

8-2022

Could Cultures Determine the Course of Epidemics and Explain Waves of COVID-19?

Md Salman Rahman
The University of Texas Rio Grande Valley

Follow this and additional works at: <https://scholarworks.utrgv.edu/etd>



Part of the [Applied Mathematics Commons](#), and the [Public Health Commons](#)

Recommended Citation

Rahman, Md Salman, "Could Cultures Determine the Course of Epidemics and Explain Waves of COVID-19?" (2022). *Theses and Dissertations*. 1089.
<https://scholarworks.utrgv.edu/etd/1089>

This Thesis is brought to you for free and open access by ScholarWorks @ UTRGV. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

COULD CULTURES DETERMINE THE COURSE OF EPIDEMICS AND EXPLAIN WAVES
OF COVID-19?

A Thesis

by

MD SALMAN RAHMAN

Submitted in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

Major Subject: Applied Statistics and Data Science

The University of Texas Rio Grande Valley

August 2022

COULD CULTURES DETERMINE THE COURSE OF EPIDEMICS AND EXPLAIN WAVES
OF COVID-19?

A Thesis
by
MD SALMAN RAHMAN

COMMITTEE MEMBERS

Dr. Tamer Oraby
Chair of Committee

Dr. Zhijun Qiao
Committee Member

Dr. Santanu Chakraborty
Committee Member

Dr. Hansapani Rodrigo
Committee Member

August 2022

Copyright 2022 Md Salman Rahman

All Rights Reserved

ABSTRACT

Rahman, Md Salman, Could Cultures Determine the Course of Epidemics and Explain Waves of COVID-19?. Master of Science (MS), August, 2022, 85 pp., 6 tables, 54 figures, references, 55 titles.

Coronavirus Disease (COVID-19), caused by the SARS-CoV-2 virus, is an infectious disease that quickly became a pandemic spreading with different patterns in each country. Travel bans, lockdowns, social distancing, and non-essential business closures caused significant economic disruptions and stalled growth worldwide in the pandemic's first year. In almost every country, public health officials forced and/or encouraged Nonpharmaceutical Interventions (NPIs) such as contact tracing, social distancing, masks, and quarantine. Human behavioral decision-making regarding social isolation significantly impedes global success in containing the pandemic. This thesis focuses on human behaviors and cultures related to the decision-making of social isolation during the pandemic. Within a COVID-19 disease transmission model, we created a conceptual and deterministic model of human behavior and cultures. This study emphasizes the importance of human behavior in successful disease control strategies. Additionally, we introduce a back engineering approach to determine whether cultures are explained by the courses of COVID-19 epidemics. We used a deep learning technique based on a convolutional neural network (CNN) to predict cultures from COVID-19 courses. In this system, CNN is used for deep feature extraction with ordinary convolution and with residual blocks. Also, a novel concept is introduced that converts tabular data into an image using matrix transformation and image processing validated by identifying some well-known function. Despite having a small and novel data set, we have achieved an 80-95% accuracy, depending on the cultural measures.

DEDICATION

Dedicated to my parents.

ACKNOWLEDGMENTS

First, I thank Almighty Allah who provided me with everything. Secondly, I want to thank Dr. Tamer Oraby, who has been a fantastic mentor to me during my time at UTRGV. He is simply outstanding, and his encouragement and support kept me going while I worked on this thesis. He is not only a brilliant individual with whom I am fortunate to have the opportunity to work but also a good human being. He was always there to respond to my text message in less than five minutes whenever I was stuck in a problem. He encouraged me to conduct quality research and taught me to learn by doing.

I would also like to thank the members of my committee, Dr. Zhijun Qiao, Dr. Santanu Chakraborty, and Dr. Hansapani Rodrigo. Dr. Qiao was the one who encouraged me to join UTRGV, and he was always available to assist me when I needed it. Dr. Rodrigo is exceptional and supportive, and I want to thank her for her help, kindness, and encouragement throughout my research. Dr. Chakraborty has also been a constructive and supportive committee member.

I would also like to thank my supervisor, Dr. George Yanev, who has been an excellent mentor to me throughout my time at UTRGV. He always believed in me and what I could accomplish and always made time to talk with me. He was very encouraging and concerned about me.

Many thanks go to Dr. Timothy Huber, the dean Dr. Vivian Incera, and Ms. Brenda Cowart (from the scholarship office) for their kindness and support. They were always by my side whenever I need assistance. They will always be an essential part of my graduate studies.

I was fortunate to have Prosanta Barai as my friend and flatmate; he cooked for me, and I can't imagine my graduate journey without him. Also, I am grateful to work with Harrinson Arrubla on mathematics and convolutional neural network.

Finally, I am grateful to my parents and sister, without whom I would not have had the wonderful opportunities I have had at UTRGV. They always believed in me, and their encouragement

helped me get through difficult times during my graduate program. Thank you for your love and encouragement.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	iv
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER I. INTRODUCTION	1
1.1 Compartmental Models of Infectious Disease	1
1.1.1 Susceptible-Infectious-Recovered (SIR) Model	2
1.1.2 Susceptible-Exposed-Infectious-Recovered (SEIR) Model	4
1.2 Bayesian Method	6
1.2.1 Bayes Theorem	6
1.2.2 Selection of Prior Distribution	7
1.2.3 Likelihood Function	8
1.2.4 Bayesian Statistical Inference	9
1.2.5 Markov chain Monte Carlo (MCMC)	11
1.2.6 MCMC Convergence Diagnostics	11
1.3 Step by Step Implementation of Bayesian Inference with R	13
1.3.1 Specifying the Ordinary Differential Equation (ODE) Model	14
1.3.2 ODE Model Implementation	14
1.3.3 Parameter Declaration and Markov chain Monte Carlo (MCMC) Inference	14
1.3.4 MCMC Diagnostics and Simulating Posterior Trajectories	14
1.4 Unsupervised Learning and Clustering	15
1.4.1 K-means Clustering	15
1.5 Deep Learning	16
1.5.1 Feedforward Networks	18
1.5.2 The Architecture of Neural Networks	19

1.5.3	Learning with Gradient Descent	19
1.5.4	Activation Function	22
1.5.5	Convolutions Neural Networks (CNN)	24
CHAPTER II. CLUSTERING: REPRESENTATIVE COUNTRY IDENTIFICATION		26
2.1	Hofstede’s Culture Dimensions	26
2.2	Elbow Method for Optimal Number of Clusters	28
2.3	Clustering and Representative Countries	28
CHAPTER III. BAYESIAN PREDICTIVE MODELING OF COVID-19 USING CULTURE AND ECONOMICAL BEHAVIOURAL FEATURES		33
3.1	Literature Review for COVID	33
3.1.1	Modeling COVID-19	33
3.1.2	Culture, Behavior and Economy in COVID-19	34
3.2	Game Theory and the Swiss Cheese Model	36
3.3	COVID-19 Disease Model	37
3.4	Data Description	39
3.4.1	COVID Cases	39
3.4.2	GDP and Population	40
3.4.3	Lockdown Stringency Index	40
3.5	Statistical Data Model	41
3.6	Model Simulation	43
3.7	Results	43
3.7.1	MCMC Inference	43
3.7.2	Trace Plots to Check MCMC Mixing	44
3.7.3	Autocorrelation and Effective Sample Size to Check Correlation Diagnostics	44
3.7.4	Posterior Distributions	46
3.7.5	Simulation and Discussion	49
3.7.6	Drawback of R for Complex Dynamical System	50
CHAPTER IV. CONVOLUTIONAL NEURAL NETWORK TO DETERMINE CULTURE FROM COVID-19		51
4.1	Tabular Data to Image Transformation	51
4.2	Development of Convolutional Neural Network (CNN)	54
4.3	Experimental Setup	55
4.4	Performance Evaluation Metrics	57

4.4.1	Results Analysis for Power Distance	57
4.4.2	Results Analysis for Individualism	59
4.4.3	Results Analysis for Masculinity	64
4.4.4	Results Analysis for Uncertainty Avoidance	67
4.4.5	Results Analysis for Long Term Orientation	72
4.4.6	Results Analysis for Indulgence	74
4.5	Discussion and Conclusion	77
	REFERENCES	80
	BIOGRAPHICAL SKETCH	85

LIST OF TABLES

	Page
Table 2.1: Culture dimension for each center of clusters (k=4)	30
Table 2.2: Culture dimension for each center of clusters (k=5)	31
Table 2.3: Centers of the cluster (k=4)	32
Table 2.4: Centers of the cluster (k=5)	32
Table 3.1: Data sources	40
Table 3.2: Priors and ranges of parameters	42

LIST OF FIGURES

	Page
Figure 1.1: Susceptible-Infectious-Recovered (SIR) model	2
Figure 1.2: Susceptible-Exposed-Infectious-Recovered (SEIR) model	5
Figure 1.3: A shallow model (left); decision list network (middle); deep learning network (right)	17
Figure 1.4: Objective function of deep learning model	20
Figure 1.5: Commonly used activation functions in deep learning systems: (a) sigmoid function; (b) ReLU and softplus function; (c) tanh function	23
Figure 2.1: Using distortion for an elbow method	29
Figure 3.1: Schematic illustration of the COVID-19 SEAIHRD model	38
Figure 3.2: MCMC trace and density plot of parameter r	44
Figure 3.3: MCMC trace and density plot of parameter α_{01}	45
Figure 3.4: MCMC trace and density plot of parameter α_{02}	45
Figure 3.5: Auto-correlation plot of parameter α_{02}	46
Figure 3.6: Posterior distributions of parameters	48
Figure 3.7: Simulation of the disease prevalence	49
Figure 4.1: State of art CNN performance analysis	52
Figure 4.2: COVID-19 new cases for the first 300 days in the United States	52
Figure 4.3: Transformed Image into 300×300 Pixel for the United States	53
Figure 4.4: Convolutional Neural Network Architecture without Residual Block	55
Figure 4.5: Convolutional Neural Network Architecture wit Residual Block	56
Figure 4.6: Confusion matrix of culture (power distance) prediction based on CNN architecture	57
Figure 4.7: Accuracy of culture (power distance) prediction with varying epoch based on CNN architecture	58
Figure 4.8: Loss of culture (power distance) prediction with varying epoch based on CNN architecture	58
Figure 4.9: Accuracy of culture (power distance) prediction with varying epoch based on CNN with residual block	59
Figure 4.10: Loss of culture (power distance) prediction with varying epoch based on CNN with residual block	60

Figure 4.11: Learning rate of culture (power distance) prediction with varying batch based on CNN with residual block	60
Figure 4.12: Confusion matrix of culture (individualism) prediction based on CNN architecture	61
Figure 4.13: Accuracy of culture (individualism) prediction with varying epoch based on CNN architecture	62
Figure 4.14: Loss of culture (individualism) prediction with varying epoch based on CNN architecture	62
Figure 4.15: Accuracy of culture (individualism) prediction with varying epoch based on CNN with residual block	63
Figure 4.16: Loss of culture (individualism) prediction with varying epoch based on CNN with residual block	63
Figure 4.17: Learning rate of culture (individualism) prediction with varying batch based on CNN with residual block	64
Figure 4.18: Confusion matrix of culture (masculinity) prediction based on CNN architecture	65
Figure 4.19: Accuracy of culture (masculinity) prediction with varying epoch based on CNN architecture	65
Figure 4.20: Loss of culture (masculinity) prediction with varying epoch based on CNN architecture	66
Figure 4.21: Accuracy of culture (masculinity) prediction with varying epoch based on CNN with residual block	66
Figure 4.22: Loss of culture (masculinity) prediction with varying epoch based on CNN with residual block	67
Figure 4.23: Learning rate of culture (masculinity) prediction with varying batch based on CNN with residual block	68
Figure 4.24: Confusion matrix of culture (uncertainty avoidance) prediction based on CNN architecture	68
Figure 4.25: Accuracy of culture (uncertainty avoidance) prediction with varying epoch based on CNN architecture	69
Figure 4.26: Loss of culture (uncertainty avoidance) prediction with varying epoch based on CNN architecture	70
Figure 4.27: Accuracy of culture (uncertainty avoidance) prediction with varying epoch based on CNN with residual block	70
Figure 4.28: Loss of culture (uncertainty avoidance) prediction with varying epoch based on CNN with residual block	71
Figure 4.29: Learning rate of culture (uncertainty avoidance) prediction with varying batch based on CNN with residual block	71

Figure 4.30: Confusion matrix of culture (long term orientation) prediction based on CNN architecture	72
Figure 4.31: Accuracy of culture (long term orientation) prediction with varying epoch based on CNN architecture	73
Figure 4.32: Loss of culture (long term orientation) prediction with varying epoch based on CNN architecture	73
Figure 4.33: Accuracy of culture (long term orientation) prediction with varying epoch based on CNN with residual block	74
Figure 4.34: Loss of culture (long term orientation) prediction with varying epoch based on CNN with residual block	75
Figure 4.35: Learning rate of culture (long term orientation) prediction with varying batch based on CNN with residual block	75
Figure 4.36: Confusion matrix of culture (indulgence) prediction based on CNN architecture	76
Figure 4.37: Accuracy of culture (indulgence) prediction with varying epoch based on CNN architecture	76
Figure 4.38: Loss of culture (indulgence) prediction with varying epoch based on CNN architecture	77
Figure 4.39: Accuracy of culture (indulgence) prediction with varying epoch based on CNN with residual block	78
Figure 4.40: Loss of culture (indulgence) prediction with varying epoch based on CNN with residual block	78
Figure 4.41: Learning rate of culture (indulgence) prediction with varying batch based on CNN with residual block	79

CHAPTER I

INTRODUCTION

1.1 Compartmental Models of Infectious Disease

Disease models play an essential role in understanding and managing pathogen transmission dynamics. They can describe spatial and temporal patterns of disease prevalence and investigate or better understand the factors that influence disease incidence. Modeling is crucial in determining which treatments and interventions are most effective, how cost-effective these approaches are, and what specific factors must be considered when attempting to eradicate the disease. The findings of those models can help policymakers implement practical, real-world solutions.

Compartmental modeling is very effective and popular for infectious disease modeling, where the population is assigned into compartments. Epidemiologists frequently use a set of models known as compartmental models to understand the complex dynamics underlying disease transmission. These models, which were developed in the early twentieth century, divide a population into groups based on their risk or infection status. Compartmental modeling is very effective and popular for infectious disease modeling, where the population is assigned into compartments. A system of differential equations governs compartmental models, which track the population as a function of time, stratifying it into different groups based on risk or infection status [6, 5].

Compartmental models are deterministic (could also be stochastic), which means they produce the same results every time given the same inputs. They can predict the various properties of pathogen spread, estimate the duration of epidemics, and be used to understand how different situations or interventions affect pathogen spread outcomes.

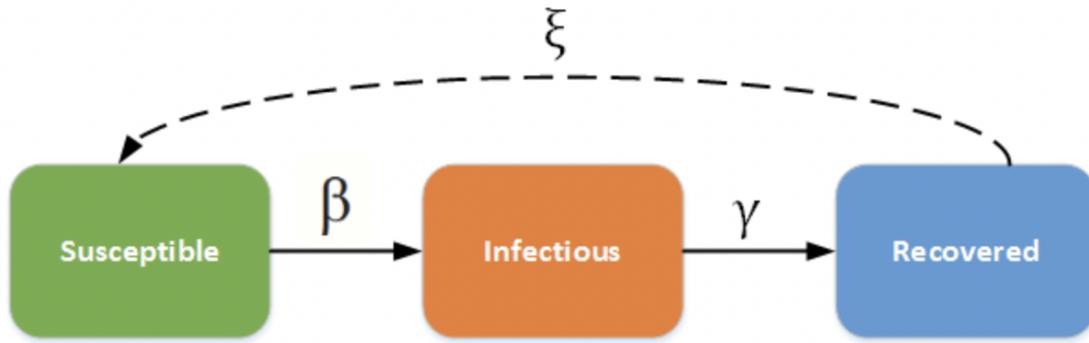


Figure 1.1: Susceptible-Infectious-Recovered (SIR) model [2].

1.1.1 Susceptible-Infectious-Recovered (SIR) Model

In 1927, W. O. Kermack and A. G. McKendrick developed the SIR (Susceptible, Infectious, and Removed) model; the total population was three compartment. Each variable ($S(t)$, $I(t)$, $R(t)$) indicates the number of people in each compartment at a particular time t . Susceptible (S) generally indicate the number of people who are vulnerable to the disease. Infectious (I) shows the number of infected individuals who have the capability of infecting other people in the S compartment. Removed (R) indicates the number of people who have recovered or died. When people from the S and I compartments come into infectious contact, then the people from the S compartment become infected and move into the I compartment.

Commonly the model consists of deterministic ordinary differential equations, but in some cases, it may be stochastic, which is complicated to analyze. The compartmental model generally attempts to predict quantities related to the disease.

The SIR/SIRS diagram in Figure 1.1 depicts how individuals move through the model's compartments. The dashed line represents how the SIR model transforms into the SIRS (Susceptible - Infectious - Recovered - Susceptible) model, in which recovery does not confer lifelong immunity and individuals can become susceptible again. The transmission rate β is the rate of a successful contact in transmitting the disease. The average duration of infection, D , determines the recovery rate, $\gamma = \frac{1}{D}$. The SIR model can be written as the following ordinary differential equation (ODE)

when births and deaths are included [2]:

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} - \nu S \quad (1.1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \nu I \quad (1.2)$$

$$\frac{dR}{dt} = \gamma I - \nu R \quad (1.3)$$

with total population $N = S + I + R$ and μ is the birth rate and ν is the death rate.

Because of the nonlinear dynamics, it is challenging to derive exact analytical solutions to the previous equations. The key metrics that control the spread, on the other hand, can be numerically calculated. For a disease to spread, the following conditions must be met at the time of infection seeding:

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \nu I > 0 \quad (1.4)$$

If the number of infections is low at the start, S will be automatically equal to N , and the condition becomes:

$$\frac{\beta}{\gamma + \nu} > 1 \quad (1.5)$$

where $\frac{\beta}{\gamma + \nu}$ is named as the reproduction number R_0 . In a fully susceptible population, R_0 is the average number of secondary cases generated by an index case in a completely susceptible population. When $R_0 > 1$, the disease spreads in the population and dies out when $R_0 < 1$. This applies to all types of compartmental models.

- **SIR Without Vital Dynamics:** If the infection's course is short (emergent outbreak) in comparison to an individual's lifetime and the disease is not fatal, vital dynamics (birth and death) can be ignored. The SIR model can be written as the following ordinary differential

equation (ODE) in deterministic form:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \quad (1.6)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \quad (1.7)$$

$$\frac{dR}{dt} = \gamma I \quad (1.8)$$

An epidemic will eventually die out in a closed population with no vital dynamics due to a lack of susceptible individuals to sustain the disease. Because of the existing population's lifelong immunity, infected individuals who are added later will not start another epidemic.

- **SIR With Vital Dynamics:** On the other hand, new births in a population with vital dynamics can provide more susceptible individuals to the population, sustaining an epidemic or allowing new introductions to spread throughout the population. Disease dynamics will reach a steady-state in a realistic population like this. This is the case when a disease is endemic to a particular region. Considering μ and ν represent the model's birth and death rates, respectively. Assume $\mu = \nu$ the population constant. In a stable state $\frac{dI}{dt} = 0$.

1.1.2 Susceptible-Exposed-Infectious-Recovered (SEIR) Model

SEIR model consists of four compartments similar to the SIR model with an additional exposed compartment which indicates the fraction of exposed individuals. Many diseases have a dormant phase in which the person is infected but not infectious. This time lag between infection and infectious state can be incorporated into the SIR model by including a latent/exposed population, E , and allowing infected (but not yet infectious) individuals to move from S to E and then from E to I .

The differential equations that govern the classic deterministic SEIR compartmental models are discussed in this topic. The SEIR/SEIRS diagram in Figure 1.2 depicts how individuals move through the model's compartments. The dashed line describes how the SEIR model evolves into the SEIRS (Susceptible - Exposed - Infectious - Recovered - Susceptible) model, in which

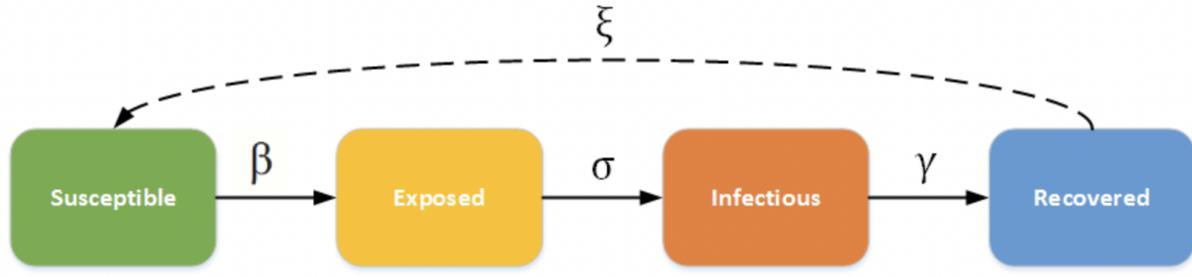


Figure 1.2: Susceptible-Exposed-Infectious-Recovered (SEIR) model [1].

recovered individuals may become susceptible again (recovery does not confer lifelong immunity). Rotavirus and malaria, for example, are diseases with long incubation periods and where recovery only provides temporary immunity. The transmission rate, β , determines the rate of spread. The incubation rate σ is the rate of latent individuals becoming infectious (the average duration of incubation is $\frac{1}{\sigma}$). The average duration of infection, D , determines the recovery rate, ($\gamma = \frac{1}{D}$). In SEIRS model ξ is the rate at which recovered individuals revert to a susceptible state due to immunity loss.

- **SEIR without Vital Dynamics:** In a closed population with no births or deaths, the SEIR model takes the following form [1]:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{1.9}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \gamma E \tag{1.10}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{1.11}$$

$$\frac{dR}{dt} = \gamma I \tag{1.12}$$

After a period of rapid expansion, the epidemic depletes the susceptible population. The virus eventually runs out of new susceptible people and dies. The introduction of the incubation period does not affect the total number of infected people.

- **SEIR with Vital Dynamics:** As with the SIR model, enabling vital dynamics (births and deaths) can either sustain an epidemic or allow new introductions to spread because new

deliveries provide more vulnerable individuals. Disease dynamics will reach a steady-state in a realistic population like this. Where μ and ν represent the birth and death rates, respectively, and are assumed to be equal to keep the population constant, the ODE becomes:

$$\frac{dS}{dt} = \mu N - \nu S - \frac{\beta SI}{N} \quad (1.13)$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \nu E - \sigma E \quad (1.14)$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \nu I \quad (1.15)$$

$$\frac{dR}{dt} = \gamma I - \nu R \quad (1.16)$$

1.2 Bayesian Method

1.2.1 Bayes Theorem

Definition 1 Consider a sample space S and an associated sigma algebra \mathcal{B} , and a probability measure P with domain \mathcal{B} . If A and B are events in \mathcal{B} , and $P(B) > 0$ then the conditional probability of A given B is defined by,

$$P(A|B) = \frac{P(A \cup B)}{P(B)} \quad (1.17)$$

Law of total probability: If $\{A_n\}_{n=1}^{\infty} \in \mathcal{B}$ is a partition of sample space S , $B \in \mathcal{B}$, then

$$P(B) = \sum_{n=1}^{\infty} P(B|A_n)P(A_n) \quad (1.18)$$

Theorem 2 Let $\{A_n\}_{n=1}^{\infty} \in \mathcal{B}$ is a partition of sample space S , and let B be any set such that $P(B) > 0$. Then, for each $i = 1, 2, \dots$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} \quad (1.19)$$

where $P(B) = \sum_{j=1}^{\infty} P(B|A_j)P(A_j)$, it is not required to begin with the knowledge of $P(B)$ [17]. The

normalization constant $P(B)$ is independent of the model parameters and ensures the fraction will be between 0 and 1.

Bayesian analysis is quantifying and updating the uncertainty/degree of belief by probability. In Bayesian analysis, the observed data are used to update the prior information or beliefs resulting in posterior belief. Bayesian methods generally follow the following steps:

- Specify prior distribution parameters θ ; the uncertainty needs to be expressed as a probability distribution called prior distribution $P(\theta)$.
- After observing the data D , the updated knowledge about the parameter θ is called posterior distribution, denoted by $P(\theta|D)$. The posterior distribution quantifies the relative uncertainty that each potential value is the actual value, expressing the uncertainty regarding the value of θ .

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

1.2.2 Selection of Prior Distribution

One crucial question in Bayesian statistics is: what criteria do we use to select a prior distribution? Prior selection reflect on our viewpoints, opinions, and uncertainties. Theoretically, we're defining the parameter's cumulative distribution function. If one gets enough data, the information included in the data will outnumber the earlier belief. Any appropriate prior distribution will yield a posterior distribution that is close to the exact value of the parameter. There are, however, a few things that can go wrong. Calibration is a valuable idea for prior selection—predictive intervals calibration. So, if we design an interval in which we estimate that 95% of new data points will occur, we may declare that this interval will contain 95% of new data points. It would be ideal if 95 percent of new data points truly fell inside that range. How can we adjust our expectations to reality? Although this is increasingly common as a notion, the results must mirror reality for practical purposes. In general, prior to the model parameters over the predictor features are based on expert knowledge, naive, non-informative priors, Jeffrey's priors.

- **Jeffrey's Priors:** Jeffrey's argued that a non-informative prior should be invariant to the

parameterization used. Jeffrey's prior is the prior distribution that satisfies $P(\theta) \propto J(\theta)^{\frac{1}{2}}$, where $J(\theta)$ is the Fisher information for θ . Its essential characteristic is that it is invariant for the parameter vector θ when the coordinates are changed. That is, regardless of the parameterization used to define the Jeffrey's prior, the relative probability assigned to a volume of a probability space using a Jeffrey's prior will be the same. As a result, it is particularly interesting to be used for scale parameters.

- **Non-informative Priors:** With "really no information," there is no such thing as a prior. The term "uninformative" prior is a misnomer. Any previous distribution comprises a specification corresponding to a certain amount of data.

1.2.3 Likelihood Function

Conditional distribution of the response and predictor features given the model will be data-driven and the number of sample data increases the likelihood overwhelms the prior distribution. A statistical analysis aims to estimate the proposed model's unknown parameter(s). The likelihood function is essential to estimate the unknown parameters. For count data, older and less sophisticated methods include the method of moments and the method of minimum chi-square. These estimators are not always efficient, and their sampling distributions are frequently intractable mathematically. Let X_1, X_2, \dots, X_n have a joint density function $f(X_1, X_2, \dots, X_n | \theta)$. Given $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ is the observed data and the function θ is defined by:

$$\mathcal{L}(\theta) = \mathcal{L}(\theta | x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta)$$

is the likelihood function. It assesses the data's support for each possible value of the parameter. If we compare the likelihood function at two parameter points and discover that $L(\theta_1 | x) > L(\theta_2 | x)$, the sample we observed is more likely to have occurred if $\theta = \theta_1$ than if $\theta = \theta_2$. This can be interpreted as θ_1 being a more likely value than θ_2 .

1.2.4 Bayesian Statistical Inference

Bayesian inference is a statistical inference method that uses Bayes' theorem to update the probability of a hypothesis as new evidence or information becomes available. Bayesian inference is a crucial technique in statistics, particularly mathematical statistics. Bayesian updating is significant in the dynamic analysis of a data sequence. Bayesian inference has been used in various fields, including science, engineering, philosophy, medicine, sports, and law. Bayesian inference is closely related to subjective probability, which is often referred to as "Bayesian probability" in decision theory.

