-most-polluted-city-2976026, Accessed: 2022-3-29, Mar. 2022.

[44] Q. Zhang, L. Tian, F. Fu, H. Wu, W. Wei, and X. Liu, "Real-time and image-based aqi estimation based on deep learning," *Advanced Theory and Simulations*, p. 2 100 628, 2022. DOI: https://doi.org/10.1002/adts.202100628.

[45] M. Brownlee, *What is happening with iphone camera?* Jan. 2023. [Online]. Available: https://youtu.be/88kd9tVwkH8.

[46] *Sdg goal 11*, en, https://sdgs.un.org/goals/goal11, Accessed: 2023-1-7, 2023.

[47] *Sdg goal 13*, en, https://sdgs.un.org/goals/goal13, Accessed: 2023-1-7, 2023.

[48] *Air cognizer: Predicting air quality with TensorFlow lite*, https://blog.tensorflow.org/2019/02/air-cognizer-predicting-air-quality.html, Accessed: 2020-1-22.

[49] *MaxPooling2D layer*, en, https://keras.io/api/layers/pooling_layers/max_pooling2d/, Accessed: 2022-9-18.

[50] *Zero padding in convolutional neural networks explained*, en, https://deeplizard.com/learn/video/qSTv_m-KFk0, Accessed: 2022-9-18.