- **Credible Intervals:** Credible intervals are an important concept in Bayesian statistics. Its primary goal is to describe and summarize the uncertainty associated with the unknown parameters that we are attempting to estimate. In this regard, it may appear to be very similar to the frequentist Confidence Intervals. While their goals are similar, their statistical definitions and meanings are very different. Indeed, while the latter is obtained through a complex algorithm full of rarely-tested assumptions and approximations, the credible intervals are fairly simple to compute. Because Bayesian inference yields a distribution of possible effect values (the posterior), the credible interval is simply a range containing a specific percentage of probable values. The 95 percent credible interval, for example, is simply the central portion of the posterior distribution that contains 95 percent of the values.
- **Highest Density Interval (HDI):** The highest density interval, abbreviated HDI, is another way of summarizing a distribution that we frequently use. The HDI indicates which points of a distribution are the most credible and cover most of the distribution. As a result, the HDI summarizes the distribution by specifying an interval that spans the majority of it, say 95 percent, such that every point inside the interval has greater credibility than any point outside the interval.
- **Bayes Factor:** Bayes factors are an essential component of the Bayesian approach to hypothesis testing. Hypothesis testing in frequentist statistics is as simple as selecting a test

statistic and calculating the p-value. The p-value assesses the strength of evidence in favor of (or against) a null hypothesis. The smaller its value, the more substantial the evidence against the null hypothesis, but the more precise interpretation of p-values is a matter of convention. For example, $p < 0.05$ is widely regarded as a yardstick: any higher value of p is regarded as insufficiently strong evidence against the null hypothesis to reject it. When interpreting p-values and communicating their meaning to others, researchers frequently fall into the trap of saying that $p < 0.05$ means the null hypothesis has a less than 5% chance of being true. This is incorrect. Unfortunately, p-values do not have a direct interpretation as evidence measures.

We can only discuss the likelihood that a hypothesis is correct within the Bayesian framework. The essence of a Bayesian hypothesis test is to compute and report on this probability. Given data x and a hypothesis H , we need the posterior probability $P(H|x)$ that H is true, which can be calculated using Bayes' theorem as follows:

$$P(H|x) = \frac{P(H)P(x|H)}{P(H)P(x|H) + P(H^c)P(x|H^c)}$$

Consider the posterior odds in favor of H , we define [39],

$$\frac{P(H|x)}{P(H^c|x)} = \frac{P(H)}{1 - P(H|x)} = \frac{P(H)}{P(H^c)} \times \frac{P(x|H)}{P(x|H^c)}$$

The final factor on the right is the Bayes factor, $B_H(x)$.

- **Bayesian Information Criterion (BIC):** BIC is a popular model assessment approach that can be used in situations where the fitting is done by maximization of a log-likelihood. The generic form of BIC [22],

$$BIC = -2 \cdot \log \text{lik} + (\log N) \cdot d$$

1.2.5 Markov chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) methods are widely used in various disciplines to fit complex models. A Google search for "Markov chain Monte Carlo" yields over 11.5 million results. MCMC's popularity stems primarily from its use in computational physics and Bayesian statistics, though it is also used in frequentist inference. Because of its flexibility, ease of use, and generality, Markov chain Monte Carlo (MCMC) is one of the most practical approaches to scientific computing. In fact, MCMC is required for performing Bayesian analysis. The basic idea behind MCMC is that if simulating from a target density π is difficult enough that the conventional Monte Carlo method based on independent and identically distributed (iid) samples cannot be used to make inference on π , it may be possible to build a Markov chain $X_n \geq 0$ with stationary density π for forming Monte Carlo estimators. Here is the basic summary of the Markov chain Monte Carlo:

- Class of algorithms for sampling from probability distributions
- Markov chain of samples equilibrium distribution is the target distribution
- Each Markov step is sampling the target distribution, more steps, more samples, more accurate
- Markov screening: state at $t + 1$ is only dependent on state at t
- Given we do not know about the distribution. Can we construct a set of samples that search over the range of possible outcomes and spend an amount of time in each interval proportional to the actual density of that distribution?

1.2.6 MCMC Convergence Diagnostics

Two critical questions for MCMC practitioners are where to begin and when to stop the simulation. These two tasks are concerned with determining the convergence of the underlying Markov chain to stationarity and, respectively, the convergence of Monte Carlo estimators to population quantities. It is known that for any initial value on the Markov chain, the distribution of X_n converges to the stationary distribution as $n \rightarrow \infty$ under standard conditions. Despite extensive

research into developing convergence criteria and stopping rules with solid theoretical foundations, MCMC users frequently decide on convergence by using practical diagnostic tools. This section attempts to discuss the most commonly used MCMC convergence diagnostic tools.

- **Effective Sample Size:** In the context of MCMC, the Effective Sample Size (ESS) measures a sample chain's information content or effectiveness. For example, one thousand samples with an ESS of 200 have more information than 2,000 samples with an ESS of 100.
- **Auto-correlation.** MCMC samples are dependent, and this has no effect on the validity of inference on the posterior if the sampler has enough time to explore the posterior distribution. Still, it does affect the sampler's efficiency. In other words, highly correlated MCMC samplers necessitate a more significant number of samples to produce the same level of Monte Carlo error for an estimate.
- **Gelman-Rubin Statistic** Another method for monitoring an MCMC sampler's convergence is to consider what we might expect when a chain has "converged." The theory claims that if we start multiple parallel chains with different starting values, they will eventually converge to the stationary distribution. So, it should be impossible to distinguish between the various chains after a while. They should all "appear" to be a stationary distribution. One way to evaluate this is to compare chains to variations within chains. If all chains are "the same," the divergence between chains should be close to zero.

Let us consider samples x_1^j, x_2^j, \dots from the j^{th} Markov chain and there are J chains which run in parallel with varying starting values.

- First Discard D values as burn-in and keep the remaining L values for each chain such as $x_D^j, x_{D+1}^j, \dots, x_{D+L-1}^j$.

– Calculate

$$\text{Chain mean, } \bar{x}_j = \frac{1}{L} \sum_{t=1}^L x_t^j \quad (1.20)$$

$$\text{Grand mean, } \bar{x} = \frac{1}{J} \sum_{j=1}^J \bar{x}_j \quad (1.21)$$

$$\text{Between chain variance, } B = \frac{L}{J-1} \sum_{j=1}^J (\bar{x}_j - \bar{x})^2 \quad (1.22)$$

$$\text{Within chain variance, } s_j^2 = \frac{1}{L-1} \sum_{t=1}^L (x_t^j - \bar{x}_j)^2 \quad (1.23)$$

$$W = \frac{1}{J} \sum_{j=1}^J s_j^2 \quad (1.24)$$

– Finally, the Gelman-Rubin statistics

$$R = \frac{\frac{L-1}{L}W + \frac{1}{L}B}{W}$$

Because the Gelman-Rubin statistic is a ratio and thus unit free, it is a straightforward summary for an MCMC sampler. Furthermore, unlike Monte Carlo standard errors, it can be implemented without first specifying a parameter to be estimated. As a result, it can be a valuable tool for monitoring a chain prior to making any specific decisions about the types of inferences to be drawn from the model.

1.3 Step by Step Implementation of Bayesian Inference with R

Differential equation (DE) models in R (software) are simple to implement, with readable code and access to R's many high-level functions [9]. We use "deBInfer" as an R-based inference package for the Bayesian inference. However, numerically solving DE models specified as R functions takes a long time. The "deSolve" package also supports evaluating DE models written in lower-level languages like C and FORTRAN. These compiled models have the advantage of faster simulation. Because the DE model is considered numerous times during the MCMC procedure, even minor speedups from compiled models can result in significant absolute time savings.

1.3.1 Specifying the Ordinary Differential Equation (ODE) Model

The first step of implementing Bayesian inference in R with the package "deBinfer" is specifying the ODE model, which is done in section III.

1.3.2 ODE Model Implementation

After specifying the ODE equation model, the ODE is solved numerically in R using the "deSolve" package. To speed up the parameter inference, we try C programming also by creating complied models.

1.3.3 Parameter Declaration and Markov chain Monte Carlo (MCMC) Inference

After implementing the ODE model, we define the parameters of dynamical system for inference. After that for MCMC inference the declared parameters are collated using `setup_debinfer` function of the package in R. "de_mcmc" in the package runs the MCMC estimation. The MCMC procedure's progress can be tracked using the `cnt`, `plot`, and `verbose` options: The function will print out information about the current state every `cnt` iteration, and if `plot=TRUE`, trace plots of the chains will be plotted. It was setting `verbose=TRUE` causes more information to be printed. It should be noted that frequent plotting will significantly slow down the MCMC sampler and should only be used on short runs when tuning the sampler.

1.3.4 MCMC Diagnostics and Simulating Posterior Trajectories

After the MCMC inference, we plot and summarize the MCMC chains. In R, "post_prior_densplot" can be obtained by plotting the parameters one at a time with the `param` option. The x and y limits of the plots can then be adjusted to show more of the prior support, and fancy labels can be added. For the deterministic part of the model, we simulate 100 differential equation model trajectories from the posterior and calculate the 95 percent highest posterior density interval. After that, we visualize the median posterior trajectory and the highest posterior density interval. Finally, we make customized plots by accessing the simulated trajectories within the `post_traj` object.

1.4 Unsupervised Learning and Clustering

Unsupervised learning is an algorithm that uses untagged data to learn patterns. The hope is that the machine will be forced to build a compact internal representation of its world through mimicry, a basic mode of human learning, and then generate imaginative content from it. The most common unsupervised learning algorithms are density estimation, clustering (discrete), and dimensionality reduction (continuous).

Clustering is the task of grouping a set of items so that those in the same group (called a cluster) are more similar to those in other groups (clusters). Mixture models can be used to cluster data and provide a framework for building more complex probability distributions. As a result, we begin our discussion of mixture distributions with the problem of finding clusters in a set of data points, which we approach first with a non-probabilistic technique known as the K-means algorithm.

1.4.1 K-means Clustering

To begin, consider the problem of identifying groups of data points, or clusters, in a multidimensional space. Assume we have a data set x_1, \dots, x_N with N observations of a random D -dimensional Euclidean variable x . Our goal is to divide the data set into some number K of clusters, with the value K assumed for the time being. Intuitively, we might think of a cluster as a group of data points with small inter-point distances compared to distances to points outside the cluster. This concept can be formalized by introducing a set of D -dimensional vectors k , where $k = 1, \dots, K$, and k is a prototype associated with the k^{th} cluster. We can think of the k as representing the cluster centers, as we will see shortly. The goal is then to find an assignment of data points to clusters as well as a set of vectors k such that the sum of the squares of each data point's distances to its closest vector k is a minimum.

It is convenient to define some notation to describe the assignment of data points to clusters at this point. We introduce a corresponding set of binary indicator variables $r_{nk} \in \{0, 1\}$, where $k = 1, \dots, K$ describing which of the K clusters the data point x_n is assigned to, so that if data point x_n is assigned to cluster k , $r_{nk} = 1$, and $r_{nj} = 0$ for $j \neq k$. This is referred to as the 1-of- K

coding scheme. Equation 1.4, which represents the sum of the squares of the distances of each data point to its assigned vector k , can then be used to define an objective function known as a distortion measure.

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (1.25)$$

We aim to find values for the r_{nk} and k that minimize J . We can accomplish this through an iterative procedure in which each iteration consists of two successive steps corresponding to successive optimizations of the r_{nk} and the k . First, we choose some initial k values. Then, in the first phase, we minimize J with respect to r_{nk} while keeping k constant. In the second phase, we minimize J with k while keeping r_{nk} constant. This two-stage optimization process is then repeated until convergence is reached.

1.5 Deep Learning

Scientists have long cherished creating a machine that can think and make decisions like humans. We human can do some task effortlessly like recognizing digits, images, but the same task become becomes extremely difficult if we try to make an attempt to write a computer program to recognize digit like human. This adores humans to invent neural networks and deep learning, which allow computers to learn from experience and understand the world in terms of the hierarchy of concepts. The concept accomplished some fantastic tasks, like IBM's Deep Blue chess-playing system defeating the world champion of chess Garry Kasparov in 1997 (citation). One pioneer of deep learning is called multilayer perceptron (MLP), also called feedforward deep learning. MLP is nothing but a mathematical mapping of some set of inputs values to output values. Deep learning dates back to the 1940s, and it calls cybernetics(1940-1960s) on that time and connectionism in 1980-1990s, and people started telling it deep learning from the beginning of 2006. The current wave of deep learning starts with this paper (citation). Deep learning is an engineering system inspired by biological neurons.

Deep learning is a broad class of machine learning techniques in which hypotheses are represented as complex algebraic circuits with tunable connection strengths. The term "deep" refers to the

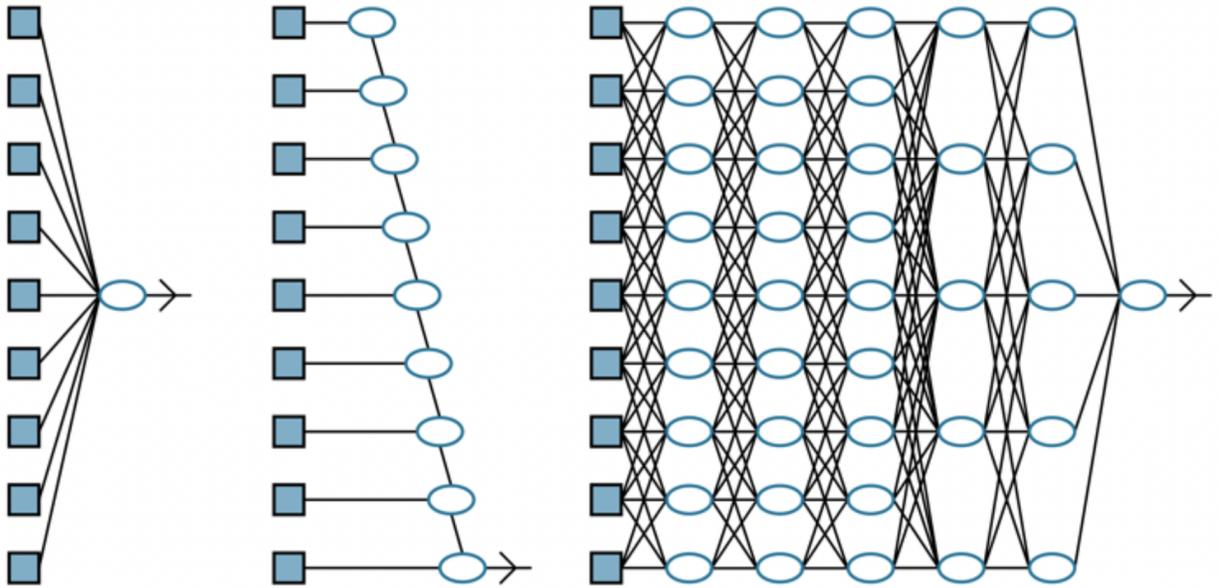


Figure 1.3: A shallow model (left); decision list network (middle); deep learning network (right) [43].

fact that circuits are typically organized into many layers, implying that computation paths from inputs to outputs are lengthy. Deep learning enables computational models of multiple processing layers to learn data representations with varying degrees of abstraction. These methods have significantly advanced the state-of-the-art in speech recognition, visual object recognition, object detection, and various other fields such as drug discovery and genomics. Deep learning discovers intricate structures in large data sets by using the backpropagation algorithm to indicate how a machine’s internal parameters used to compute each layer’s representation from the previous layer’s representation should be changed. Deep convolutional nets have significantly advanced in image, video, speech, and audio processing, whereas recurrent nets have shed light on sequential data such as text and address [29].

Deep learning has its roots in early work that attempted to model neural networks in the brain with computational circuits (McCulloch and Pitts, 1943). As a result, deep learning networks are frequently referred to as neural networks, despite their superficial resemblance to actual neural cells and structures [43]. While the exact reasons for deep learning’s success have yet to be fully explained, it has advantages over some traditional machine learning methods, particularly for

high-dimensional data such as images or texts. Although linear and logistic regression methods can handle many input variables, the computation path from each input to the output is short: multiply by a single weight, then add to the aggregate output. Furthermore, the various input variables contribute to the output independently Figure 1.3(left). This severely restricts the expressive power of such models. They can only represent linear functions and boundaries in the input space, whereas most real-world concepts are far more complex.

Long computation paths that depend on many input variables are possible with decision lists and decision trees, but only for a small fraction of the possible input vectors (Figure 1.3(middle)). If a decision tree has long computation paths for a significant fraction of the possible inputs, the number of input variables must be exponentially large. Deep learning's basic idea is to train circuits so that the computation paths are long, allowing all of the input variables to interact in complex ways (Figure 1.3(right)). These circuit models are found to be sufficiently expressive to capture the complexity of real-world data for many significant learning problems.

Today deep learning reaches a level for surpassing humans in medical, vision, natural language, and robotics. The key reasons are the increase of the dataset and model size and the high-performance GPU computing system.

1.5.1 Feedforward Networks

This section focuses on simple feedforward and recurrent networks, their components, and the learning fundamentals in such networks. As the name implies, a feedforward network has connections only in one direction, forming a directed acyclic graph with designated input and output nodes. Each node in the network computes a function from its inputs and passes the result to its successors. The network has no loops as information flows from the input to the output nodes. On the other hand, a recurrent network feeds its intermediate or final outputs back into its own inputs. The network's signal values form a dynamic system with an internal state or memory.

Feedforward networks are illustrated by Boolean circuits, which implement Boolean functions. The inputs in a Boolean circuit are limited to 0 and 1, and each node implements a simple Boolean function of its inputs, yielding a 0 or a 1. The input values in neural networks are typically

continuous, and nodes take continuous inputs and produce continuous outputs. Some of the inputs to nodes are network parameters; the network learns by adjusting the values of these parameters to fit the training data as a whole.

1.5.2 The Architecture of Neural Networks

Two most important artificial neurons are perceptrons and sigmoid neurons. Perceptrons were developed by Frank Rosenblatt in 1950s and 1960s inspired by the work of Warren McCulloch and Walter Pitts. Generally, a perceptron takes binary inputs x_1, x_2, \dots , and produces a single binary output. Rosenblatt proposed initialization of weights w_1, w_2, \dots , real numbers expressing the relative importance of the inputs to the output. The output of the neuron's is determined by whether the weighted sum $\sum_i w_i x_i$ is less than or greater than some threshold value. In perceptron's bias is a measure of how easy it is to get the perceptron to output a 1. In biological point of view, we can define bias as how easy it is for perceptron to fire. Considering the output is defined by the following equation below (citation):

$$\text{output} = \begin{cases} 0 & \text{if } \sum_i w_i x_i + \text{bias} \leq 0 \\ 1 & \text{if } \sum_i w_i x_i + \text{bias} > 0 \end{cases}$$

It is obvious that perceptron is not a complete model of human brain decision making but perceptron can weight up different kinds of evidence and many layers of perceptrons can make complex and subtle decisions. We can able to devise the learning algorithms which can automatically adjust the weight and bias of artificial neurons.

1.5.3 Learning with Gradient Descent

Any machine/deep learning model's ultimate objective is minimizing the loss function to a local minimum, as shown in Figure 1.4. The univariate linear model has the excellent property of being simple to find an optimal solution with zero partial derivatives. However, this is not always the case, so we present here a method for minimizing loss that does not rely on solving to find zeroes of derivatives and can be applied to any loss function, no matter how complex. Modifying the parameters incrementally allows us to search through a continuous weight space. We called the

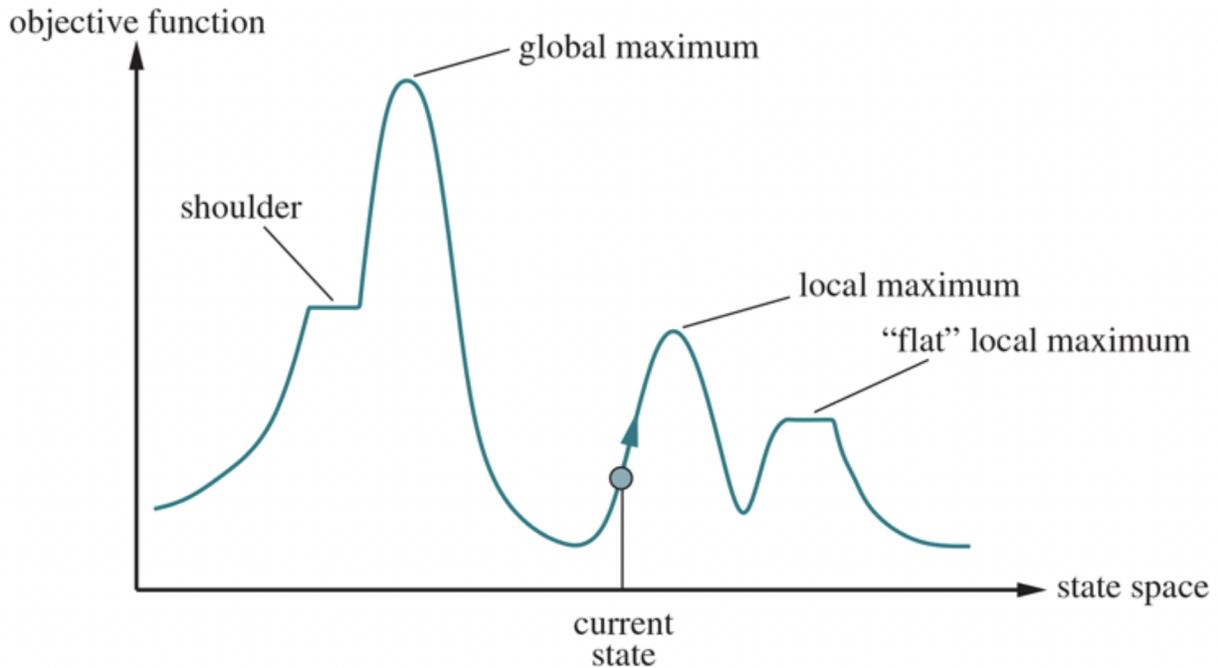


Figure 1.4: Objective function of deep learning model [43].

algorithm hill climbing there, but since we're minimizing loss rather than maximizing gain, we'll call it gradient descent. When we try to minimize the loss in a learning problem, the parameter α we call the step size in Algorithm 1 is the learning rate. It can be a constant or decay over time as the learning process progresses.

Let us define a univariate linear function with parameter \mathbf{w} to be the vector $\langle w_0, w_1 \rangle$,

$$h_{\mathbf{w}}(x) = w_1x + w_0 \quad (1.26)$$

Here, the partial derivatives - the slopes -in the simplified case of only one training example, (x, y) :

$$\frac{\partial}{\partial w_i} Loss(\mathbf{w}) = 2(y - h_{\mathbf{w}}(\mathbf{x})) \frac{\partial}{\partial w_i} (y - (w_1x + w_0)) \quad (1.27)$$

For N training example, we want to minimize the sum of the individual losses for each training example. Because the derivative of a sum is the sum of its derivatives, we get:

$$w_0 \leftarrow w_0 + \alpha \sum_j (y_j - h_{\mathbf{w}}(x_j)) \quad w_1 \leftarrow w_1 + \alpha \sum_j (y_j - h_{\mathbf{w}}(x_j)) \times x_j \quad (1.28)$$

The batch gradient descent learning rule for univariate linear regression is comprised of these updates (also called deterministic gradient descent). Because the loss surface is convex, there are no local minima to get stuck in. Convergence to the global minimum is guaranteed (as long as we don't choose α that is so large that it overshoots), but it may be prolonged: we must sum over all N training examples for each step, and there may be many steps. The problem is exacerbated if N is larger than the processor's memory size. An epoch is a step that covers all of the training examples. Stochastic

Algorithm 1 Gradient descent algorithm

```

1: procedure GRADIENT DESCENT( $w_i$ )
2:    $w \leftarrow$  any point in the parameter space
3:   while not converged do
4:     for each  $w_i$  in  $w$  do
5:        $w_i \leftarrow w_i - \alpha \frac{\partial}{\partial w_i} \text{Loss}(w)$ 

```

gradient descent (SGD) is a faster variant that randomly selects a small number of training examples at each step and updates according to Equation 1.27. The original version of SGD only used one training example for each step, but using a minibatch of m out of N examples is now more common. Assume we have $N = 1000$ examples and select a size $m = 100$ minibatch. Then, on each step, we reduced the amount of computation by a factor of 100; however, because the standard error of the estimated mean gradient is proportional to the square root of the number of examples, the expected error increases by only a factor of 10. Even if we have to take ten more steps before convergence, minibatch SGD is still ten times faster than full batch SGD in this case.

We can use parallel vector operations with some CPU or GPU architectures, making a step with m examples almost as fast as a step with only a single sample. With these constraints m would be treated as a hyperparameter that should be tuned for each learning problem.

SGD can be helpful in an online setting where new data is arriving all at once, and the stationarity

assumption may not be valid. (In fact, SGD is another name for online gradient descent.) A good model with perfect α will gradually evolve, remembering what it learned in the past while also adapting to the changes represented by new data. SGD is widely used in models other than linear regression, most notably neural networks. Even when the loss surface is not convex, the method is effective in locating good local minima near the global minimum.

The same gradient descent approach can be used to learn the weights in computation graphs. The gradient calculation is essentially the same for the weights leading into units in the output layer that generate network output. The process is slightly more complicated for weights leading into units in the hidden layers that are not directly connected to the outputs.

1.5.4 Activation Function

Each node in a network is referred to as a unit. According to McCulloch and Pitts' design, a unit computes the weighted sum of the inputs from predecessor nodes and then applies a nonlinear function to produce its output. Let a_j denote the unit's output j and $w_{i,j}$ be the weight attached to the unit i to unit j link; then, we have

$$a_j = g_j\left(\sum_i w_{i,j}a_i\right) = g_j(\mathbf{w}^T \mathbf{x}) \quad (1.29)$$

where g_j is a nonlinear activation function associated with unit j ; \mathbf{w} is the vector of weights leading into unit j ; \mathbf{x} is the vector of inputs to unit j .

The fact that the activation function is nonlinear is critical because any composition of units would still represent a linear function if it were not. Because of the nonlinearity, sufficiently large networks of units can express arbitrary functions. According to the universal approximation theorem, a network with only two layers of computational units, one nonlinear and one linear, can approximate any continuous function to an arbitrary degree of accuracy. The proof shows that an exponentially large network can represent an exponentially large number of "bumps" of varying heights at various locations in the input space, approximating the desired function. In other words, large enough networks can implement a lookup table for continuous functions, just as large enough decision trees

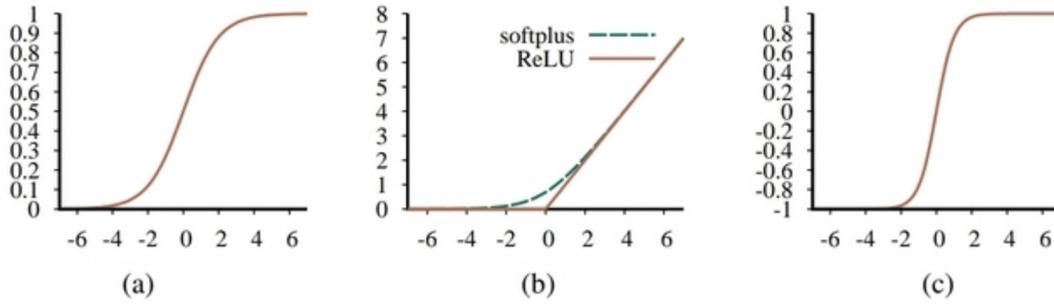


Figure 1.5: Commonly used activation functions in deep learning systems: (a) sigmoid function; (b) ReLU and softplus function; (c) tanh function. [43].

can implement a lookup table for Boolean functions. The most common activation function (Figure 1.5 are listed below:

- **Sigmoid Function:** The logistic function, also known as the sigmoid function, is used in logistic regression.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (1.30)$$

- **Softplus Function:** The softplus function is a smoother version of the ReLU function. The sigmoid function is the softplus function's derivative.

$$\text{softplus}(x) = \log(1 + e^x) \quad (1.31)$$

- **ReLU Function:** The ReLU function, an abbreviation for a rectified linear unit,

$$\text{ReLU}(x) = \max(0, x) \quad (1.32)$$

- **tanh Function:** The tanh function,

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (1.33)$$

1.5.5 Convolutions Neural Networks (CNN)

A convolutional neural network (CNN) has spatially local connections, at least in the early layers, and weight pattern replication across units in each layer. A kernel is a pattern of weights replicated across multiple local regions. Convolution is applying the kernel to the pixels of an image (or to spatially organized units in a subsequent layer).

- **Convolution:** A mathematical operation used to extract features from an image is convolution. An image kernel defines the convolution. The image kernel is essentially a small matrix. A 3×3 kernel matrix is very common most of the time. Some standard filters, such as the Sobel filter, have values 1, 2, 1, 0, 0, -1, -2, -1. This advantage is that it gives a little bit more weight to the central row, the central pixel, possibly making it a little bit more robust. Another filter used by computer vision researchers is the Scharr filter, which is 3, 10, 3, and then -3, -10, and -3. Furthermore, this has slightly different properties and can be used for vertical edge detection. If it is rotated 90 degrees, it will behave similarly to horizontal edge detection.
- **Padding:** Convolution raises two new concerns. After each convolution operation, the original image size shrinks; after multiple convolution operations, our original image becomes extremely small; however, we do not want the image to shrink every time. The second issue is that when the kernel moves over original images, it touches the edge less frequently and the middle more regularly and overlaps in the middle. As a result, the image's corner or edge features are rarely used in the output. Padding is a new concept introduced to address these two issues. Padding keeps the original image size.
- **Strides:** The number of pixels shifted across the input matrix is referred to as the stride. Padding p , filter size $f \times f$, input image size $n \times n$, and stride's' then final image dimension will be:

$$[(n + 2pf + 1)/s + 1] \times [(n + 2pf + 1)/s + 1]$$

- **Pooling:** A pooling layer is another component of a CNN. Pooling Its purpose is to gradually reduce the spatial size of the representation to reduce network complexity and computational cost. In the CNN layer, two pooling types are commonly used: Average Pooling, Maximum Pooling. Max pooling is a rule that takes the maximum of a region and aids in moving on to the image's most important features. Max pooling selects the image's brighter pixels. It is useful when the image's background is dark, and we are only interested in its lighter pixels. Average Pooling differs from Max Pooling in that it retains much information about the "less important" elements of a block or pool. Whereas Max Pooling discards them by selecting the highest value, Average Pooling blends them in. This can be useful in various situations where such information would be beneficial.

CHAPTER II

CLUSTERING: REPRESENTATIVE COUNTRY IDENTIFICATION

Clustering is one of the most common exploratory data analysis approaches. A cluster is a collection of data points that have been grouped due to commonalities. Here, the clustering aims to identify the representative country where we can show the Bayesian analysis and deep learning simulation result of the representative country. We develop the k means clustering based on the six cultural dimensions: Power Distance, Individualism, Masculinity, Uncertainty Avoidance, Long Term Orientation, and Indulgence.

2.1 Hofstede's Culture Dimensions

Culture dimensions data of 115 countries was obtained from Hofstede's cultural dimensions. Geert Hofstede established Hofstede's cultural aspects theory as a foundation for cross-cultural communication. It uses a structure drawn from component analysis to highlight the influence of a society's culture on its members' values and how these values connect to behavior. The original idea proposed four dimensions to examine cultural values along masculinity-femininity, individualism-collectivism, uncertainty avoidance, power distance (social hierarchy strength), and masculinity-femininity (task orientation versus person-orientation). The country's scores on the dimensions are relative, in the sense that we are all human while yet being distinct. To put it another way, culture can only be meaningfully employed through comparison. The following is a list of the model's dimensions:

- **Power distance index (PDI):** This dimension expresses the extent to which society's less powerful individuals tolerate and expect unequal power distribution. The primary problem here is how a society deals with social inequality. People in civilizations with a high level of

Power Distance accept a hierarchical structure in which everyone has a place, and no more reason is required. People in societies with a low Power Distance attempt to equalize power distribution and demand reason for power imbalances.

- **Individualism vs. collectivism (IDV):** On the high end of this scale, individualism is characterized as a preference for a loosely woven social structure in which people are expected to look for just themselves and their immediate family. On the other hand, collectivism is a preference for a close-knit social structure in which people may expect their relatives or members of a certain ingroup to look after them in exchange for unquestioned allegiance. Whether people's self-image is defined as "I" or "us" reflects a society's perspective on this dimension.
- **Uncertainty avoidance (UAI):** The Uncertainty Avoidance dimension expresses how uncomfortable members of a society are with uncertainty and ambiguity. The central issue here is how a society deals with the fact that the future can never be predicted: should we try to control it or just let it happen? Countries with strong UAI adhere to rigid codes of belief and behavior and are intolerant of unconventional behavior and ideas. Weak UAI societies have a more relaxed attitude in which practice takes precedence over principles.
- **Masculinity vs. femininity (MAS):** The Masculinity aspect of this dimension denotes a societal desire for achievement, heroism, assertiveness, and material rewards for success. The general public is more competitive. On the other hand, Femininity represents a preference for cooperation, modesty, compassion for the vulnerable, and high quality of life. The general public is more consensus-oriented. In the corporate world, masculinity against Femininity is sometimes referred to as "tough versus delicate" cultures.
- **Long-term orientation vs. short-term orientation (LTO):** Every society must maintain ties to its past while dealing with present and future challenges. These two existential goals are prioritized differently by societies. Low-scoring societies, for example, prefer to uphold time-honored traditions and norms while viewing societal change with suspicion. Those with

a high-scoring culture, on the other hand, take a more pragmatic approach: they encourage thrift and investment in modern education as a means of preparing for the future. This dimension is known in the business world as "(short-term) normative versus (long-term) pragmatic" (PRA). Monumentalism versus Flexhumility is a term that is sometimes used in academic circles.

- **Indulgence vs. restraint (IND):** Indulgence represents a society that allows for the relatively free gratification of basic and natural human drives related to having fun and enjoying life. Restraint means a society suppresses the need for gratification and regulates it through strict social norms.

2.2 Elbow Method for Optimal Number of Clusters

The optimal number of clusters into which the data can be grouped is a crucial stage in any clustering technique (unsupervised learning). The Elbow Method is one of the most prominent approaches for determining the ideal value of k . When it comes to determining the k value, there is no simple solution. The elbow method is one of the methods. To begin, calculate the sum of squared errors (SSE) for a given value of k . SSE is the sum of the squared distances between the cluster's centroid and each member. Then, against the SSE graph, plot a k . We will notice that as k increases, SSE drops, and distortion diminishes. This approach aims to find the value of k at which the graph drastically decreases. Using the Python Sklearn module, we implement the suggested strategy using the K-Means clustering methodology. To find the optimal number of clusters, we must find the value of k at the "elbow," the point at which the distortion begins to decrease linearly. As a result, we find that the best number of clusters for the given data is vary between four to five (Figure 2.1).

2.3 Clustering and Representative Countries

After defining the optimal number of clusters ($k = 4$ or 5), k means clustering is adopted to find the cluster. Table 2.1 and 2.2 indicate the country with four and five clusters respectively. Also, Armenia, Malawi, Luxembourg, and Jordan are the four representative countries from each cluster nearest to the cluster center for $k=4$. Table ?? and ?? describe the cluster center with respect

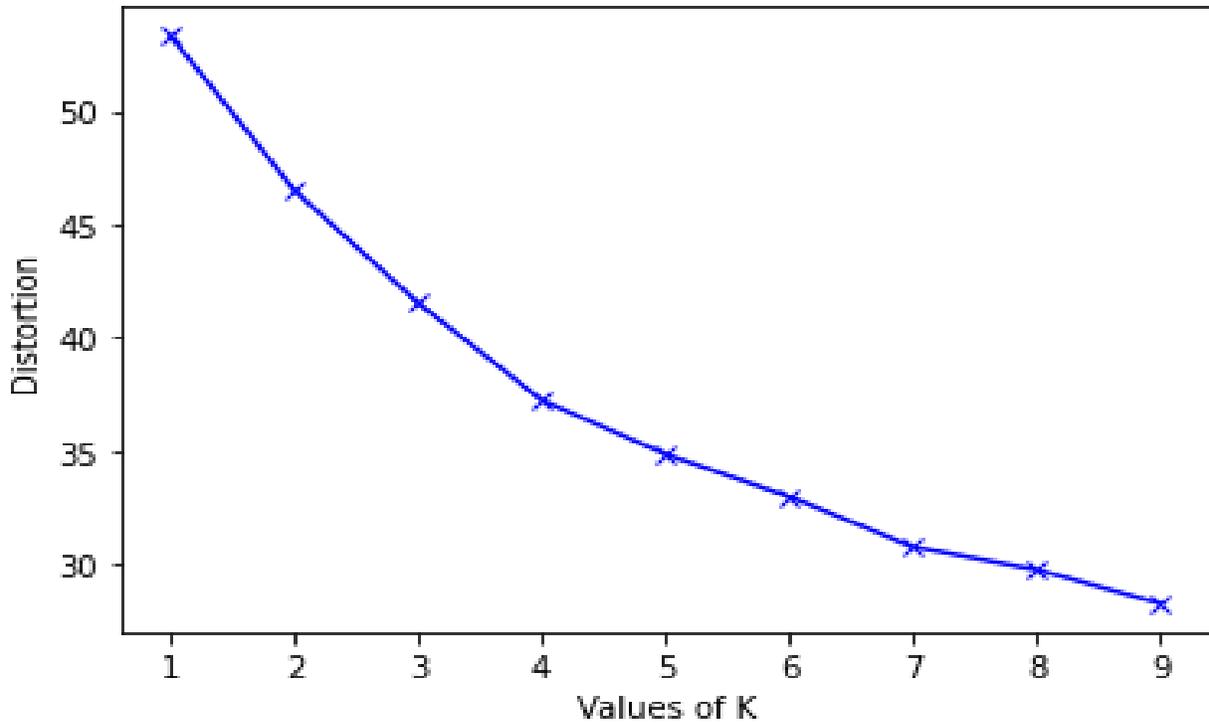


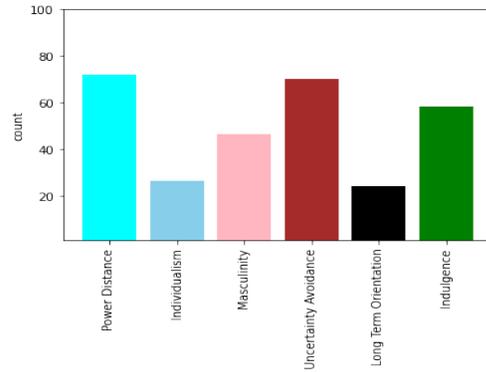
Figure 2.1: Using distortion for an elbow method.

to six cultural dimensions. Jordan, Czechia, Malawi, Armenia, and Canada are five representative countries from each cluster close to the center for $k=5$. From the expert judgment, we proceed with $k=5$ clustering as it represents the best cluster compared to $k=4$. We also verified our clustering result with K-Means++ (which is helpful when initializing the cluster centroids), and K-Means++ gives a similar outcome.

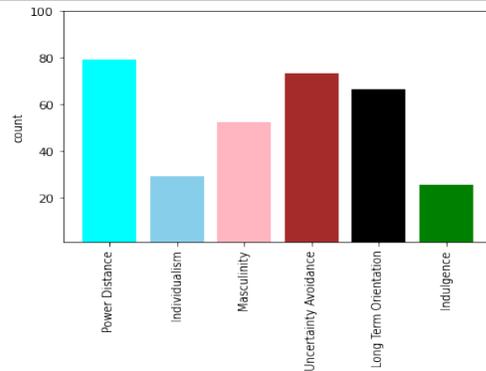
Table 2.2 shows the country belongs to five clusters and the six culture plot of the cluster's center. Also, the first cluster contains the maximum number of countries, whereas the minimum country is available under cluster 5. Culture dimension, power distance is minimum for cluster 5 and maximum for cluster 2. Individualism is maximum for cluster 5 and minimum for cluster 2. Masculinity is maximum for cluster 3 and minimum for cluster 5. Uncertainty avoidance is maximum for cluster 3 and minimum for cluster 5. Long-term orientation is maximum for cluster 2, minimum for cluster 4, and indulgence is maximum for cluster 5 and minimum for cluster 3. Thus, the table 2.2 shows a brief picture of the culture of 115 countries.

Table 2.1: Culture dimension for each center of clusters (k=4)

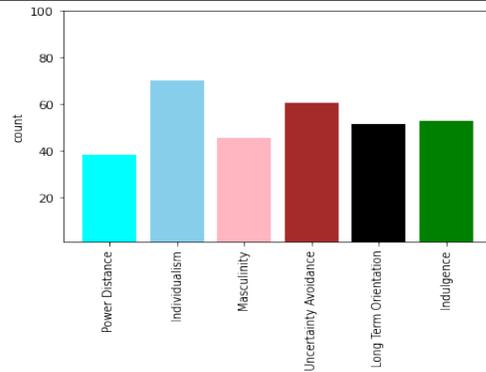
Albania, Armenia, Azerbaijan, Bangladesh, Belarus, Bosnia and Herzegovina, Bulgaria, China, Croatia, Czechia, Hong Kong, India, Indonesia, Japan, Kazakhstan, Moldova, Montenegro, North Macedonia, Pakistan, Poland, Romania, Russia, Serbia, Singapore, Slovakia, South Korea, Ukraine, Vietnam



Bhutan, Burkina Faso, Costa Rica, Ecuador, Egypt, Fiji, Guatemala, Honduras, Iraq, Jamaica, Kenya, Kuwait, Lebanon, Malawi, Morocco, Namibia, Nepal, Panama, Qatar, Senegal, Sierra Leone, Sri Lanka, Suriname, Tunisia, United Arab Emirates



Australia, Austria, Belgium, Canada, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Israel, Italy, Latvia Lithuania, Luxembourg, Malta, Netherlands, New Zealand, Norway, South Africa, Spain, Sweden, Switzerland, United Kingdom, United States



Algeria, Angola, Argentina, Bolivia, Brazil, Cape Verde, Chile, Colombia, Dominican Republic, El Salvador, Ethiopia, Georgia, Ghana, Greece, Iran, Jordan, Libya, Malaysia, Mexico, Mozambique, Nigeria, Paraguay, Peru, Philippines, Portugal, Sao Tome and Principe, Saudi Arabia, Slovenia, Tanzania, Thailand, Trinidad and Tobago, Turkey, Uruguay, Venezuela, Zambia

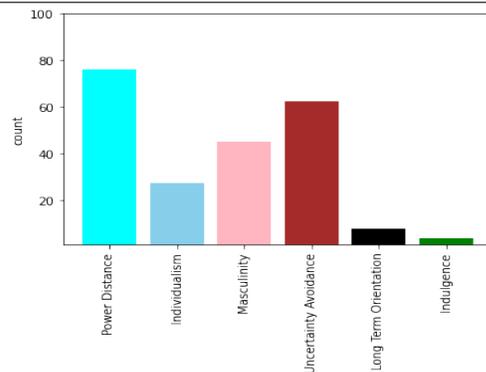
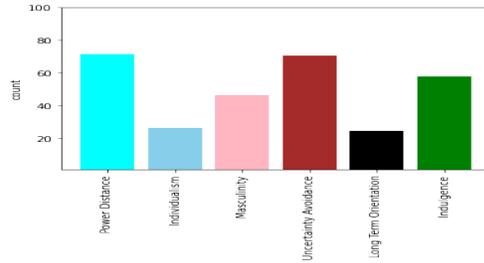
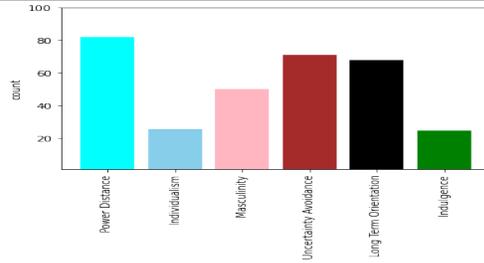


Table 2.2: Culture dimension for each center of clusters (k=5)

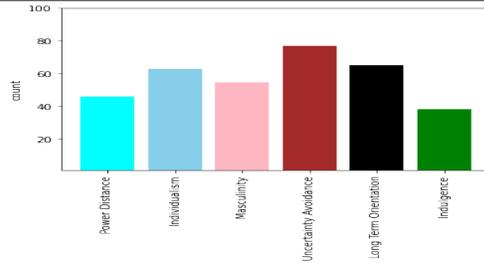
Cluster 1: Algeria, Angola, Argentina, Bolivia, Brazil, Cape Verde, Chile, Colombia, Dominican Republic, El Salvador, Ethiopia, Georgia, Ghana, Greece, Iran, Jordan, Libya, Malaysia, Mexico, Mozambique, Nigeria, Paraguay, Peru, Philippines, Portugal, Sao Tome and Principe, Saudi Arabia, Slovenia, Tanzania, Thailand, Trinidad and Tobago, Turkey, Uruguay, Venezuela, Zambia



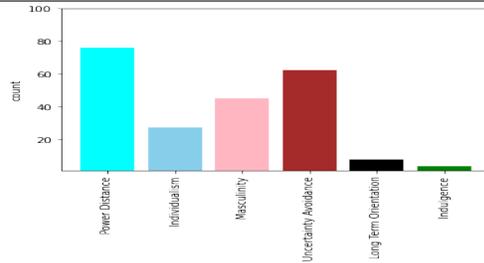
Cluster 2: Albania, Armenia, Azerbaijan, Bangladesh, Belarus, Bosnia and Herzegovina, Bulgaria, China, Croatia, Hong Kong, India, Indonesia, Kazakhstan, Moldova, Montenegro, North Macedonia, Pakistan, Romania, Russia, Serbia, Singapore, Slovakia, South Korea, Ukraine, Vietnam



Cluster 3: Austria, Belgium, Czechia, Estonia, France, Germany, Hungary, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Malta, Poland, Spain, Switzerland



Cluster 4: Bhutan, Burkina Faso, Costa Rica, Ecuador, Egypt, Fiji, Guatemala, Honduras, Iraq, Jamaica, Kenya, Kuwait, Lebanon, Malawi, Morocco, Namibia, Nepal, Panama, Qatar, Senegal, Sierra Leone, Sri Lanka, Suriname, Tunisia, United Arab Emirates



Cluster 5: Australia, Canada, Denmark, Finland, Iceland, Ireland, Netherlands, New Zealand, Norway, South Africa, Sweden, United Kingdom, United States

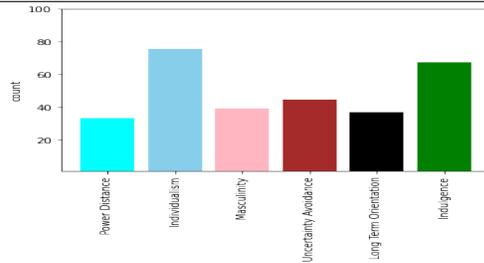


Table 2.3: Centers of the cluster (k=4)

Cluster Center						
Cluster	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	Long Term Orientation	Indulgence
1	71.57	26.57	46.46	70.54	24.63	57.69
2	79.5	28.61	52.18	72.68	67.32	25.39
3	38.26	69.89	45.30	60.30	51.56	52.63
4	75.92	27.16	45.24	62.24	7.68	3.56

Table 2.4: Centers of the cluster (k=5)

Cluster Center						
Cluster	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	Long Term Orientation	Indulgence
1	71.57	26.57	46.46	70.54	24.63	57.68
2	81.88	25.48	49.8	71.04	67.56	24.44
3	45.88	62.94	54.71	77.06	65.11	38
4	75.92	27.16	45.24	62.24	7.68	3.56
5	33.23	75.46	39.15	44.38	37	67.31

CHAPTER III

BAYESIAN PREDICTIVE MODELING OF COVID-19 USING CULTURE AND ECONOMICAL BEHAVIOURAL FEATURES

Bayesian predictive modeling is one of the most common and practical approaches for compartmental disease modeling see e.g. [33]. The goal of this part is to determine the courses of the epidemic and explain waves of COVID-19 from cultures using the Bayesian predictive modeling approach.

3.1 Literature Review for COVID

The SARS-CoV-2 virus causes Coronavirus Disease (COVID-19), an infectious disease that infects almost all of the world. Most of those infected with the virus will have mild to moderate respiratory symptoms and recover without needing medical attention. On the other hand, some will become highly unwell and require medical assistance. Severe sickness is more likely to strike the elderly and those with underlying medical disorders such as cardiovascular disease, diabetes, chronic respiratory disease, or cancer. COVID-19 can make anyone sick and cause them to get very ill or die at any age. As a result, the COVID-19 pandemic has emerged as one of the most challenging public health emergencies to which international attention has been drawn. This literature review provides a summary of previous and continuing research on COVID-19.

3.1.1 Modeling COVID-19

From the beginning of the pandemic, researchers tried to figure out scenarios and several parameters that directly or indirectly affect COVID-19 using various disease modeling, machine learning, and deep learning approach [28]. The IHMR COVID-19 forecasting team used COVID-19 case and mortality data from 1 February 2020 to 21 September 2020, as well as a deterministic

SEIR (susceptible, exposed, infectious, and recovered) disease compartmental framework, to model possible trajectories of the severe acute respiratory syndrome of coronavirus 2 infections and the effects of non-pharmaceutical interventions in the United States from 22 September 2020 to 28 February 2021 at the state level[3]. They visualize the social distancing mandates and levels of mask use scenario using SEIR model and find that achieving universal mask use (95%) in public could be sufficient to ameliorate the worst effects of epidemic resurgence in many states. They utilize the SEIR model to illustrate the social distancing mandates and levels of mask wear and conclude that obtaining universal mask use (95 percent) in public could be enough to mitigate the worst consequences of epidemic revival in many states. The Traditional SEIR epidemic model with specific dynamic compartments and epidemic parameters of COVID-19 shows that COVID-19 spreads in an age-heterogeneous community, and closures, mobility restrictions, and social distancing improve the pandemic control, especially in a sustainability context[44].

In the wake of the COVID-19 pandemic, governments worldwide are faced with the task of devising personalized epidemic control methods that provide reliable health protection while allowing for societal and economic activities. In, [18] the authors propose an extension of the epidemiological SEIR model to enable a detailed analysis of commonly discussed tailored measure of epidemic control. Along with that they introduce groups into the SEIR model that may differ both in their underlying parameters as well as in their behavioral response to public health interventions.

3.1.2 Culture, Behavior and Economy in COVID-19

After the World Health Organization declared the COVID-19 outbreak a public health emergency of international concern (PHEIC), almost all countries began implementing a variety of Nonpharmaceutical Interventions (NPIs) such as contact tracing, social distancing, mask-wearing, self-isolation, school closures, business closures, and countrywide lockdowns at various levels of rigor to halt the disease's spread. Several NPIs can be used by public health professionals at the start of a pandemic to reduce disease transmission until a viable vaccine or antiviral treatment becomes available. Individuals are subjected to constraints as a result of public health measures, therefore it's critical to understand how they react and if they're likely to follow or breach new laws. Measures

might possibly have a significant impact on the transmission of the pathogen. However, a variety of issues, including human choice and self-interest over charity, can stymie Non-pharmaceutical Interventions(NPI) effectiveness and disease control efforts.

Lockdowns and self-isolation (self-quarantine), for example, can be quite efficient in reducing transmission, but they might have socioeconomic and emotional consequences for the entire population [10]. Extensive isolation has been linked to adverse outcomes in a variety of people, including children and adolescents, immigrant workers, and adults [14, 11]. Changes in food patterns, sleep disruptions, depression, and anxiety symptoms were also reported by children [20]. Mental health concerns, anxiety, stigma, despair, alcohol-related injury, and domestic violence were all reported by adults [11, 31].

Engagement with quarantine, lockdown, and compliance with public health directives about personal protective practices is influenced by various demographic, social, and psychological factors. Perceptions of vulnerability to infection, the severity of infection, the effectiveness of ongoing public health efforts, and the ability to do the activity safely (self-efficacy) are among the factors [8]. The belief of a lesser risk of disease or having fewer risk factors is one of the fundamental causes of non-adherence to quarantine and self-isolation documented in the research literature [53]. Psychological exhaustion has also been considered a cause of NPI non-compliance [46]. While fear may challenge cultural and social influences, economic hardship suffered by some groups, particularly minorities in some places, plays a role in the human decision [15]. This could explain why minorities in the United States, Australia, Canada, and the United Kingdom have disproportionately high COVID-19 incidence and mortality [7]. Migrant workers in low-income nations are an economically vulnerable demographic[50]. As a result, cultural factors can significantly impact NPI uptake and adherence, as well as disease transmission and mortality.

Individual compliance is aided by infrastructure, resources, stockpiles, inter-pandemic planning, authoritative source communication, and the country's capacity. People fearful of developing a viral infection will employ the best hygiene precautions, wear masks, avoid crowded places, and exercise social distance. While such efforts can help slow the spread of viral infections like COVID-19, they

cannot protect the public. Public health directives aimed at reducing population-level risk factors and disease transmission are strongly associated with the idea of everyone doing their part to achieve high compliance by implementing the most nuanced hygiene measures. Indeed, GDP gives a gauge of the pre-existing infrastructure for maintaining and enforcing law and order, regulating economic activity, and providing public goods during a protracted pandemic wave [4]. Many countries in less developed parts of the world lack this capability, making them more sensitive to system shocks such as pandemics, which can halt economic growth and cut GDP [34].

3.2 Game Theory and the Swiss Cheese Model

The Swiss Cheese Model was created by British psychologist James Reason two decades ago to understand how breakdowns in complex systems emerge [41]. Multiple defenses can be in place, according to his model, to protect persons against threats, but these defenses also have intrinsic flaws. Multiple protections or obstacles are like Swiss cheese slices with a lot of transitory holes in them. In most cases, having holes in any one 'slice' does not result in a terrible consequence. If the gaps in several layers line up so that an accident opportunity can travel across the layers, it provides for hazard exposure and casualties. Active failures and latent conditions both contribute to the weaknesses in in-built defenses. Almost of adverse outcomes are caused by a combination of these two causes.

Regarding pandemic defenses or safeguards, the Swiss Cheese Model demonstrates that there are two degrees of protection: personal and interpersonal precautions. When applying the Swiss Cheese Model to COVID-19, the early NPIs such as social distancing, self-isolation, and lock-downs are the pandemic barriers that can collapse. We refer to these NPIs as "social isolation" barriers in the model. The topic of this chapter is the human behavior of social isolation decision-making during the pandemic and its impact on socio-economic growth. Using a Swiss Cheese Model method, we build a conceptual framework to examine the problem by combining evolutionary game theory, an economic growth model, and a deterministic disease transmission model.

3.3 COVID-19 Disease Model

We develop a vector values compartmental model of the type SEAIHRD to model country-wise COVID-19 compartment are Susceptible-Exposed-Asymptomatic-Infected-Hospitalized-Recovered-Dead 3.1. Arrows are labeled by the rates of transmission of individuals in each compartment. SD in model is social distance strategy with proportion x and NSD is no-social distance strategy with proportion $1 - x$. Rational agents compare the difference between benefits of the other strategy to their and if positive they switch strategy. In the model the parameter is defined as: disease transmission rate (β), rate of leaving exposed state (α), probability of becoming symptomatic (p), recovery rate of asymptomatic (μ_A), recovery rate of infectious (μ_I), recovery rate of hospitalized (μ_r), rate of hospitalization (μ_H), death rate from disease (μ_D), imitation rate (r), cost of infection (c_1), fear of death (c_2), cost of government strict policy (c_3), sensitivity to relative economic loss (c_4), pandemic fatigue rate (c_5), cost of hospitalization (c_h), investment rate (σ), elasticity (γ), depreciation rate (δ), initial per-capita GDP (K_0), and fraction of labor working with social isolation (q). The system of the ODE represent the continuous time in-flow and out-flow of those compartment. We use imitation dynamical equation from behavioral game theory to model rational decision making and include a pandemic fatigue rate.

The disease model depends on solving deterministic model of ordinary differential equation as specified below [40]. There are seven compartments (Figure 3.1), E = exposed but not infectious, A = infected but asymptomatic, I = infected and symptomatic, H= isolated or hospitalized, D = dead, and R = recovered. Also, x and k represent proportion of those who maintain social isolation

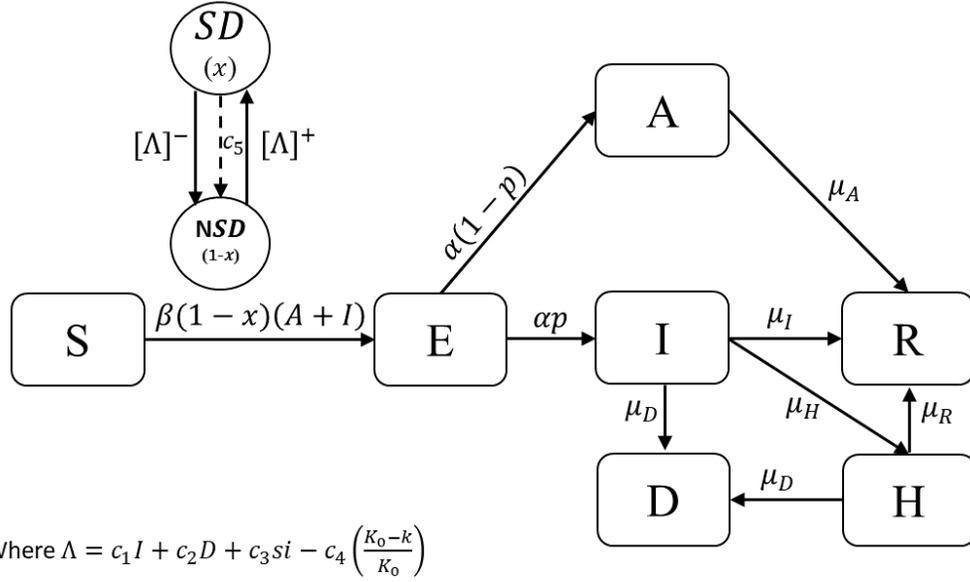


Figure 3.1: Schematic illustration of the COVID-19 SEAIHRD model.

and k is the per-capita GDP ($k = GDP/N$).

$$\frac{dS}{dt} = -\beta(1-x)(A+I)S \quad (3.1)$$

$$\frac{dE}{dt} = \beta(1-x)(A+I)S - \alpha E \quad (3.2)$$

$$\frac{dA}{dt} = \alpha(1-p)E - \mu_A A \quad (3.3)$$

$$\frac{dI}{dt} = \alpha p E - \mu_I I - \mu_H p_{50} I - \mu_D I \quad (3.4)$$

$$\frac{dH}{dt} = \mu_H p_{50} I - \mu_R H - \mu_D H \quad (3.5)$$

$$\frac{dD}{dt} = \mu_D I + \mu_D H \quad (3.6)$$

$$\frac{dR}{dt} = \mu_A A + \mu_I I + \mu_R H \quad (3.7)$$

$$\frac{dx}{dt} = rx(1-x)(c_1 I + c_2 D + c_3 s i - c_4 (K_0 - k)/K_0) - c_5 x \quad (3.8)$$

$$\frac{dk}{dt} = \sigma((S+A+R)((1-x)+qx))^{\gamma} k^{1-\gamma} - \delta k - c_h H \quad (3.9)$$

where

$$\log(c_j) = \alpha_{0j} + \sum_{i=1}^6 \alpha_{ij} X_i$$

for $j = 1, 2, 3, 4, 5$, α_{0j}, α_{ij} is the log linear coefficient of c_j , and $(X_{1k}, X_{2k}, \dots, X_{6k})$ is the input vector of six culture dimension for each country k for $k = 1, 2, \dots, 115$. Here, the initial condition is the number of cases and their values from populations and Runge-Kutta method is used to solve the ODE. S=Susceptible, E=Exposed but not infectious, A=Asymptomatic, I=Infected and Symptomatic, H=Isolated or Hospitalized, D=Dead, and p_{50} is the proportion of those 50 years or older in the population, and R=Recovered. $\beta(1 - x)$ is effective transmission rate which is dependent on x (proportion of the people who maintain social isolation).

In the ODE model, we use the dynamical population equation as specified by the Equation 3.8 to model the dynamic changes of x where people choose social isolation comparing the economic loss and risk of infection and fear of death. But people who choose social isolation may break it after an average of $1/c_5$ days due to fatigue raise from social isolation. We assume that c_5 (pandemic fatigue rate) is dependent on the six cultural dimension specified by Hofstede more especially long term orientation, individualism, and uncertainty avoidance. On the other side, we think the economic loss is more related to long-term orientation, masculinity, and socioeconomic status. Equation 3.9 (economic growth and decline) is modeled using the Solow economic model of per capita GDP combining with per-capita GDP combining with Cobb-Douglas functional form of production.

3.4 Data Description

3.4.1 COVID Cases

We use daily data from 115 countries from the beginning of January 2020 to the end of January 2021 (360 days for each country), depending upon the land and availability of data. We use the daily numbers reported on ourworldindata web page for new cases and death [42]. The number of hospitalization data was also obtained from ourworldindata, but these data appear to be available for only 28 countries. Lockdown stringency index data was collected for 115 countries from www.statistica.com

Table 3.1: Data sources

Types of Data	Sources
COVID cases, mortality, hospitalization	https://ourworldindata.org/
Lockdown stringency index	https://www.statista.com/
Hofstede’s cultural dimension	https://www.hofstede-insights.com/
Per-capita GDP	https://data.un.org/
Population Size	https://data.worldbank.org/

3.4.2 GDP and Population

The GDP and population data was obtained from the UN and Worldbank GDP and population database, respectively, from 2007 to 2017.

3.4.3 Lockdown Stringency Index

The Stringency Index is calculated using nine metrics:

- School closures
- Workplace closures
- Public event cancellations
- Restrictions on public gatherings
- Public transportation closures
- Stay-at-home requirements
- Public information campaigns
- Internal movement restrictions
- International travel controls

On any given day, the index is calculated as the mean of the nine metrics, each of which has a value between 0 and 100. A higher score indicates a more stringent response (i.e., 100 =

most stringent). If policies differ at the subnational level, the index is displayed as the strictest sub-response region's level. Because government policies vary depending on vaccination status, a stringency index is calculated for three categories: those who are vaccinated, those who are not vaccinated, and a national average weighted by the proportion of vaccinated people.

3.5 Statistical Data Model

- **Likelihood Function:** The likelihood function describes the joint probability of observed data as a function of the statistical model's parameters. Here, likelihood indicates the probability of observing infected, death, and total number of hospitalization case data given the proposed model. In this study, we adopt the Gaussian distribution as the likelihood for infected, dead, and hospitalization cases.

- Gaussian likelihood function

$$L_j(\Theta, \sigma_{obs,j} | \{\tilde{y}_{t,i,j}\}) = \prod_{i=1}^{115} \prod_{t=1}^T \left(\frac{1}{\sqrt{2\pi}\sigma_{obs,j}} \exp\left(-\frac{(\tilde{y}_{t,i,j} - y_{t,i,j}(\Theta))^2}{2\sigma_{obs,j}^2}\right) \right)$$

of the parameters $\Theta = (\beta, \alpha, p, \dots, \alpha_{6,5}), \sigma_{obs,j}$, for $j = 1, 2, 3$ given the observed proportions $\tilde{y}_{t,i,j}$ of infected ($j = 1$), dead ($j = 2$), and hospitalized ($j = 3$) people in country i .

- where $\tilde{y}_{t,i,j}$ are the observations and $y_{t,i,j}(\Theta)$ are the solutions of the differential equation model representing the mean values.
- That is $y_{t,i,1}(\Theta) = I_i(t)$, $y_{t,i,2}(\Theta) = D_i(t)$, and $y_{t,i,3}(\Theta) = H_i(t)$ for $i = 1, \dots, 115$, and $t = 1, \dots, T$.

- **Bayesian Approach:** We adopt Bayesian inference which calculate the posterior inference of parameters by updating the prior beliefs given by a data-driven likelihood.
- **Priors Selection:** Prior subject matter knowledge is encoded in parameter estimation using parameter prior distributions. The differential equation model parameters, as well as their

Table 3.2: Priors and ranges of parameters

Parameters	Fixed value/Prior	Range	Reference
Disease transmission rate (β)	0.78	0.2-4	[38][49][21][45]
Rate of leaving exposed state (α)	$\frac{1}{7}$	$\frac{1}{7}$ - $\frac{1}{5}$	[26][52]
Probability of becoming symptomatic (p)	0.75	0.5-0.85	[35][48][24]
Recovery rate of asymptomatic (μ_A)	$\frac{1}{14}$	$\frac{1}{14}$ - $\frac{1}{7}$	[54][48]
Recovery rate of infectious (μ_I)	$\frac{1}{14}$	$\frac{1}{14}$ - $\frac{1}{7}$	[54][48]
Recovery rate of hospitalized (μ_R)	0.78	0.7-0.88	[30][19][32]
Rate of hospitalized (μ_H)	$\frac{4.6}{10^5}$		[16]
Death rate from disease (μ_D)	$\frac{2.3}{100}$		[25]
Imitation rate (r)	Lognormal(0,1)		[55]
Fraction of labor working with social isolation (q)	0.279	0.2-0.35	[36][13]

associated priors and MCMC parameters, are declared. The prior settings for the model parameter are listed in Table 1. In this study, we consider uniform prior for β and p and lognormal distribution for $\alpha, \mu_A, \mu_I, \mu_R, \mu_H, \mu_D, r, q$.

Choosing $\alpha_{ij} \sim \text{Laplace}(0, \tau^2)$ results in a weighted- L_1 regularization that seeks to minimize

$$\sum_{i,j,t} \left(\frac{(\tilde{y}_{t,i,j} - y_{t,i,j}(\{\alpha_{ij}\}))^2}{2\sigma_{obs,j}^2} + \lambda \sum_{i,j} |\alpha_{ij}| \right)$$

The parameter $\lambda = 2^{1.5}/\tau$. The maximum *a posteriori* probability (MAP) estimate is a least absolute shrink and selected.

Predictive Posterior: Posterior predictive distribution at a test point x follows Gaussian distribution [37]:

$$p(y|x, \mathcal{D}, \sigma^2) = \int \mathcal{N}(y|\mathbf{x}^T \mathbf{w}, \sigma^2) * \text{priors} d\mathbf{w} \quad (3.10)$$

- **Monte-Carlo Cross-validation:** The model evaluation is one of the most critical steps in the machine/statistical learning approach. There is a dilemma regarding which data should be used for training and testing, and the solution is to select data for training and testing in an unbiased way called cross-validation. Cross-validation uses all points for both in an

iterative method in training or testing, and it solves the potential over-fitting and data leakage problem. We randomly split the entire data set into 70% training and 30% testing set. While splitting the data set, we ensure the country with hospitalization data remains on the training data set. After separating, training data contain 80 countries, and testing data includes 35 countries. And we use the mean square loss function as our error function to evaluate the model performance.

3.6 Model Simulation

We simulated the model using the ordinary differential equation function ode using Runge-Kutta hybrid 4 and 5 methods in R. A total number of 360 days is considered to simulate the model. We assume that the epidemic started with 100 exposed, 50 asymptomatic, and 30 infected individuals in a population size of 11,000,000.

3.7 Results

The main goal is to parameter inference for a SEAIHRD (Susceptible-Exposed-Asymptomatic-Infected-Hospitalized-Recovered- Dead) model described in section 3.3. This model has also further understood the influence of culture on COVID-19 courses of the epidemic. The analysis results from an R package called deBInfer: Bayesian inference for dynamical models of biological systems in R.

3.7.1 MCMC Inference

We use the `setup_debinfer` function of deBInfer to collate the declared parameters and run the MCMC estimation using `de_mcmc`. We run the MCMC for several 1000 iterations as it takes too long. But for better convergence, we will run it for more iteration and possibly put the system into parallel computing GPU. For complex dynamical systems with a large number of data, running the system's parallel will be a good option as MCMC takes longer. In the section 3.7.2 and 3.7.4 we will discuss about MCMC model diagnostic and final posterior distribution of 1000 iteration respectively.

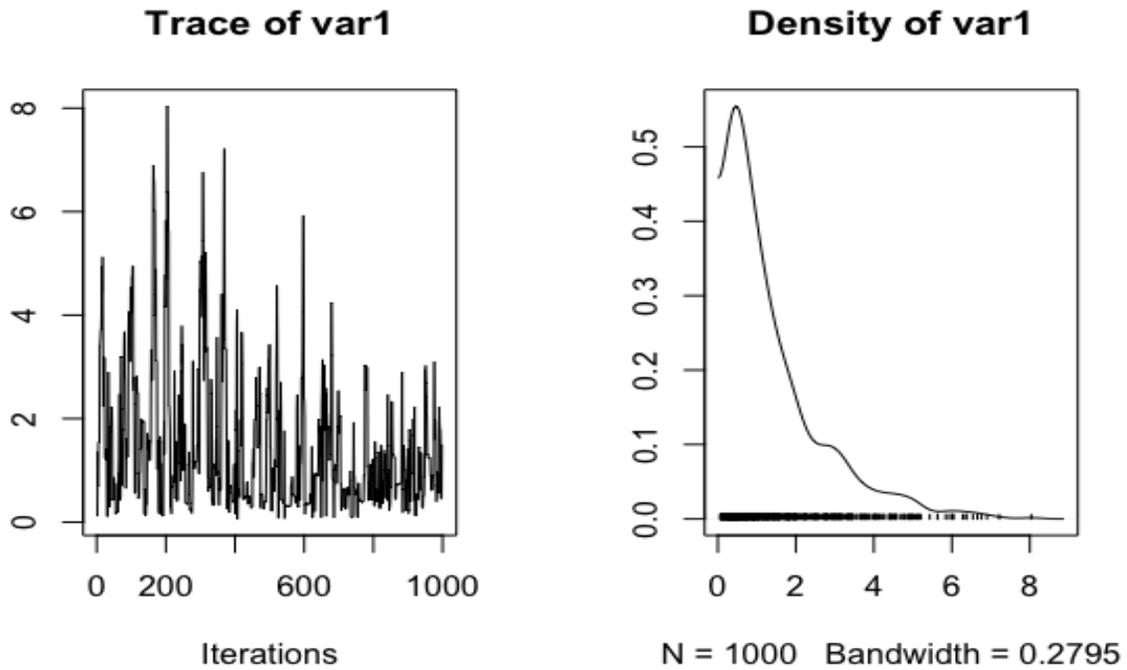


Figure 3.2: MCMC trace and density plot of parameter r .

3.7.2 Trace Plots to Check MCMC Mixing

Trace plots depict the chain's progression over time. The ideal trace plot has little "stickiness" and a lot of movement around the mode from iteration to iteration. Trace plots with a clear trend indicate that the MCMC sampler has not yet converged and should be run for a longer period of time. Extreme stickiness in trace plots indicates that the chain has not yet explored the entire parameter space and needs to run longer.

In this dynamical system, there are 39 parameters in total which we try to estimate using Bayesian inference. And, Figure 3.2, 3.3, and 3.4 shows the MCMC traces of some sample parameter r , α_{01} , and α_{02} and the MCMC chain progress over time. But for better convergence, it needs to run for a longer iteration, which will be the study's next step.

3.7.3 Autocorrelation and Effective Sample Size to Check Correlation Diagnostics

Independent (Monte Carlo) draws have no autocorrelations. Dependent draws (MCMC) have non-zero, usually positive autocorrelations. Large autocorrelations indicate that the chain is not

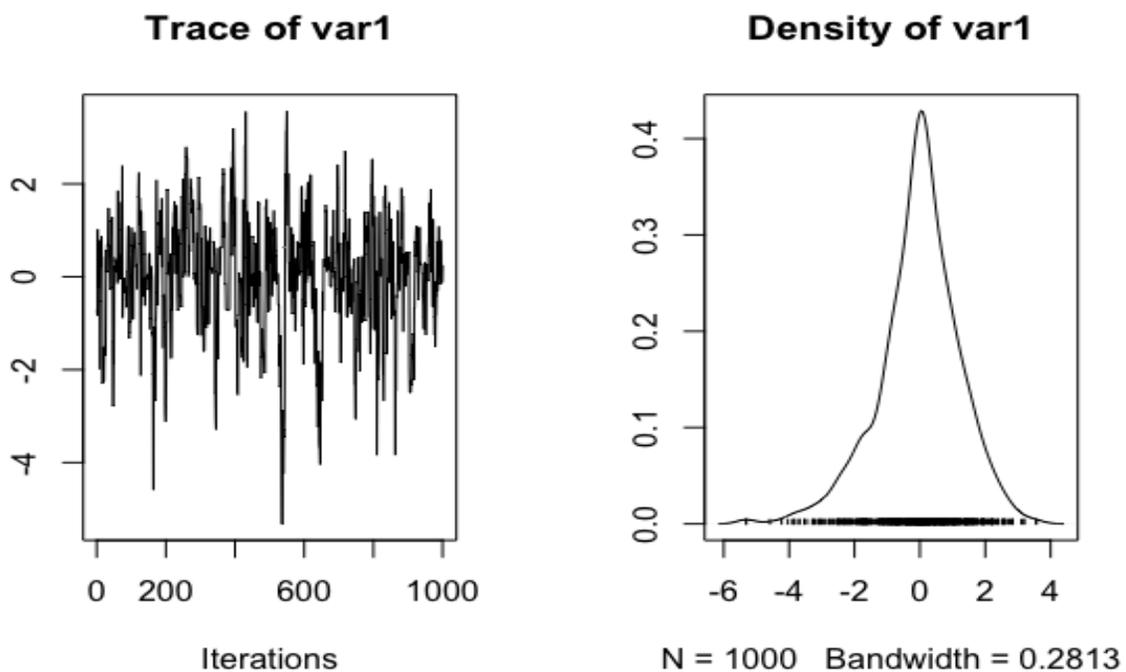


Figure 3.3: MCMC trace and density plot of parameter α_{01} .

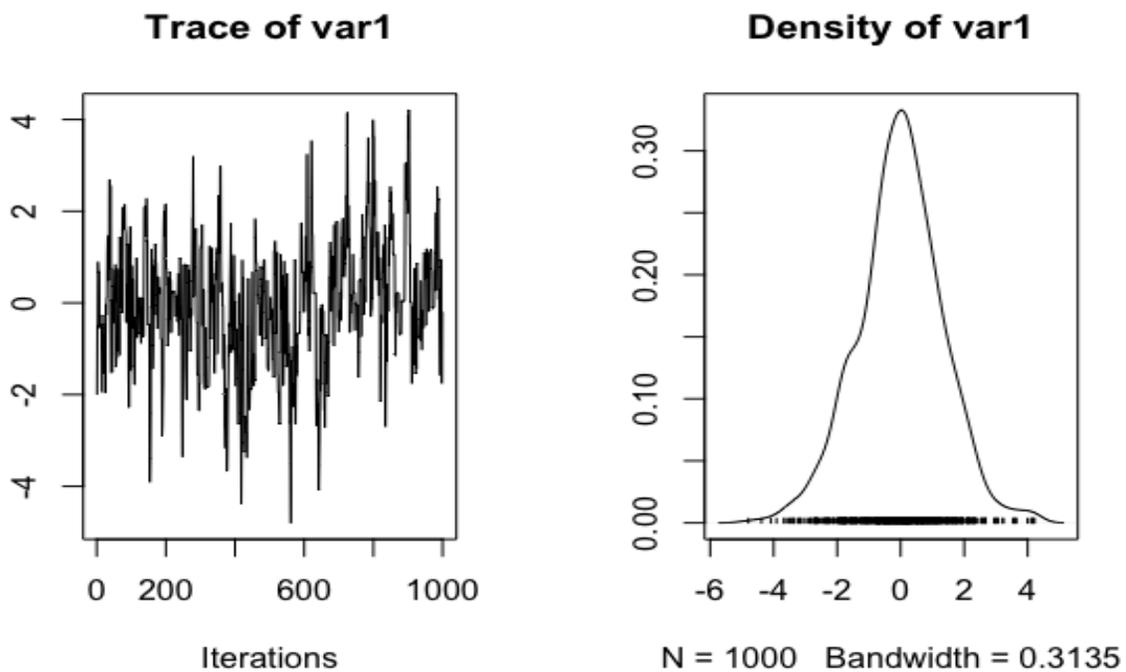


Figure 3.4: MCMC trace and density plot of parameter α_{02} .

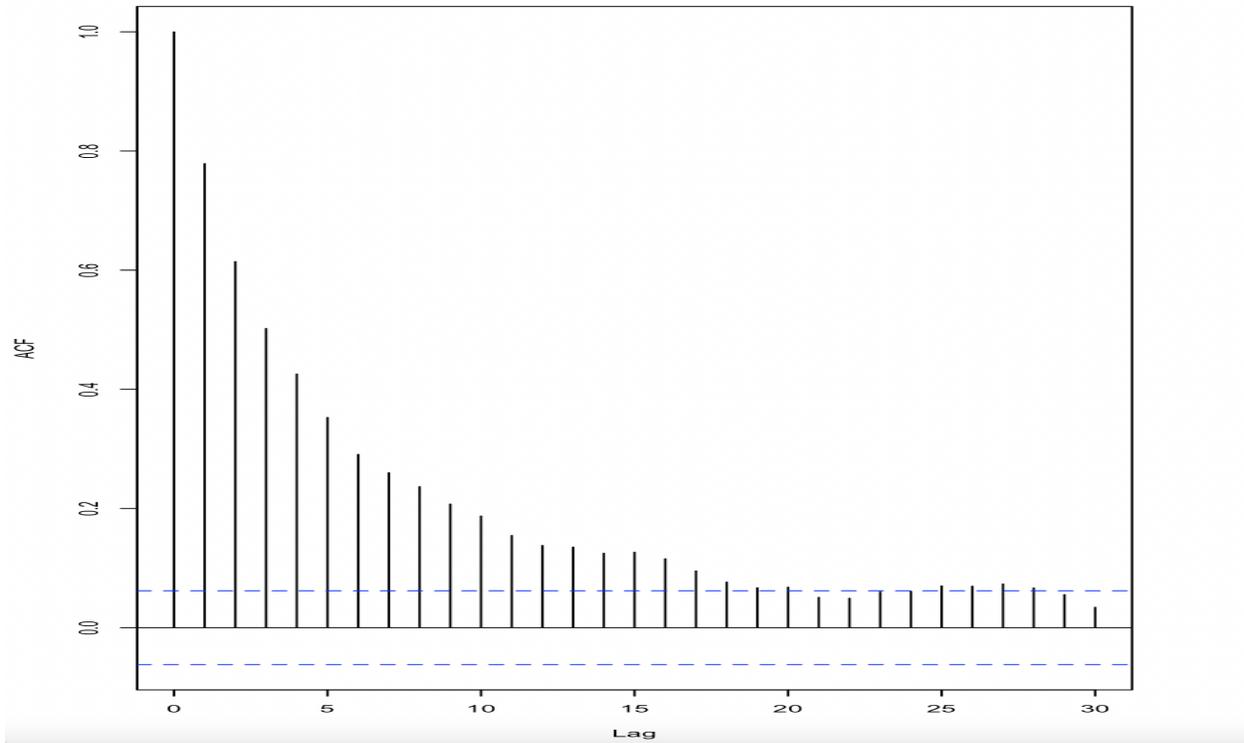


Figure 3.5: Auto-correlation plot of parameter α_{02} .

mixing well, implying that the chain has not explored the entire space of the posterior distribution (has not converged). When the parameters are highly correlated, large autocorrelations occur frequently in multi-parameter MCMC algorithms.

Figure 3.5 shows the autocorrelation plot of only one parameters α_{02} and there exists autocorrelation. To solve the problem, it is required to run the MCMC for a long iteration and thinning so that MCMC mix properly and there will be minimum autocorrelation among the parameter. We also perform the effective sample size test, which varies from 86-150 depending on the parameter, which also directs us to run MCMC for longer iterations for better convergence.

3.7.4 Posterior Distributions

The Bayes theorem is used to calculate the conditional distribution for unobserved quantities given data, also known as the posterior distribution. After seeing the data, the posterior distribution expresses our uncertainty about the parameter.

Understanding the mechanisms underlying the COVID disease model and ultimately predicting

their behaviors in a changing environment requires overcoming the gap between mathematical models and experimental or observational data. We believe Bayesian inference provides a powerful tool for fitting dynamical models and selecting between competing models. Posterior distribution of all the parameter is shown in Figure 3.6. It indicates that the culture parameter does not have high influence on the courses on epidemics except for some. But to come to the final decision, checking the MCMC convergence with better mixing and putting the system into parallel computing is required.

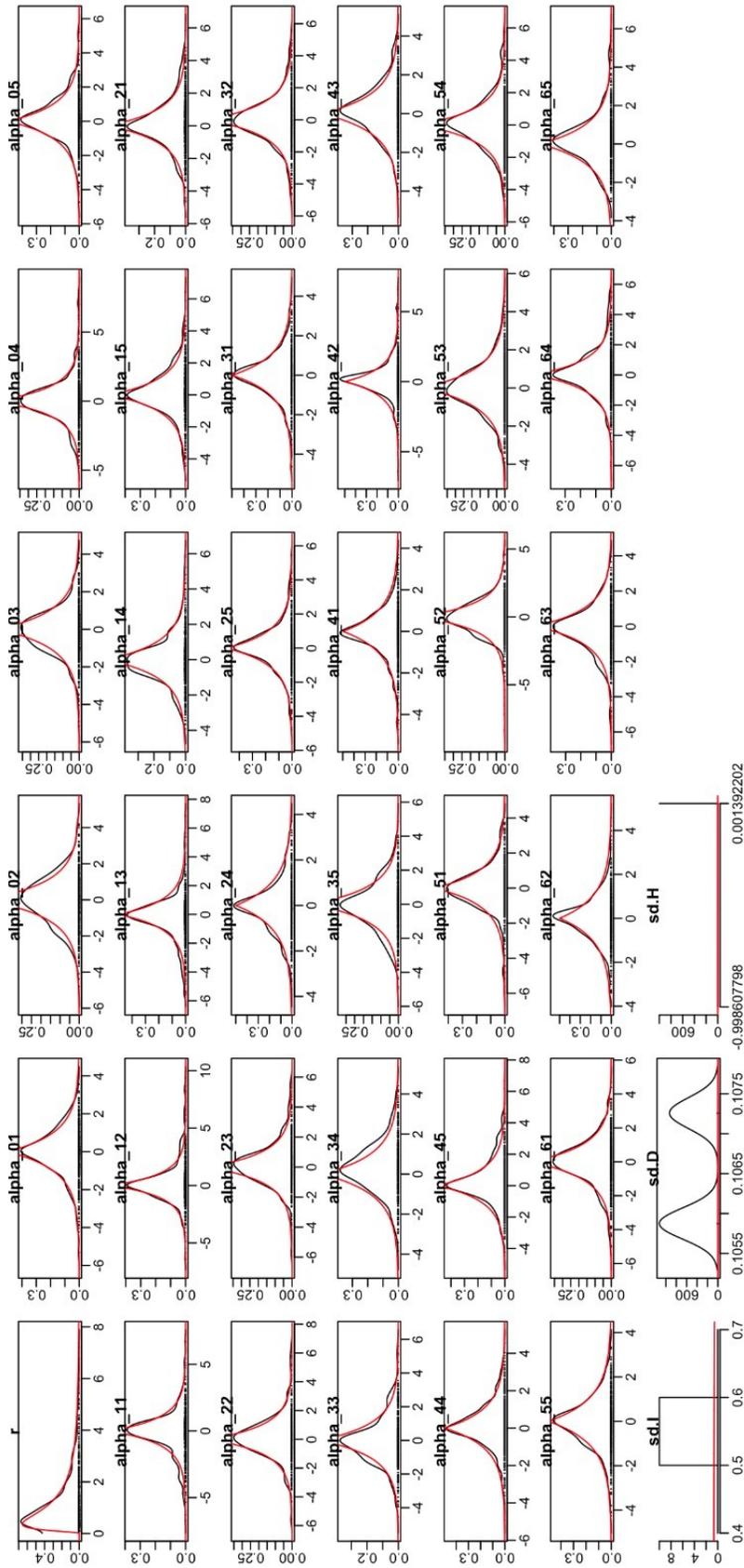


Figure 3.6: Posterior distributions of parameters.

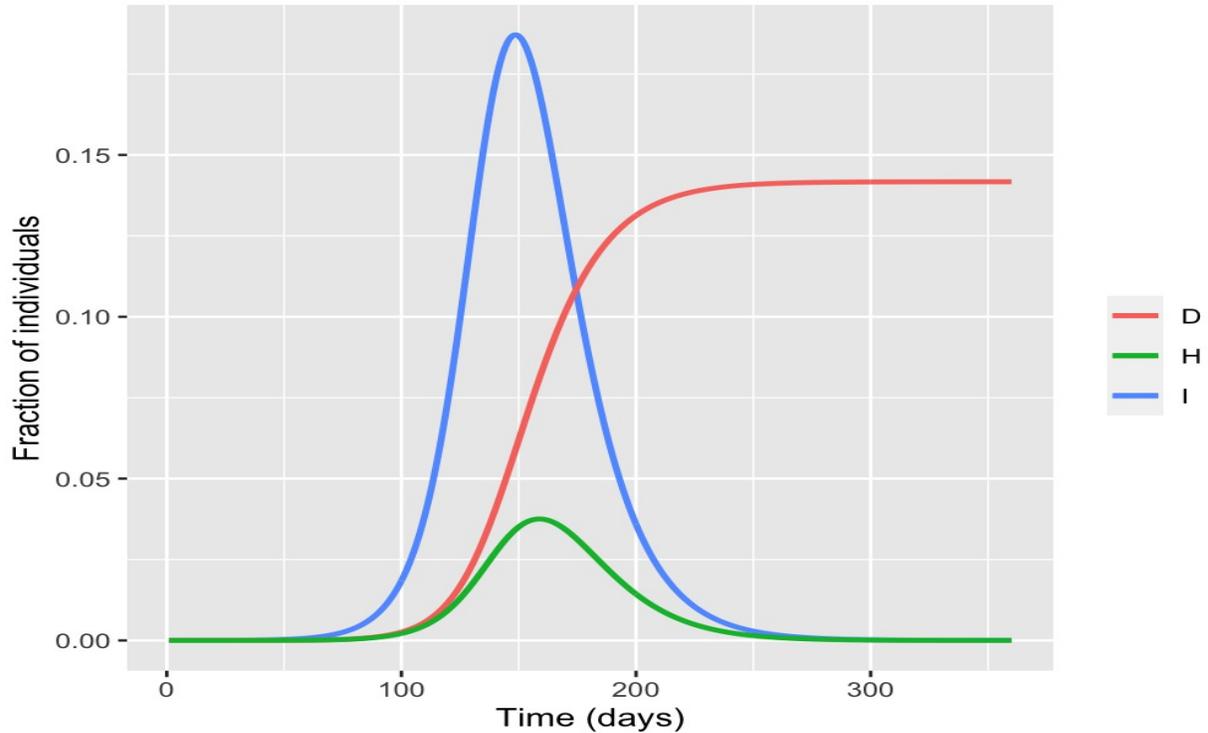


Figure 3.7: Simulation of the disease prevalence.

3.7.5 Simulation and Discussion

Simulations were performed for 360 days with the mean parameter obtained from the posterior distribution of the first cluster described in Section II and shown in Figure 3.7. According to the figure, the prevalence of the infected and hospitalization fraction of individuals shows waves with varying times. The fraction of people's death increased as the time increased, and after 200 days, it became constant. Preliminary estimation shows that the culture dimensions are not predictive of courses of COVID-19 epidemics. For a better result, it is required to run the MCMC for a longer iteration so that the parameter value converges. Upon checking the MCMC diagnostic, we may decide upon the convergence of the parameter and need to use appropriate Bayesian inference methods like Bayes Factor. Section 3.7.6 discuss the computational tools' disadvantage for solving dynamic systems. The next step will be running the MCMC for a longer iteration in parallel computing GPU to get better convergence.

3.7.6 Drawback of R for Complex Dynamical System

The MCMC sampler is written in R package `deBinfer`, which makes it significantly slower than samplers written in compiled languages, such as those found in underlying packages like Stan. The computational bottleneck for purely R-based inference is numerically solving the DE model. However, even for relatively simple models, using compiled DE models can result in a 5- to 10-fold speedup of the inference procedure. Furthermore, the `debinferMCMC` algorithm is not adaptive and must be tuned manually. Finally, sampling with the Metropolis-Hastings MCMC algorithm can be inefficient in strong parameter correlations. In these cases, alternative approaches such as Hamiltonian MC or particle-filtering methods may provide more efficient means of parameter estimation in ODEs. Nonetheless, the package can solve real-world problems in minutes to hours on current desktop hardware, which is sufficient for many applications and provides flexible inference for ODE and DDE models.

CHAPTER IV

CONVOLUTIONAL NEURAL NETWORK TO DETERMINE CULTURE FROM COVID-19

The chapter aims to develop a Convolutional Neural Network(CNN) that can identify the culture of 115 countries from COVID -19 cases (input). One of the key challenges is that the input data on COVID-19 cases and mortality are tabular times series data. So, we proposed a new method to transform the times series COVID-19 data into an $m \times m$ matrix and finally transform the matrix into the figure. The transformation matrix reflects the changes (increase/decrease) of COVID -19 new cases and mortality. This method of altering the tabular times series data into a figure is a novel contribution that can be applied to any tabular data concerned with identifying the changes. Transforming the data into a matrix and ultimately into a figure has several advantages, especially when talking about deep learning algorithms like convolutional neural networks. After the invention of AlexNet in 2012, convolutional neural networks have changed the field of image recognition and computer vision [27]. Different convolutional neural networks(ResNet, VGG, GoogLeNet, and Inception-v4) show a state of performance (human level) in image recognition [47, 23, 51]. Figure 4.1 shows top-1 single-crop validation accuracy's for top-scoring single-model architectures, as well as top-1 one-crop accuracy versus the number of operations required for a single forward pass. There are currently a few research studies showing that tabular data can transform into images and fed into convolutional neural networks to extract the underlying pattern in the data.

4.1 Tabular Data to Image Transformation

The raw data for COVID-19 is collected in time series format, and we consider only the first 300 days because vaccines and other measures may add to the system after that. As we aim to predict the culture effect on the COVID-19 pandemic, the COVID new cases data of more than 300

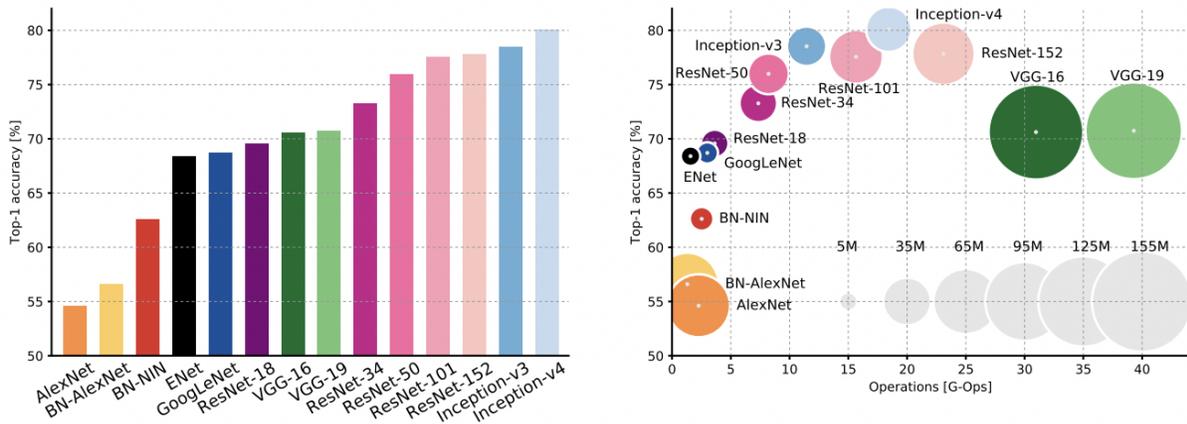


Figure 4.1: State of art CNN performance analysis: Top1 accuracy vs. network. (left) and Top1 accuracy vs. operations, size (right) [12].

days is excluded from the analysis. Figure 4.2 represent the distribution of new cases for first 300 days.

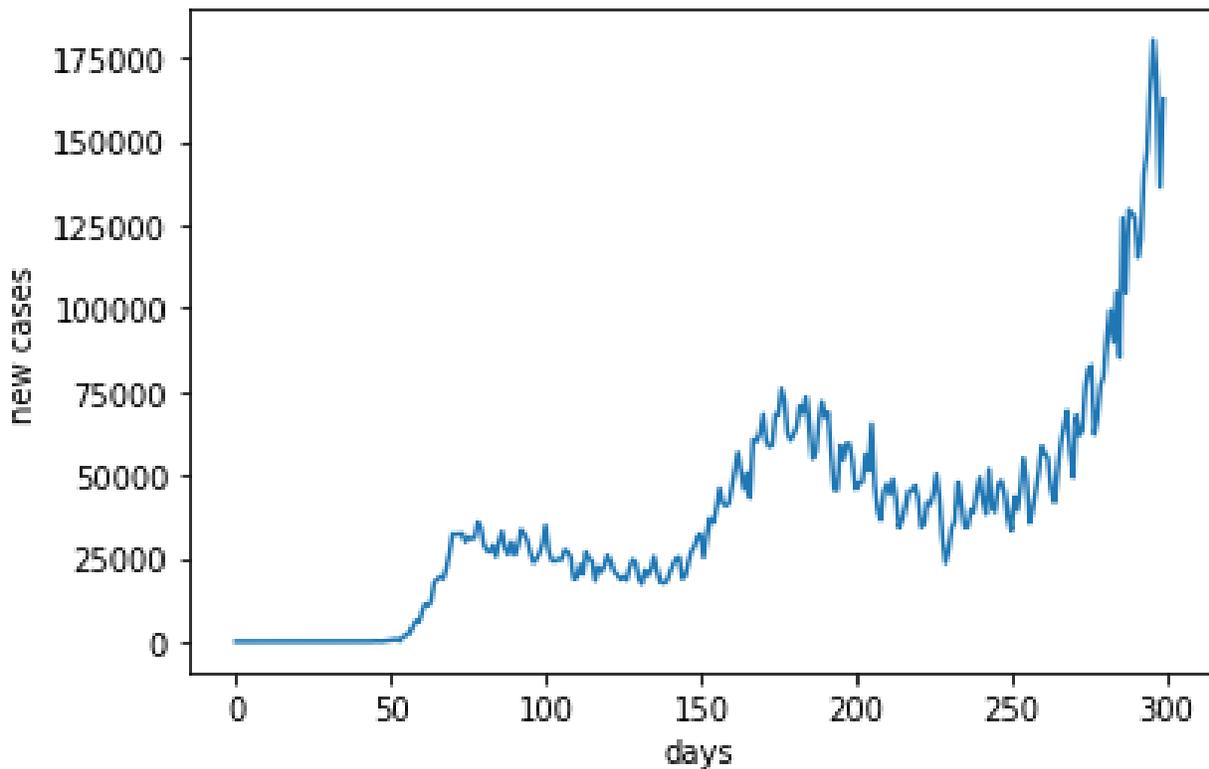


Figure 4.2: COVID-19 new cases for the first 300 days in the United States

Figure 4.3 shows the transform images from the data shown in Figure 4.2 for the United

States. To produce the image, initially, we make a transformation matrix of input size. The transformation matrix tries to reflect the changes(rise or fall) of COVID-19 new cases as time proceeds. Algorithm 2 shows the working procedure to calculate the distance matrix from tabular data. After that with the help of matplotlib colormap the converted 300×300 matrix is transform

Algorithm 2 Tabular to image transformation

```
1: procedure DISTANCE MATRIX( $d_{mn}$ )
2:    $d \leftarrow$  empty matrix of input data size
3:   for each  $i$  in  $m$  do
4:     for each  $j$  in  $n$  do
5:        $d_{ij} \leftarrow X(i) - X(j)$ 
```

into image show in Figure 4.3.

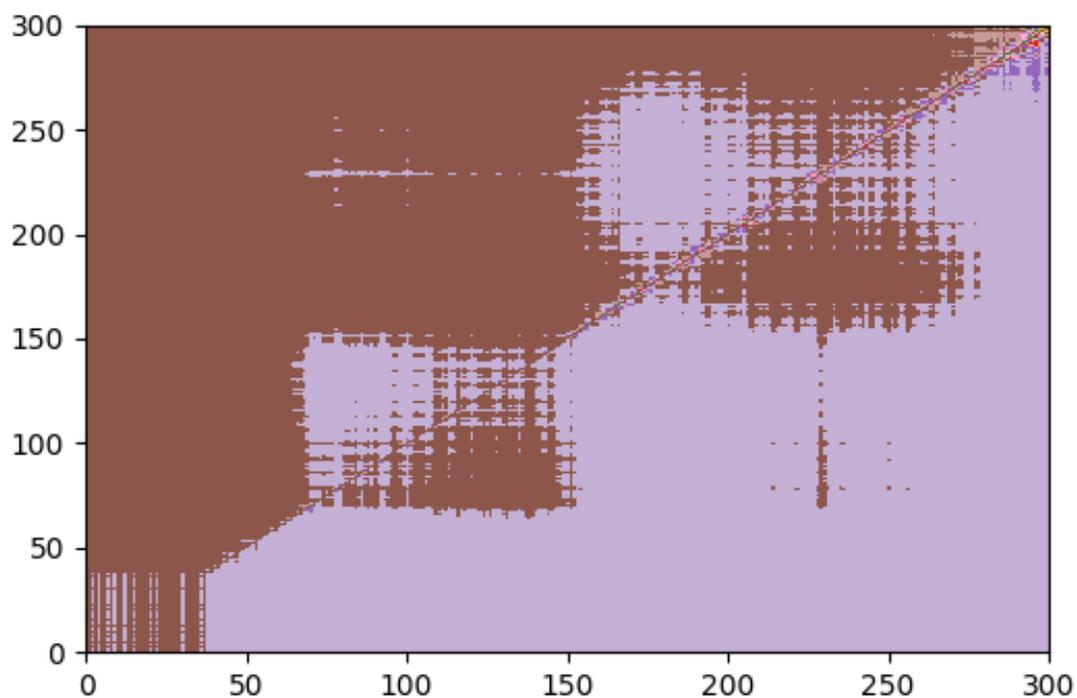


Figure 4.3: Transformed Image into 300×300 Pixel for the United States

4.2 Development of Convolutional Neural Network (CNN)

This section introduces an image classification problem, which is the task of assigning a specific category to an input image. An image is a numerical representation of a pixel value in a three-dimensional array (for RGB). Images in three-dimensional arrays of integers ranging from 0 to 255, with width \times height \times 3. The 3 represent the three color channels in the last dimension: red, green, and blue. For example, Figure 4.3 is an image of COVID-19 new cases from the United States with labels representing different cultural dimensions. Image classification is not a trivial task because of viewpoint variation, scale variation, deformation, occlusion, lighting conditions, and background clutter.

Figure 4.4 and 4.5 represents CNN architecture with and without residual block respectively. CNN with convolution only (Figure 4.4) start with a 2D convolution which is fairly a simple operation starting with a kernel (small weights matrix). This kernel (in our case 3×3) slides over the input data, performing element-wise multiplication with the portion of the input it is currently on, and then adding the results to form a single output pixel. When working with image data, there are several advantages to using convolutional layers, including a small number of convolutional layer parameters, sparsity connection, parameter sharing, and spatial invariance. After that we use a maxpooling layer after two 2D convolution layers to gradually reduce the height and width of the output tensors from each convolutional layer. The Convolution 2D layer transform a 3- layer image to a 16 channel feature map and the MaxPool2d layer halves the height and width. The pattern of convolution and pooling continues. We flatten the final process image and pass it into a neural network with the ReLU activation function at the end of the process in the last layer of CNN, giving the output of the image class.

We also use a different CNN architecture which can be powerful in capturing the relationship between the images and the labels more accurately called CNN with residual block shown in Figure 4.5. One of the fundamental changes to the previous CNN model is that here, in addition to the convolution and pooling layer, we add a residual block that adds the original input back to the output feature map obtained by passing the input through one or more convolution layer. And this

```

(network): Sequential(
  (0): Conv2d(3, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (1): ReLU()
  (2): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (3): ReLU()
  (4): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (5): Conv2d(64, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (6): ReLU()
  (7): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (8): ReLU()
  (9): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (10): Conv2d(128, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (11): ReLU()
  (12): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (13): ReLU()
  (14): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (15): Flatten(start_dim=1, end_dim=-1)
  (16): Linear(in_features=4096, out_features=1024, bias=True)
  (17): ReLU()
  (18): Linear(in_features=1024, out_features=512, bias=True)
  (19): ReLU()
  (20): Linear(in_features=512, out_features=10, bias=True)

```

Figure 4.4: Convolutional Neural Network Architecture without Residual Block.

CNN architecture is deeper than the previous one. Also, in the second CNN model, we add a batch normalization layer which normalizes the outputs of the previous layer.

4.3 Experimental Setup

The dataset was divided into 80%, 10%, and 10% for training, validation, and testing, respectively, in the experiment. The dataset contains 115 images from 115 countries, with 91, 12, and 12 images used for training, validation, and testing respectively. The data set is divided into training, validation, and testing avoiding data leakage and undersampling. A 5-fold cross-validation technique was used to obtain the results. Also, the 115 countries are divided into two classes based on high (> 50) and low culture (< 50) dimension defined by Hofstede. The proposed network is made up of 12 convolutional layers; as shown in the Figure, the learning rate is vary with batch size, and the maximum epoch number is 8 to 10, which was determined experimentally. Python and the PyTorch package were used to implement the CNN with and without residual blocks using the graphical processing unit (GPU).

```

ResNet9(
  (conv1): Sequential(
    (0): Conv2d(3, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU(inplace=True)
  )
  (conv2): Sequential(
    (0): Conv2d(64, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU(inplace=True)
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (res1): Sequential(
    (0): Sequential(
      (0): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (1): Sequential(
      (0): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU(inplace=True)
    )
  )
  (conv3): Sequential(
    (0): Conv2d(128, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU(inplace=True)
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (conv4): Sequential(
    (0): Conv2d(256, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU(inplace=True)
    (3): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (res2): Sequential(
    (0): Sequential(
      (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (1): Sequential(
      (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (1): BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (2): ReLU(inplace=True)
    )
  )
  (classifier): Sequential(
    (0): MaxPool2d(kernel_size=4, stride=4, padding=0, dilation=1, ceil_mode=False)
    (1): Flatten(start_dim=1, end_dim=-1)
    (2): Dropout(p=0.2, inplace=False)
    (3): Linear(in_features=512, out_features=10, bias=True)
  )
)

```

Figure 4.5: Convolutional Neural Network Architecture wit Residual Block.

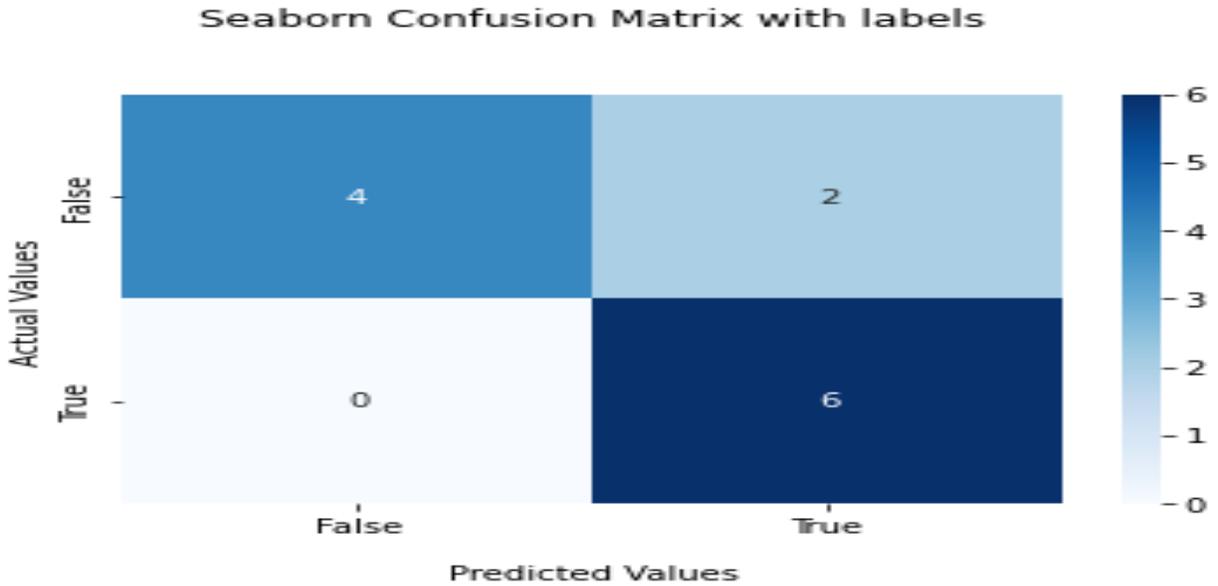


Figure 4.6: Confusion matrix of culture (power distance) prediction based on CNN architecture.

4.4 Performance Evaluation Metrics

4.4.1 Results Analysis for Power Distance

The Figure 4.6 depicts the confusion matrix of the best CNN architecture for culture classification during the testing phase. Two images from class 2 were misclassified by the CNN architecture among the 12 images from testing data. The proposed system can efficiently identify the class 2 groups of power distance. Figure 4.7 shows the performance evaluation of the CNN classifier in terms of accuracy with varying number of epochs. At the training image size is small and we didn't run it for large number of epoch as there is a chance for over-fitting. At epoch 1, the training accuracy decreases slightly but continues to increase further. The CNN architecture has a training accuracy of 0.83 for identifying the power distance. The training and validation loss decreases with the increasing number of epochs (Figure 4.8). After six epochs, the training and validation loss became almost constant at 0.6 and 0.51, respectively. The addition of the residual block, which adds the original input back to the output feature map obtained by passing the input through one or more convolutional layers, is one of the most significant changes to our CNN model this time. Figure 4.9 shows the performance evaluation of the CNN classifier with residual block in

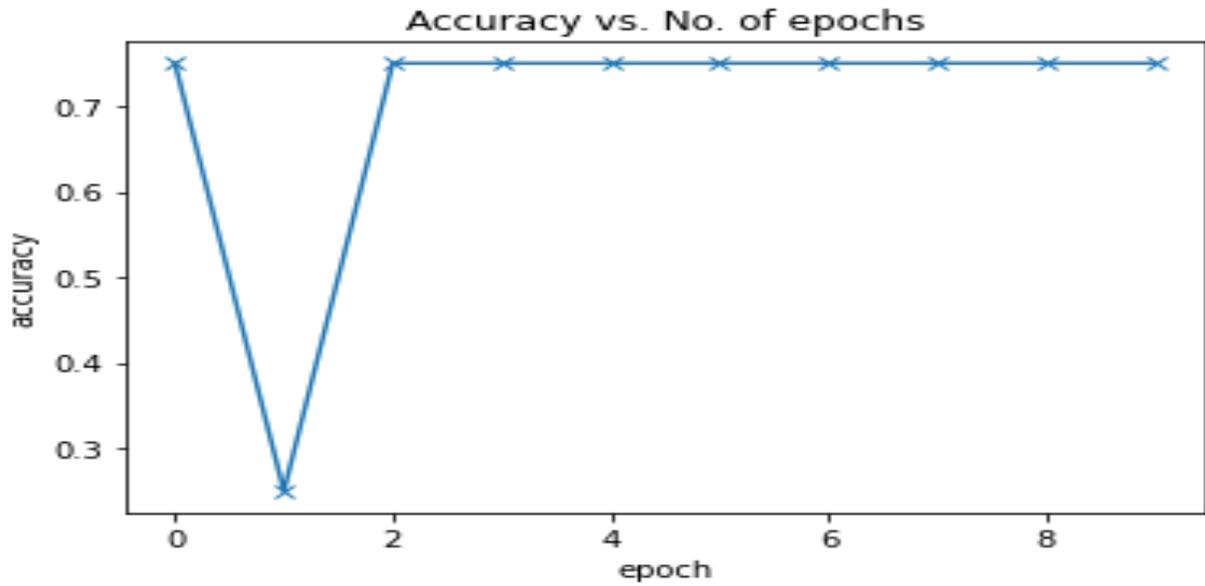


Figure 4.7: Accuracy of culture (power distance) prediction with varying epoch based on CNN architecture.

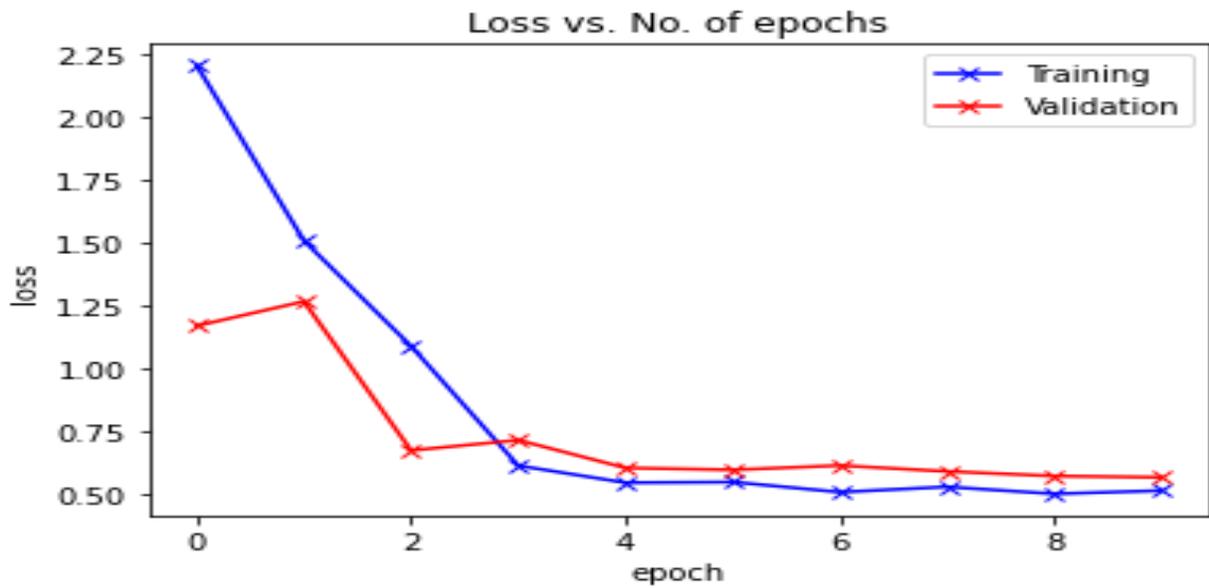


Figure 4.8: Loss of culture (power distance) prediction with varying epoch based on CNN architecture.

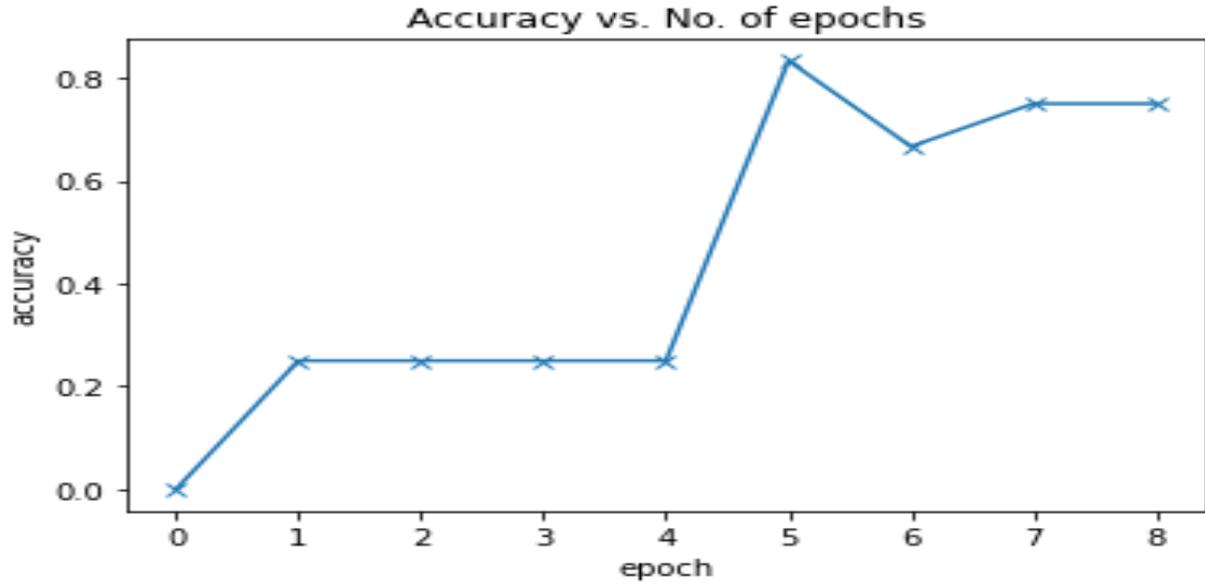


Figure 4.9: Accuracy of culture (power distance) prediction with varying epoch based on CNN with residual block.

terms of accuracy with varying number of epochs. At epoch 8, the CNN architecture with residual block has a training accuracy of 0.81 for identifying the power distance.

The training and validation loss decreases with the increasing number of epochs for CNN model with residual block (Figure 4.10). After five epochs, the training and validation loss became almost constant. Figure 4.11 shows how we change the learning rate with the varying number of batches on CNN architecture with residual block. The variation of the learning rate is completely based on how the model works (accuracy) with validation images. The overall accuracy, specificity, and sensitivity for the best CNN model to predict the power distance are 0.83, 0.67, and 1 respectively. The convolutional neural network can identify power distance from COVID-19, meaning the cultural dimension (power distance) has some impact on the courses of the epidemic.

4.4.2 Results Analysis for Individualism

During the testing phase, the confusion matrix of the best CNN architecture for culture (individualism) classification is depicted in Figure 4.12. Among the 12 images from testing data, the CNN architecture misclassified two images from class 2 and one from class 1. The proposed system can efficiently identify individualism's category two groups. Figure 4.13 depicts the CNN classifier's

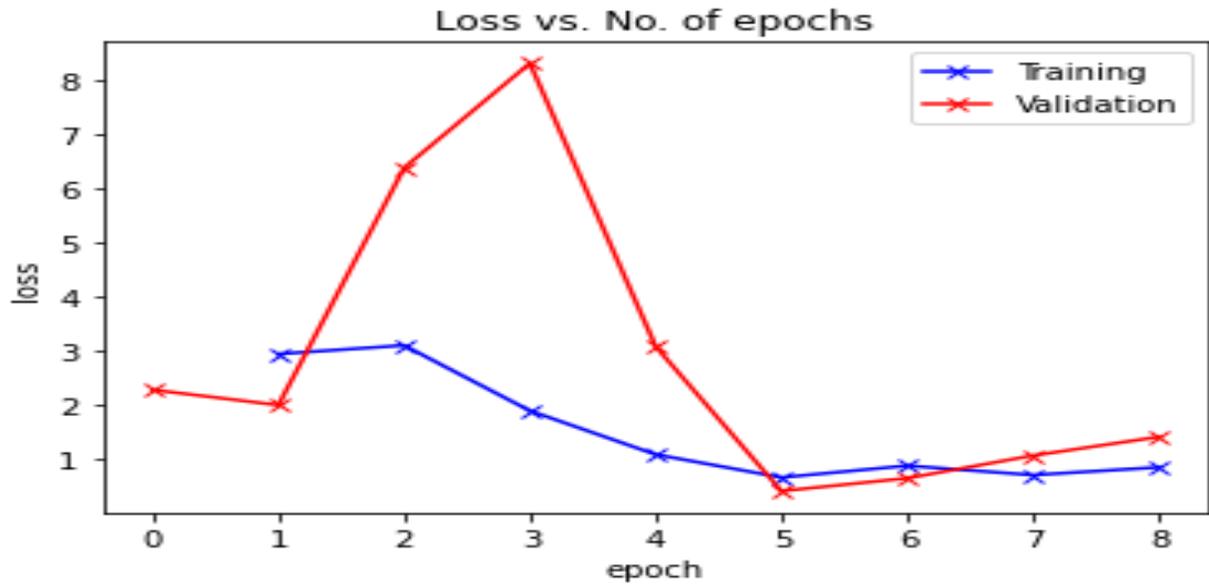


Figure 4.10: Loss of culture (power distance) prediction with varying epoch based on CNN with residual block.

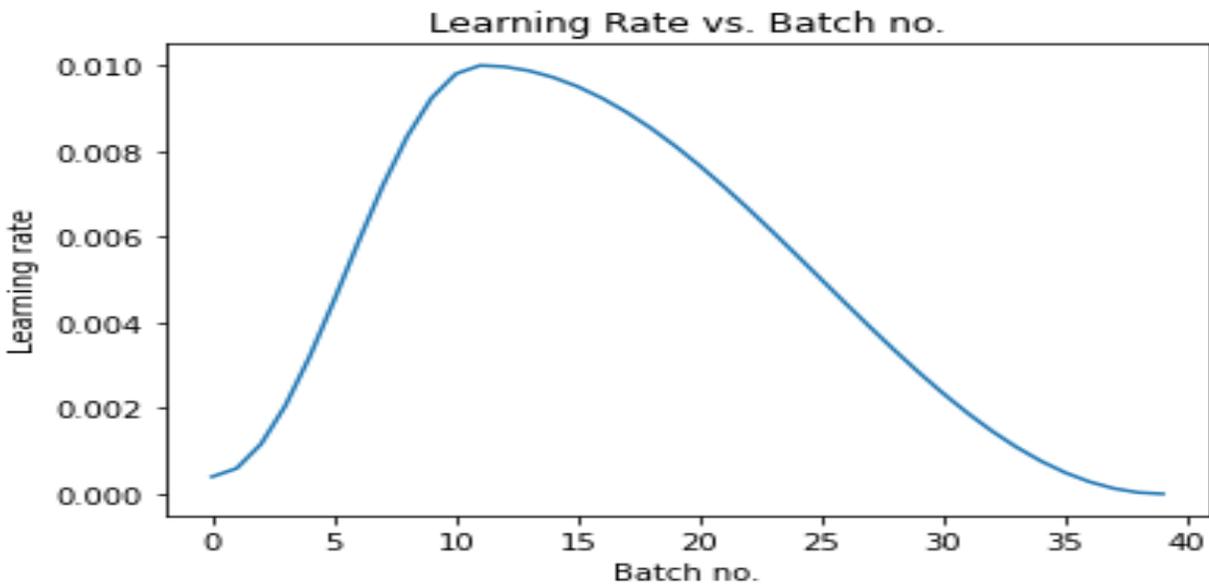


Figure 4.11: Learning rate of culture (power distance) prediction with varying batch based on CNN with residual block.

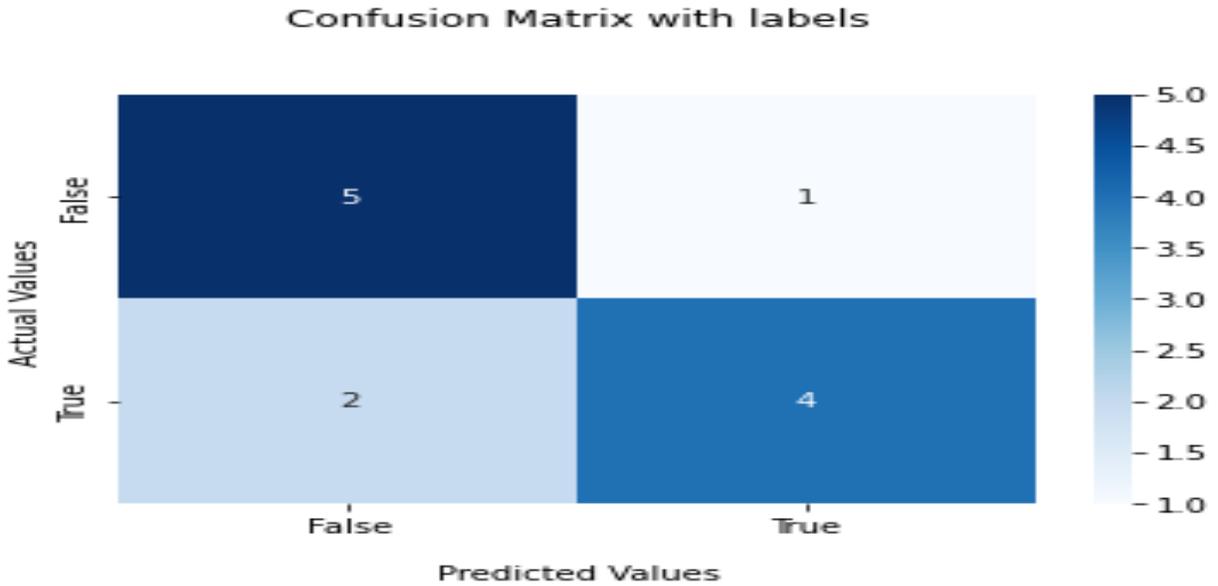


Figure 4.12: Confusion matrix of culture (individualism) prediction based on CNN architecture.

performance evaluation in terms of accuracy with varying epochs. The training image size is small, and we did not run it for a large number of epochs to avoid over-fitting. The training accuracy was initially low but has since increased. For identifying individualism, the CNN architecture has a training accuracy of 0.72. The training and validation loss decreases as the number of epochs increases (Figure 4.14). The training and validation losses were nearly constant after three epochs, at 0.5 and 0.59, respectively. The performance evaluation of the CNN classifier with the residual block in terms of accuracy with varying epochs is shown in Figure 4.15. The CNN architecture with residual block has a training accuracy of 0.79 for identifying individualism at epoch 8. For CNN models with residual blocks, the training and validation loss decreases as the number of epochs increases (Figure 4.16). The training and validation losses became nearly constant after seven epochs. Figure 4.17 shows how we change the learning rate on CNN architecture with the residual block by varying the number of batches. The learning rate varies entirely depending on how the model works (accuracy) with validation images. A batch size of 10 to 12 achieves the highest learning rate. The best CNN model for predicting individualism has an overall accuracy, specificity, and sensitivity of 0.75, 0.67, and 0.83, respectively. From the analysis, it is evident that CNN is not too good or bad at identifying individualism from the COVID-19 epidemic. One of the

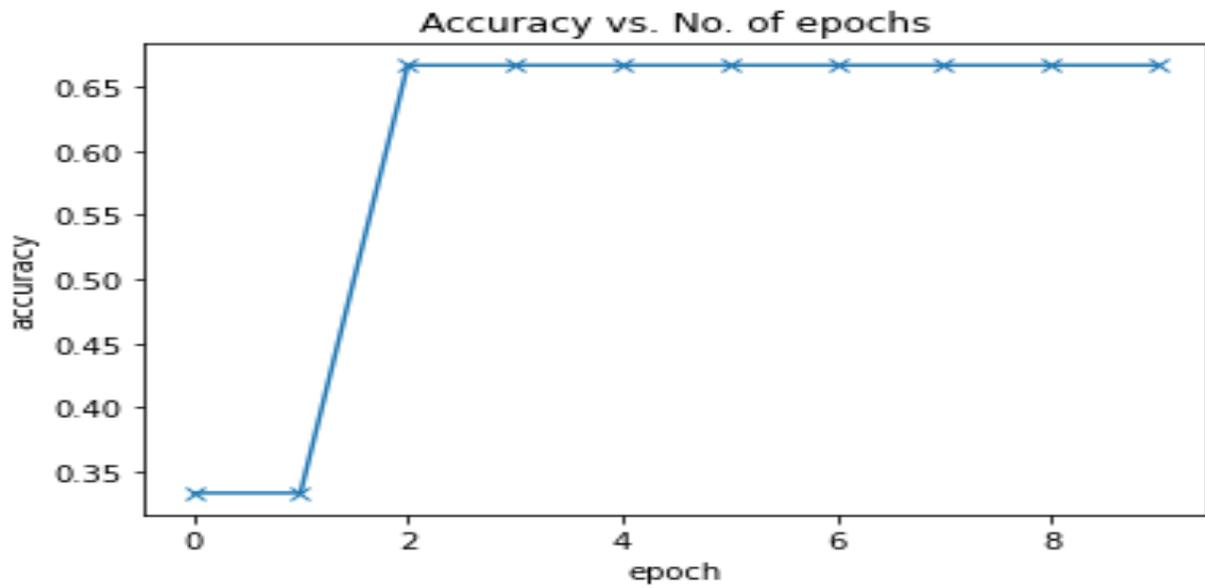


Figure 4.13: Accuracy of culture (individualism) prediction with varying epoch based on CNN architecture.

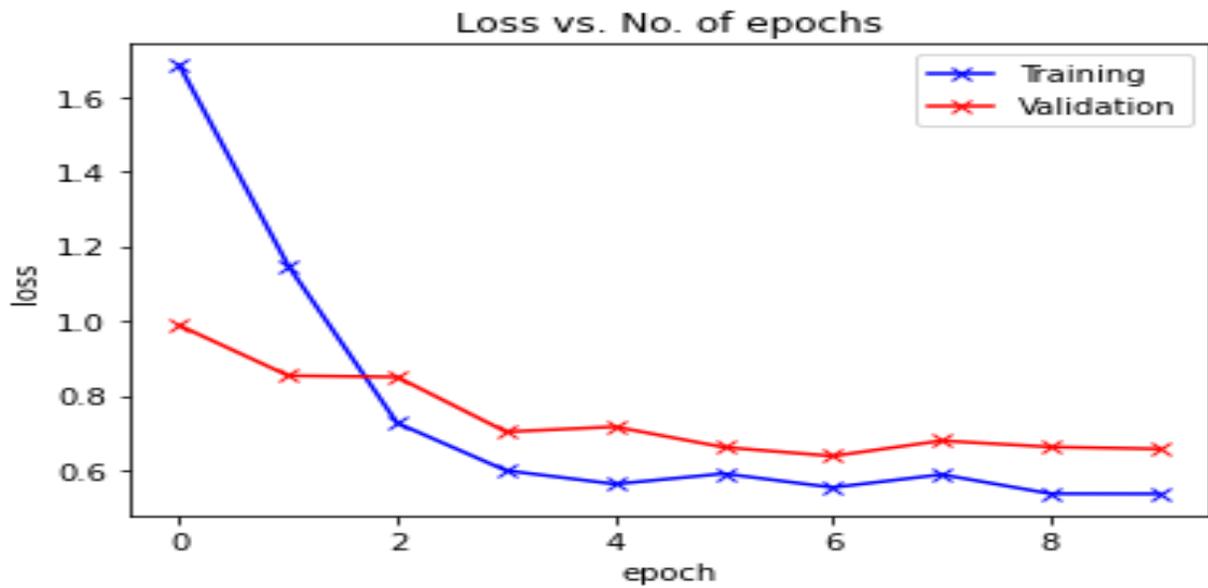


Figure 4.14: Loss of culture (individualism) prediction with varying epoch based on CNN architecture.

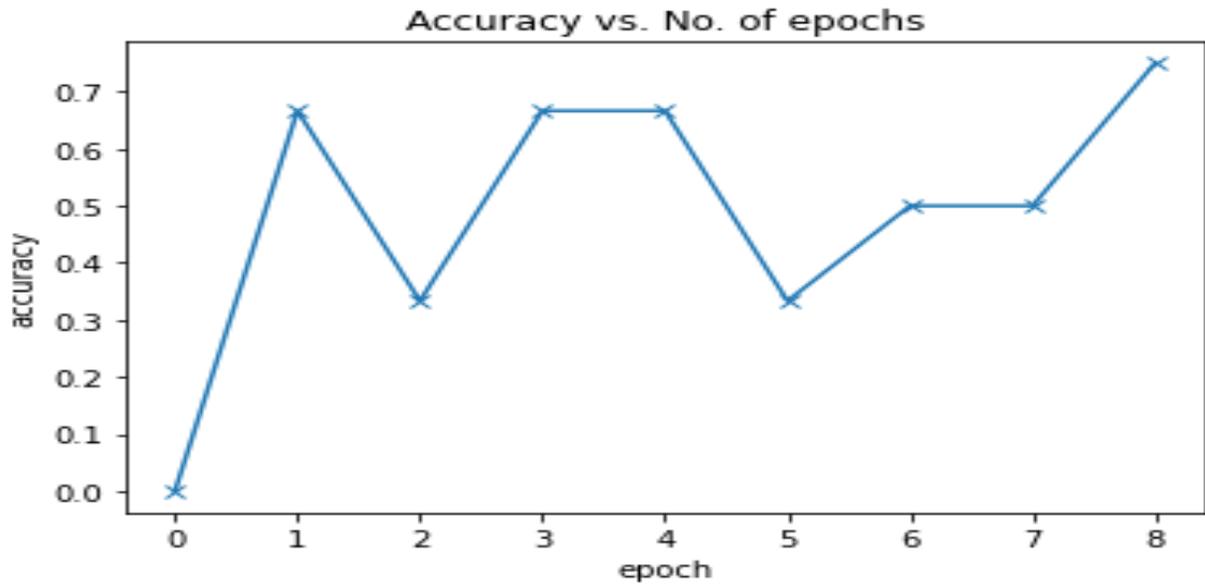


Figure 4.15: Accuracy of culture (individualism) prediction with varying epoch based on CNN with residual block.

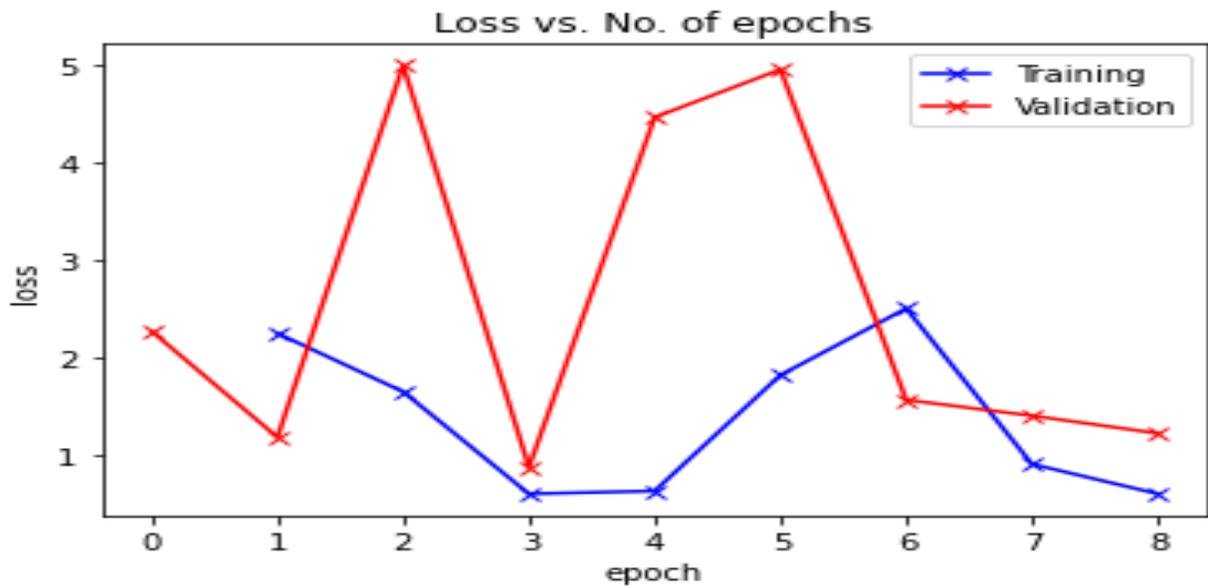


Figure 4.16: Loss of culture (individualism) prediction with varying epoch based on CNN with residual block.

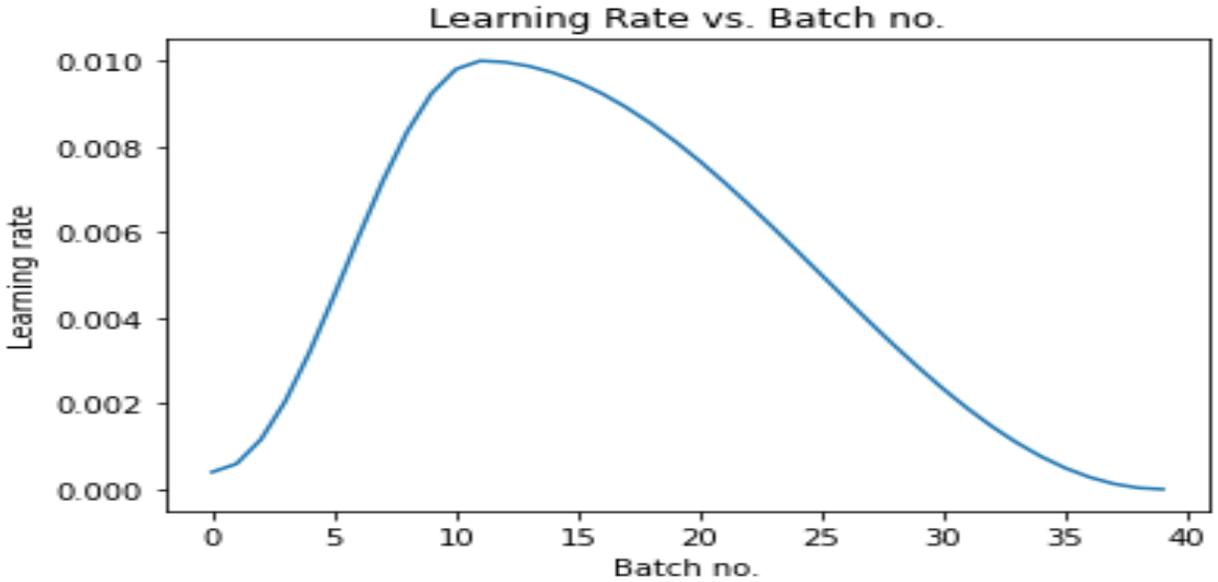


Figure 4.17: Learning rate of culture (individualism) prediction with varying batch based on CNN with residual block.

reasons may be individualism has minimal impact on the courses of the epidemic.

4.4.3 Results Analysis for Masculinity

During the testing phase, the confusion matrix of the best CNN architecture for culture (masculinity) classification is depicted in the Figure 4.18. Among the 12 images from testing data, the CNN architecture misclassified five images from class 1 and four from class 2. The proposed system is ineffective for identifying masculinity classes 1 and 2. The performance evaluation of the CNN classifier in terms of accuracy with varying epochs is shown in Figure 4.19. CNN architecture has not been shown to be particularly effective at identifying culture class masculinity. The training and validation loss do not show any specific decreasing pattern as the number of epochs increases (Figure 4.20). There could be two reasons for this: masculinity has little effect on explaining COVID-19 courses of the epidemic, or CNN architecture is unsuitable for identifying masculinity. To test the effectiveness of CNN on masculinity, we change the CNN architecture with a residual block. The performance evaluation of the CNN classifier with the residual block in terms of accuracy with varying epochs is shown in Figure 4.21. For identifying masculinity, the CNN architecture with residual block has a training accuracy of 0.5. For CNN models with residual blocks, the training

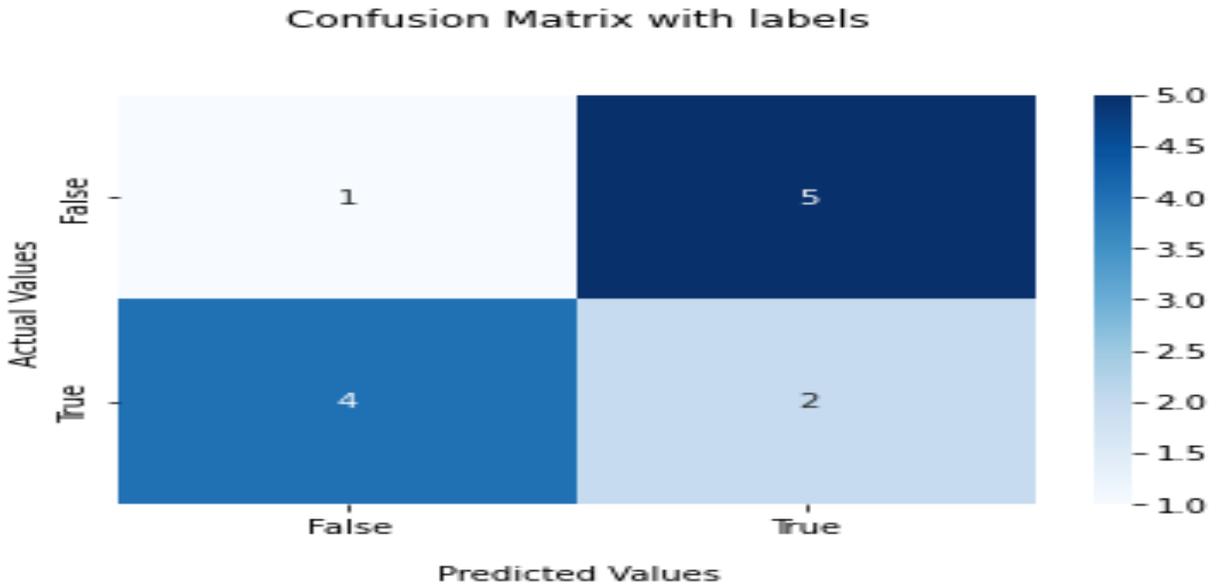


Figure 4.18: Confusion matrix of culture (masculinity) prediction based on CNN architecture.

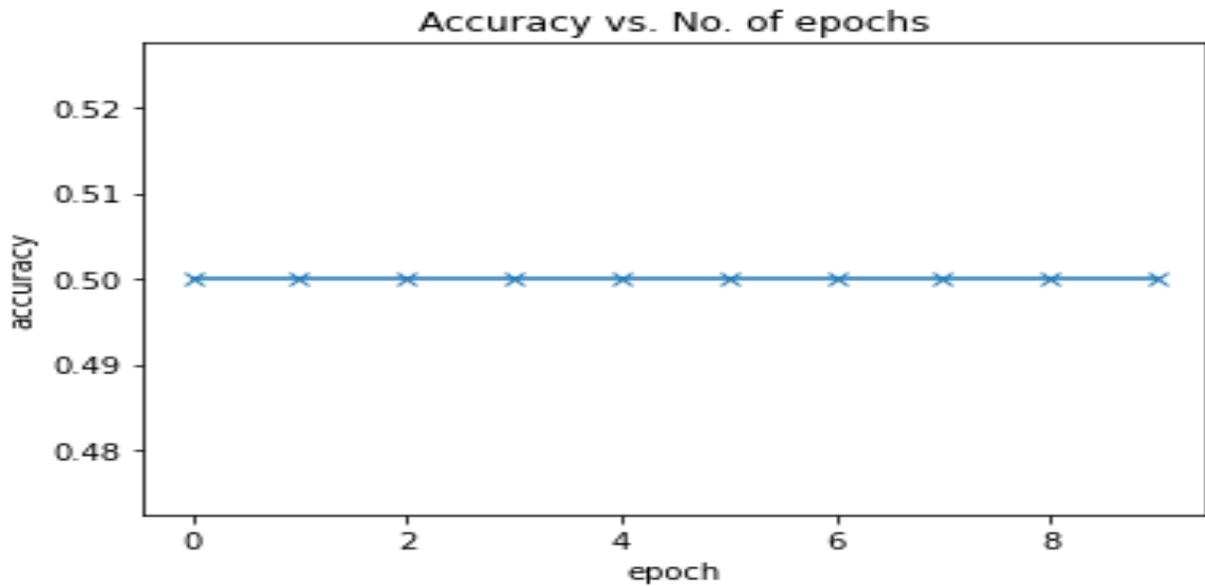


Figure 4.19: Accuracy of culture (masculinity) prediction with varying epoch based on CNN architecture.

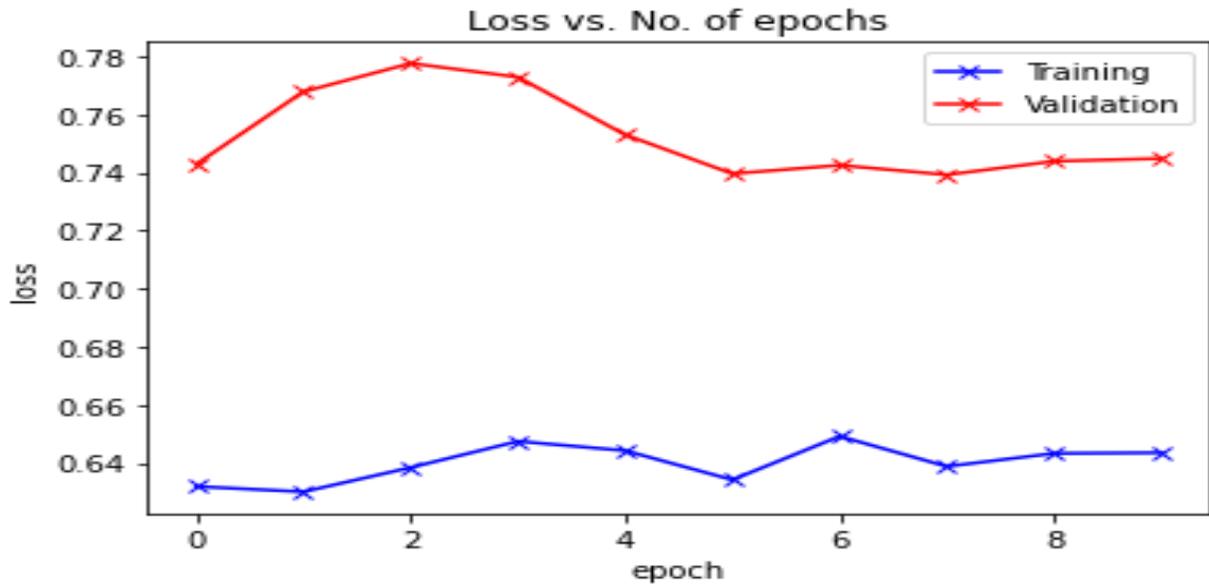


Figure 4.20: Loss of culture (masculinity) prediction with varying epoch based on CNN architecture.

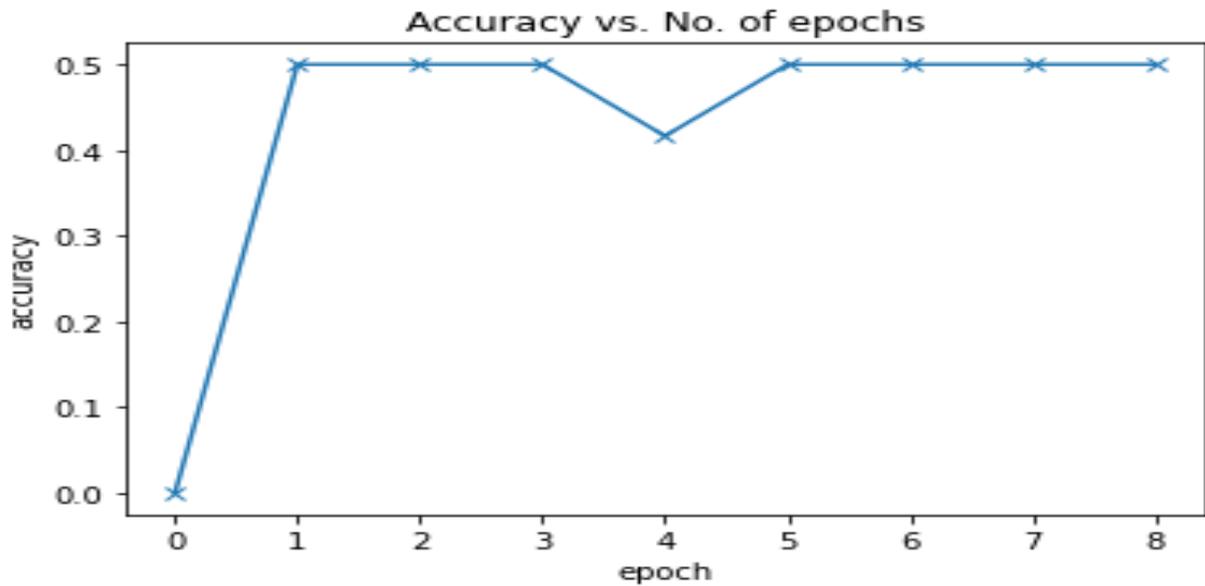


Figure 4.21: Accuracy of culture (masculinity) prediction with varying epoch based on CNN with residual block.

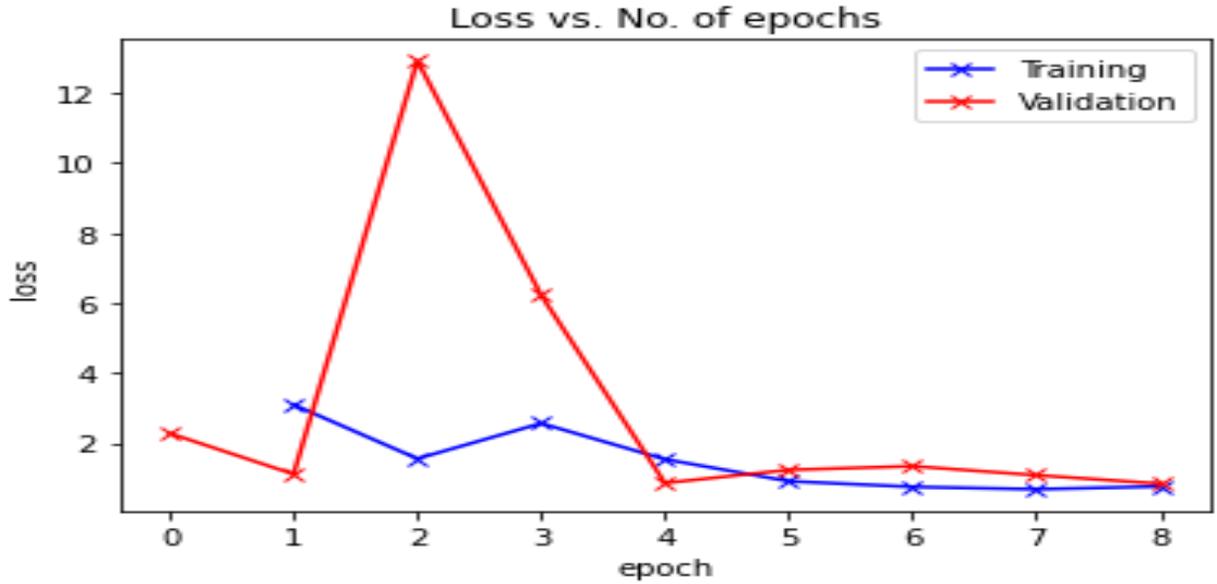


Figure 4.22: Loss of culture (masculinity) prediction with varying epoch based on CNN with residual block.

and validation loss decreases as the number of epochs increases (Figure 4.22). The training and validation losses were nearly constant after four epochs. The architecture demonstrated superior performance when compared to the previous CNN model. Figure 4.23 shows how we change the learning rate on CNN architecture with the residual block by varying the number of batches. The learning rate varies entirely depending on how the model works (accuracy) with validation images—a batch size of 12 results in the highest learning rate. Overall accuracy, specificity, and sensitivity of the best CNN model for predicting masculinity are 0.55, 0.33, and 0.16, respectively. We can conclude that masculinity is not a good feature to identify from COVID-19, implying that it may have no influence on the epidemic’s course.

4.4.4 Results Analysis for Uncertainty Avoidance

During the testing phase, the confusion matrix of the best CNN architecture for culture classification is depicted in the Figure 4.24. Among the 12 images from testing data, the CNN architecture misclassified one image from class 1 and one from class 2. The proposed system is efficient at identifying the uncertainty avoidance class 1 and 2 groups. The performance evaluation of the CNN classifier in terms of accuracy with varying epochs is shown in Figure 4.25. This CNN

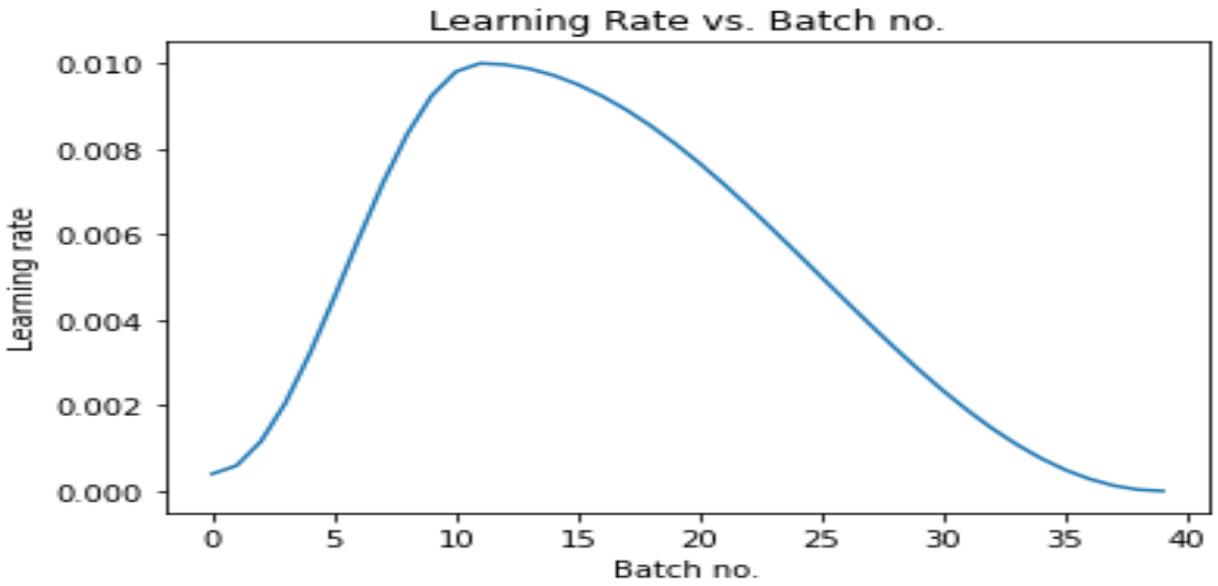


Figure 4.23: Learning rate of culture (masculinity) prediction with varying batch based on CNN with residual block.

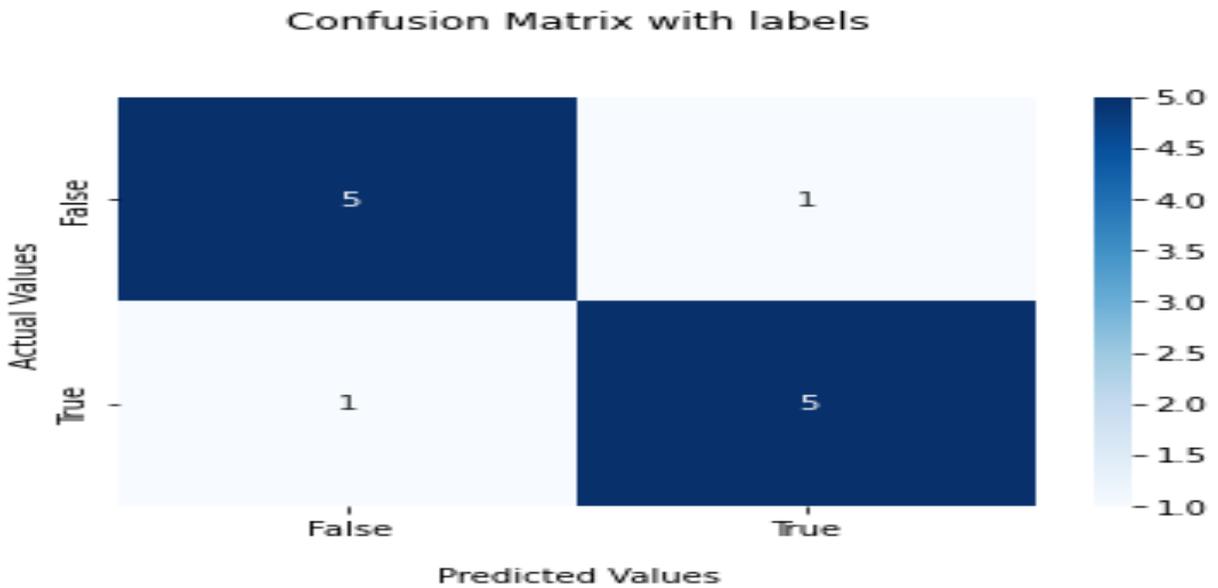


Figure 4.24: Confusion matrix of culture (uncertainty avoidance) prediction based on CNN architecture.

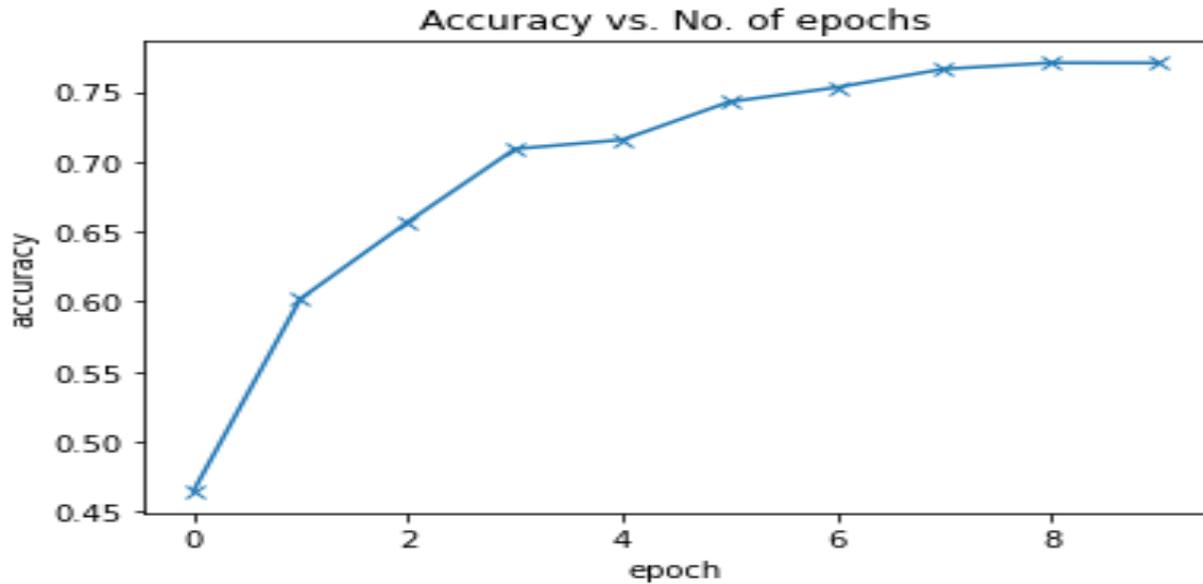


Figure 4.25: Accuracy of culture (uncertainty avoidance) prediction with varying epoch based on CNN architecture.

architecture is highly effective for classifying uncertainty avoidance and may significantly impact on epidemics. The training and validation loss decreases as the number of epochs increases (Figure 4.26). The training and validation losses became nearly constant after six epochs, at 0.40 and 0.91 respectively. One of the most significant changes to our CNN model this time is the addition of the residual block, which adds the original input back to the output feature map obtained by passing the input through one or more convolutional layers. The performance evaluation of the CNN classifier with the residual block in terms of accuracy with the varying number of epochs is shown in Figure 4.27. The CNN architecture with residual block has a training accuracy of 0.9 for identifying uncertainty avoidance at epoch 8. For CNN models with residual blocks, the training and validation loss decreases as the number of epochs increases (Figure 4.28). The training and validation losses were nearly constant after six epochs. Figure 4.29 shows how we change the learning rate on CNN architecture with residual blocks with varying batches. The learning rate varies entirely depending on how the model works (accuracy) with validation images. The best CNN model's overall accuracy, specificity, and sensitivity for predicting uncertainty avoidance are 0.9, 0.83, and 0.83, respectively. That is, uncertainty avoidance has a significant influence in explaining epidemic courses, which

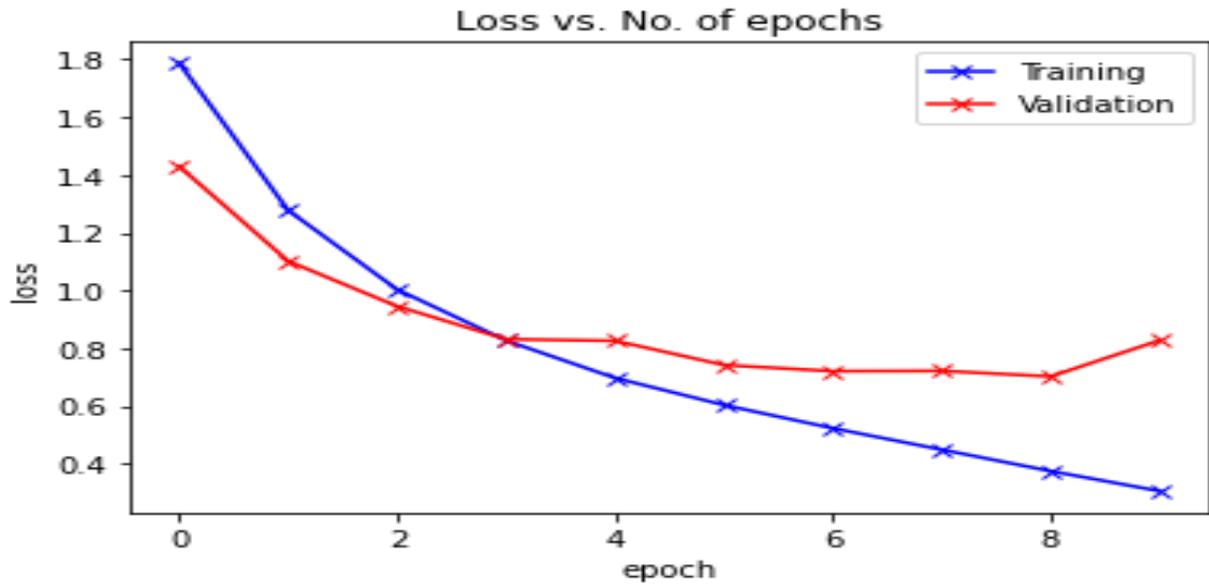


Figure 4.26: Loss of culture (uncertainty avoidance) prediction with varying epoch based on CNN architecture.

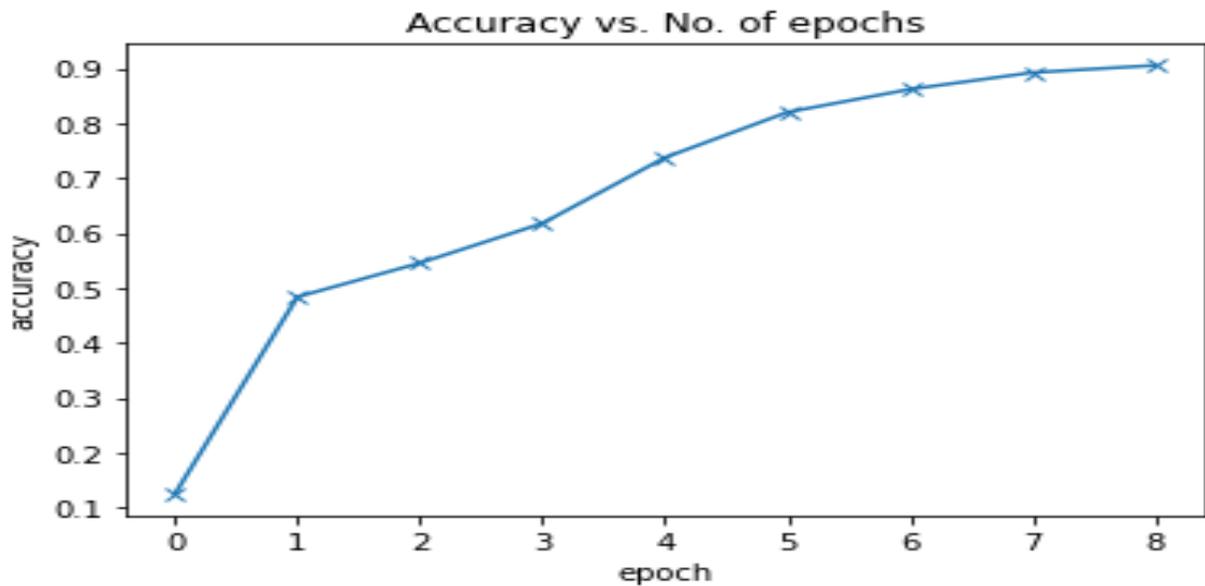


Figure 4.27: Accuracy of culture (uncertainty avoidance) prediction with varying epoch based on CNN with residual block.

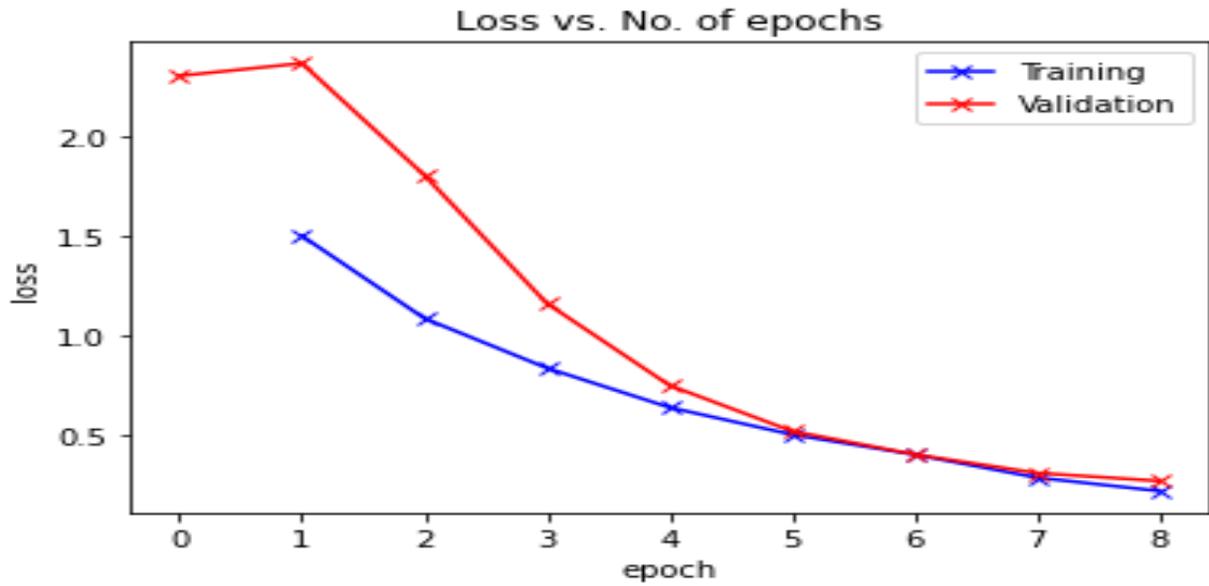


Figure 4.28: Loss of culture (uncertainty avoidance) prediction with varying epoch based on CNN with residual block.

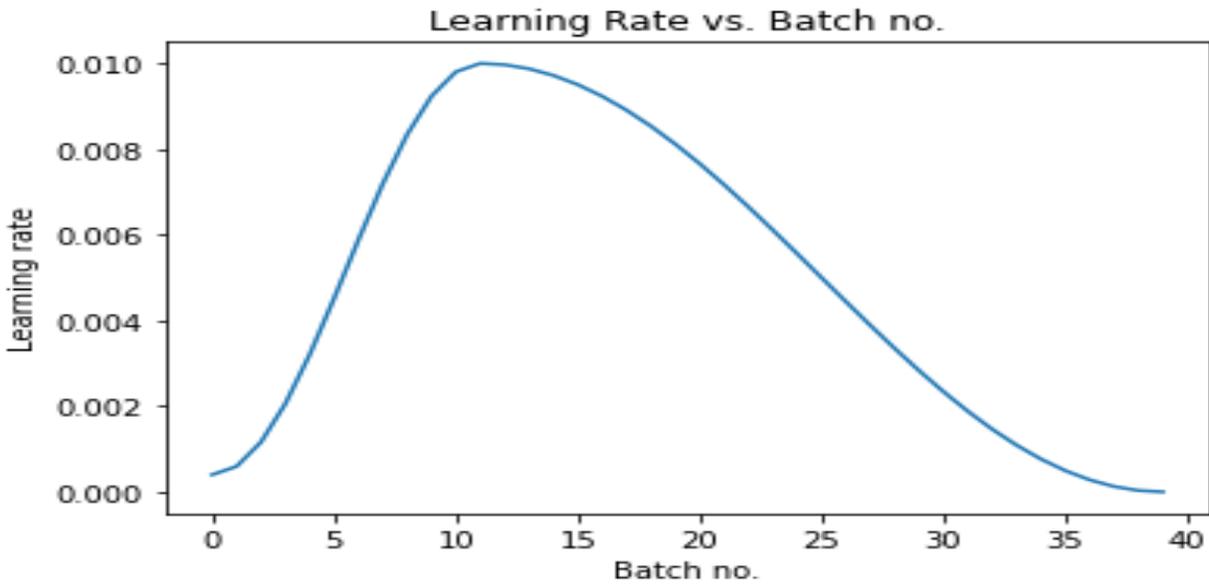


Figure 4.29: Learning rate of culture (uncertainty avoidance) prediction with varying batch based on CNN with residual block.

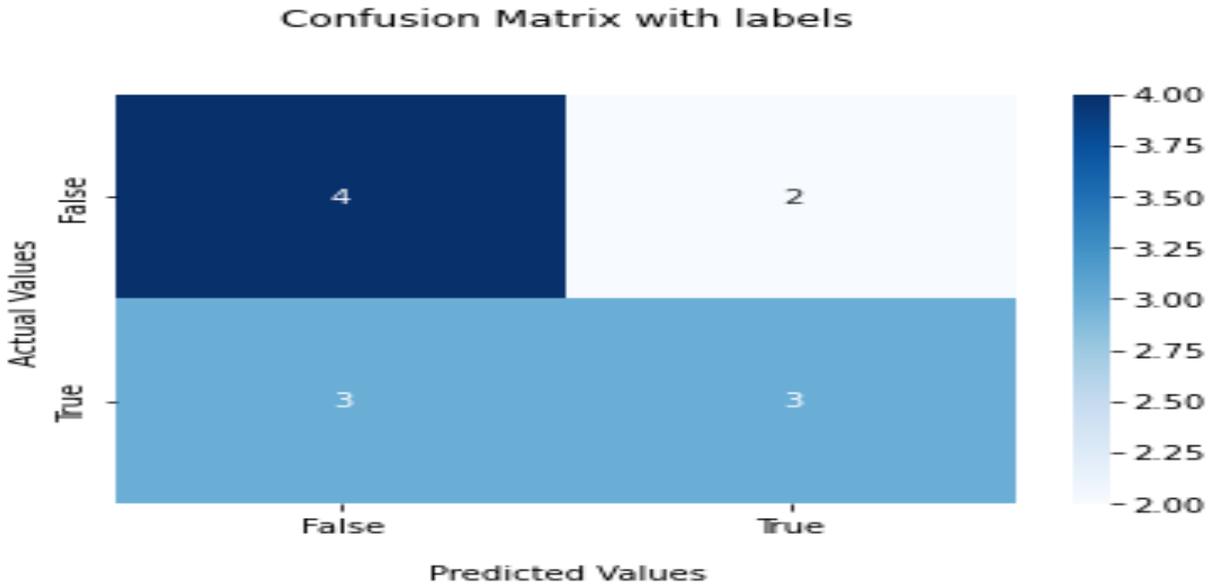


Figure 4.30: Confusion matrix of culture (long term orientation) prediction based on CNN architecture.

is understandable given that the uncertainty avoidance dimension expresses the degree to which members of a society are uncomfortable with uncertainty and ambiguity.

4.4.5 Results Analysis for Long Term Orientation

During the testing phase, the confusion matrix of the best CNN architecture for culture classification is depicted in Figure 4.30. Among the 12 images from testing data, the CNN architecture misclassified two images from class 1 and three from class 2. The proposed system is incapable of identifying class 2 long-term orientation groups efficiently. Figure 4.31 depicts the CNN classifier’s performance in terms of accuracy with varying number of epochs. The training image size is small, and we did not run it for a large number of epochs to avoid over-fitting. The training accuracy decreases slightly at epoch 4, but continues to rise. For identifying long-term orientation, the CNN architecture has a training accuracy of 0.6. The training and validation loss decreases as the number of epochs increases (Figure 4.32). The training and validation losses became nearly constant after four epochs, at 0.6 and 0.71 respectively. The performance evaluation of the CNN classifier with residual block in terms of accuracy with varying number of epochs is shown in Figure 4.33. The CNN architecture with residual block has a training accuracy of 0.59 for

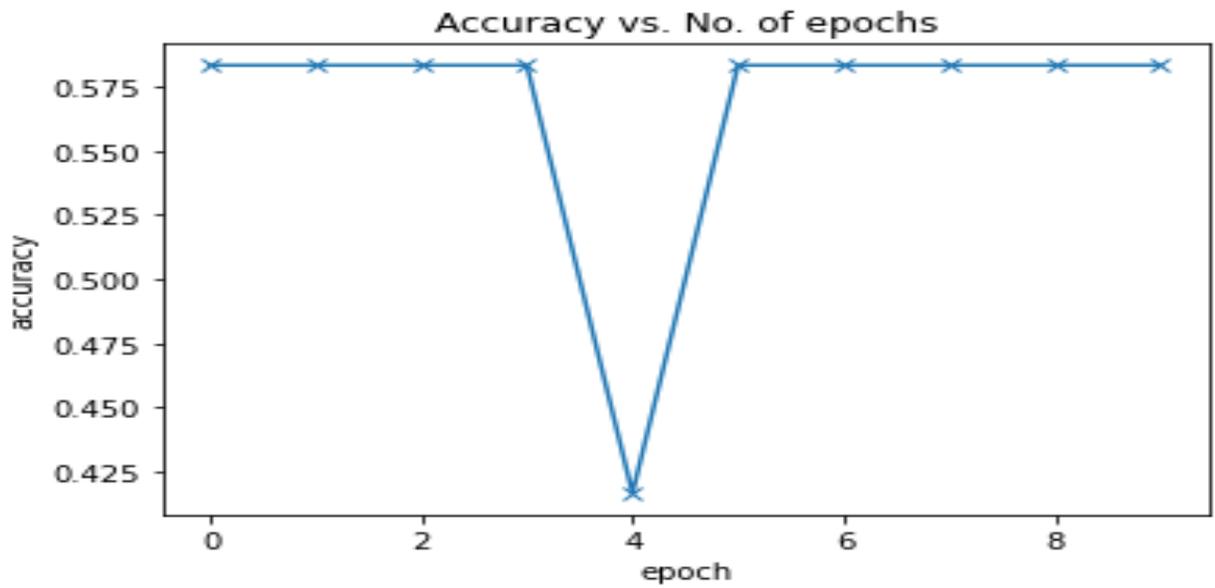


Figure 4.31: Accuracy of culture (long term orientation) prediction with varying epoch based on CNN architecture.

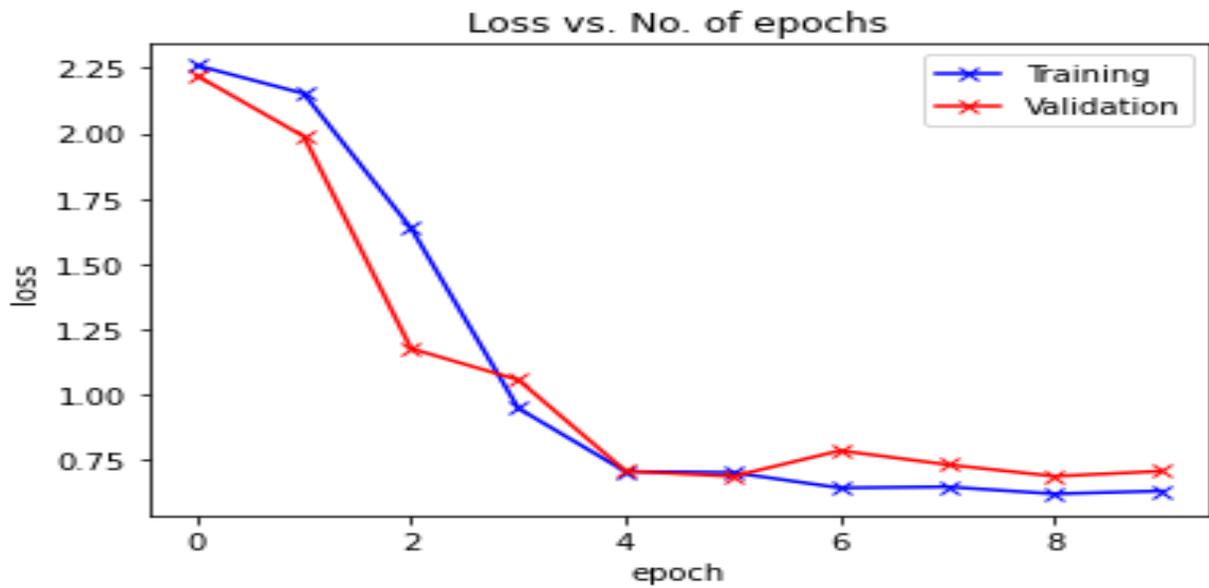


Figure 4.32: Loss of culture (long term orientation) prediction with varying epoch based on CNN architecture.

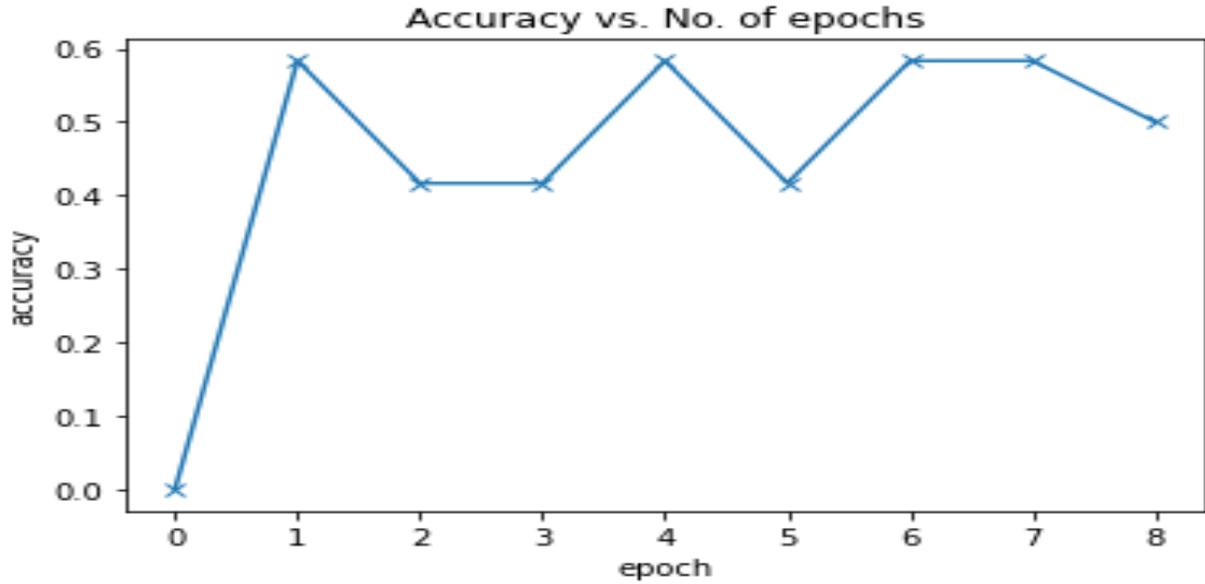


Figure 4.33: Accuracy of culture (long term orientation) prediction with varying epoch based on CNN with residual block.

identifying long term orientation at epoch 7. For CNN models with residual blocks, the training and validation loss decreases as the number of epochs increases (Figure 4.34). Figure 4.35 depicts how we change the learning rate with varying batch sizes on a CNN architecture with residual block. The learning rate varies entirely depending on how the model works (accuracy) with validation images. The best CNN model's overall accuracy, specificity, and sensitivity for predicting long-term orientation are 0.61, 0.5, and 0.67, respectively. This suggests that long-term thinking has an impact on the epidemic course.

4.4.6 Results Analysis for Indulgence

During the testing phase, the confusion matrix of the best CNN architecture for indulgence culture classification is depicted in the Figure 4.36. Among the 12 images from testing data, the CNN architecture misclassified three images from classes 1 and 2. The performance evaluation of the CNN classifier in terms of accuracy with varying epochs is shown in Figure 4.37. The training and validation loss decreases as the number of epochs increases (Figure 4.38). The training and validation losses were nearly constant after four epochs, at 0.5 and 0.54, respectively. The performance evaluation of the CNN classifier with a residual block in terms of accuracy with varying

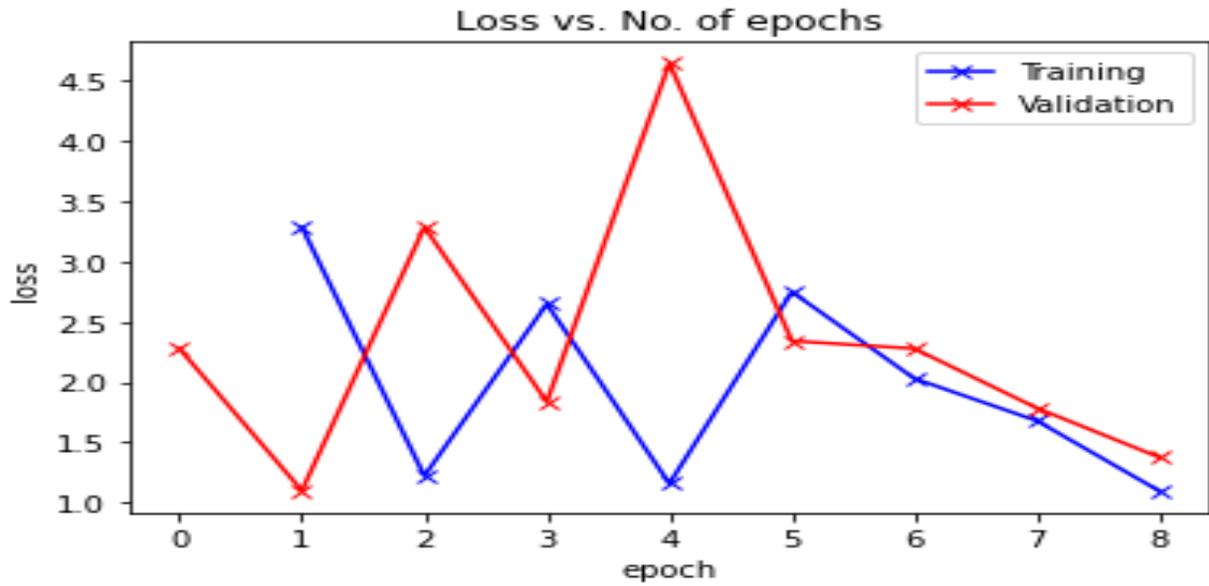


Figure 4.34: Loss of culture (long term orientation) prediction with varying epoch based on CNN with residual block.

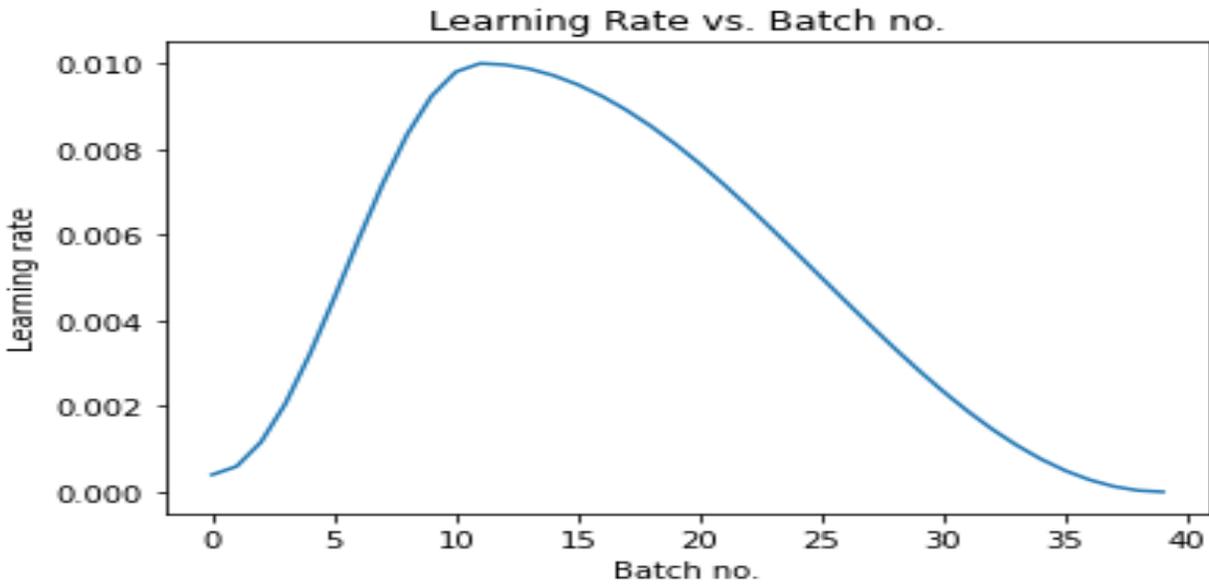


Figure 4.35: Learning rate of culture (long term orientation) prediction with varying batch based on CNN with residual block.

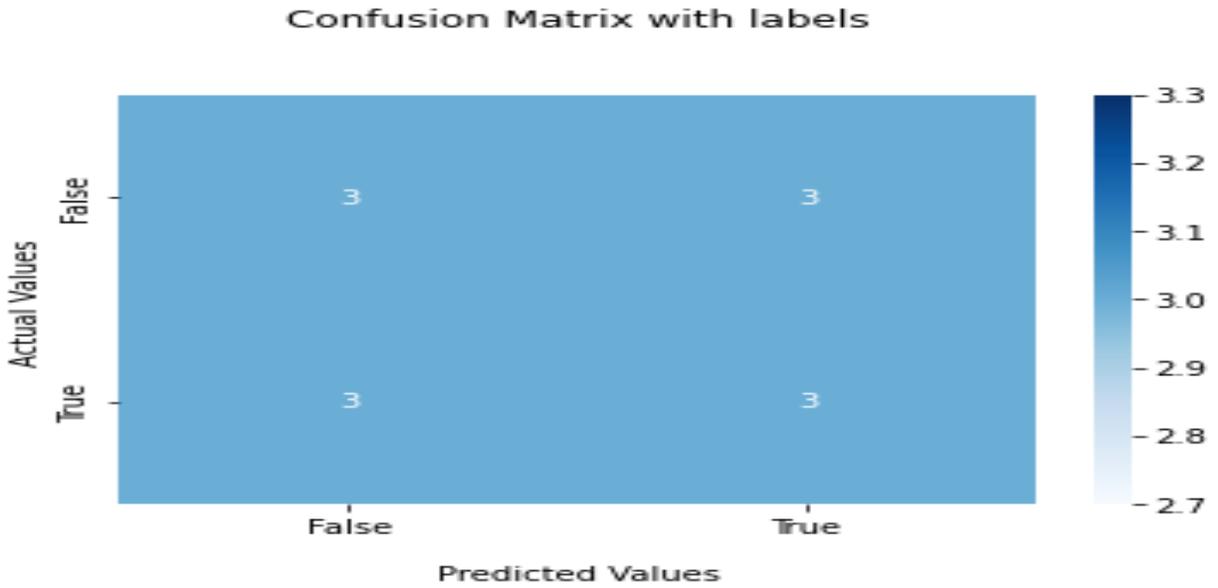


Figure 4.36: Confusion matrix of culture (indulgence) prediction based on CNN architecture.

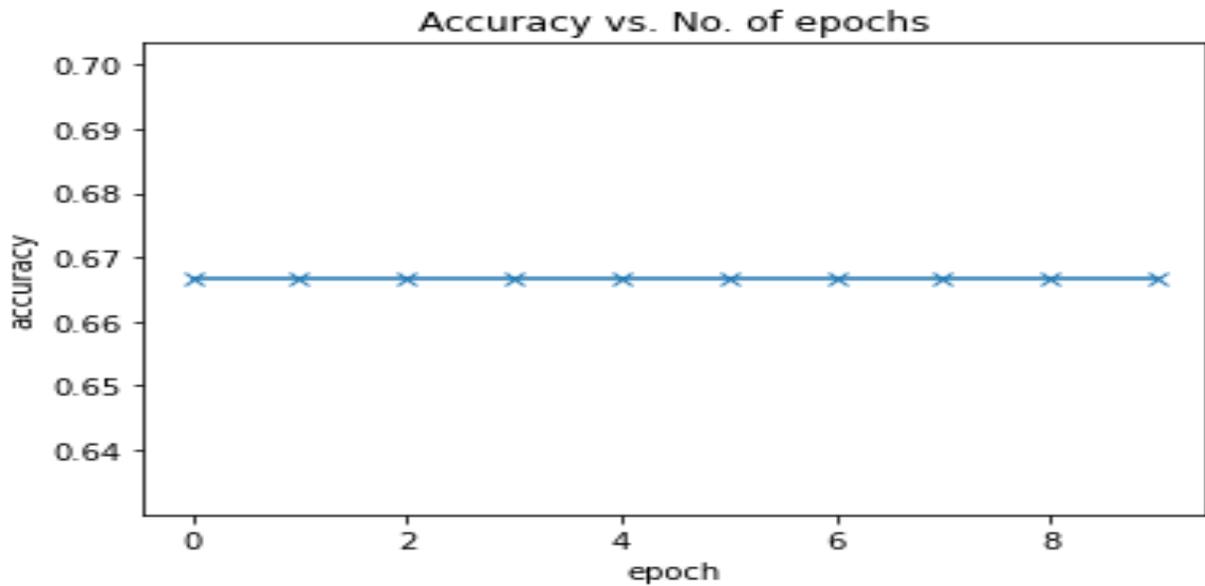


Figure 4.37: Accuracy of culture (indulgence) prediction with varying epoch based on CNN architecture.

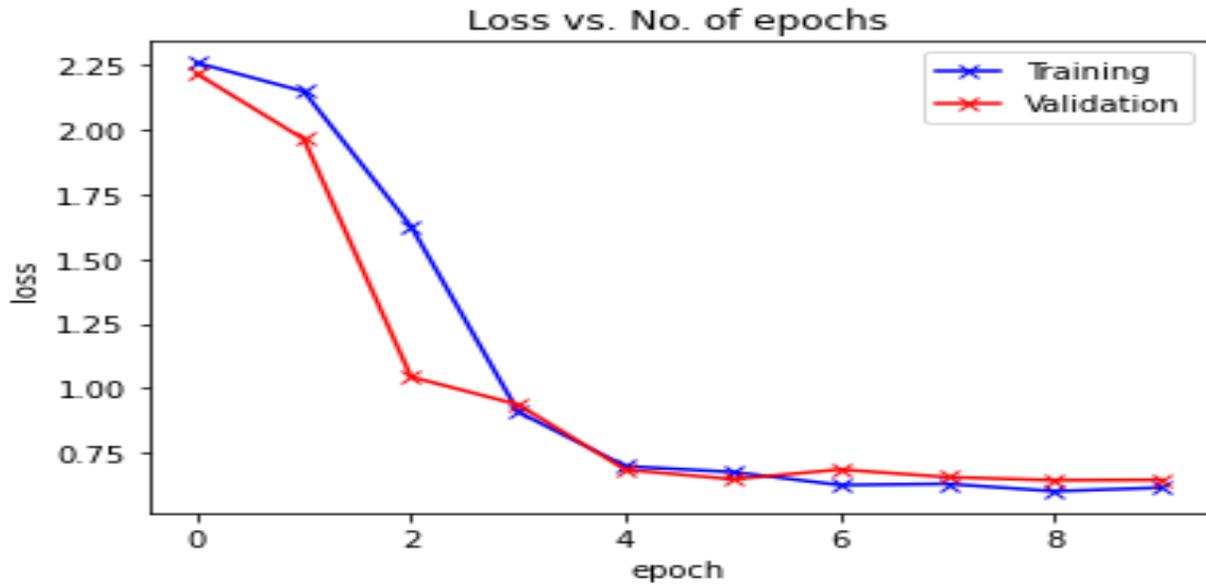


Figure 4.38: Loss of culture (indulgence) prediction with varying epoch based on CNN architecture.

epochs is shown in Figure 4.39. For CNN models with residual blocks, the training and validation loss decreases as the number of epochs increases (Figure 4.40). Figure 4.41 shows how we change the learning rate on CNN architecture with the residual block by varying the number of batches. The learning rate varies entirely depending on how the model works (accuracy) with validation images. The best CNN model for predicting indulgence has an overall accuracy, specificity, and sensitivity of 0.5, 0.5, and 0.5, respectively. It implies that the cultural dimension of indulgence has a lower impact on explaining the epidemic course, and CNN architecture is ineffective at predicting indulgence from COVID-19.

4.5 Discussion and Conclusion

After analyzing the CNN model results, it is clear that not all cultures have the same effect on COVID-19 epidemic courses. Specific cultural dimensions, such as uncertainty avoidance, power distance, and individualism, significantly affect the COVID-19 pandemic, while others are less effective. In Section III with 1000 MCMC iterations, we find that culture dimensions are not predictive of the COVID-19 epidemic course, which contradicts the CNN result. As a result, it is necessary to allow the MCMC mixing for longer iterations to achieve better convergence and

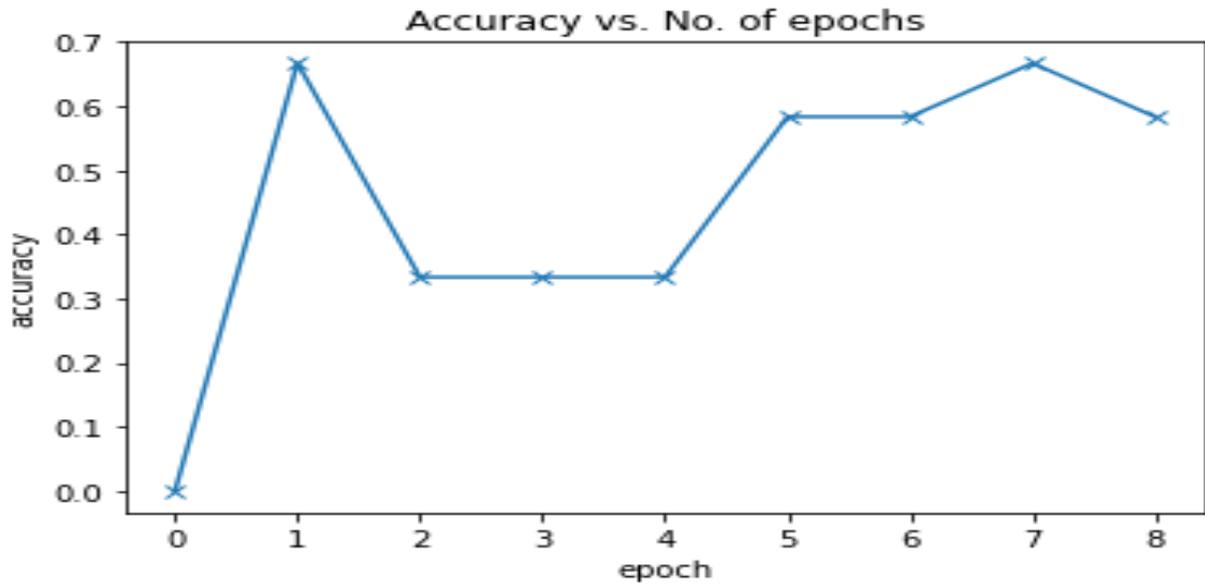


Figure 4.39: Accuracy of culture (indulgence) prediction with varying epoch based on CNN with residual block.

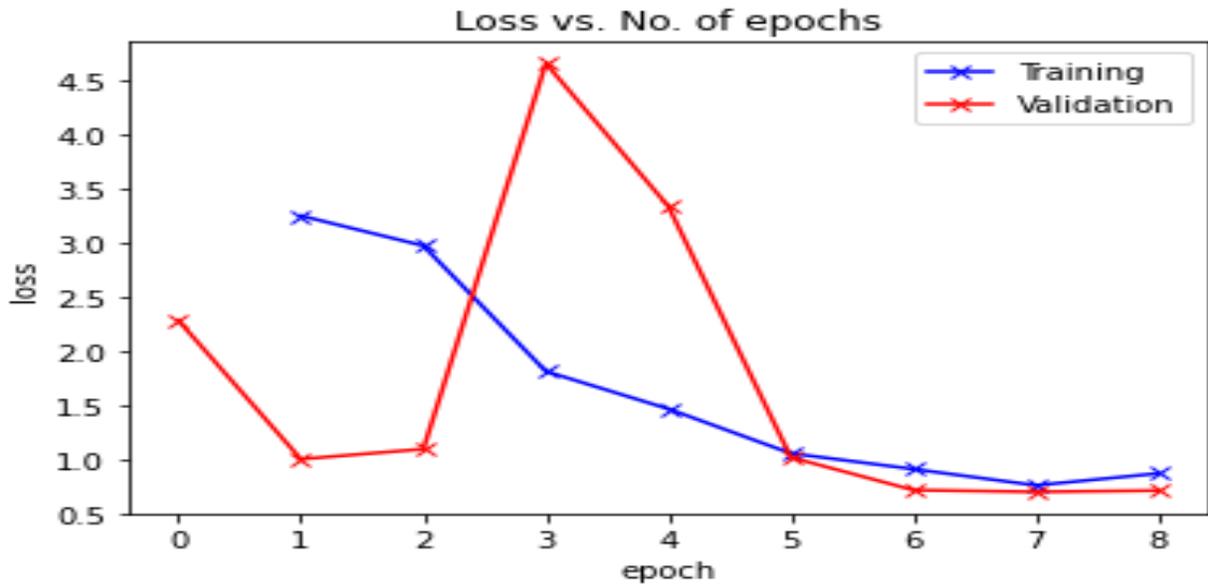


Figure 4.40: Loss of culture (indulgence) prediction with varying epoch based on CNN with residual block.

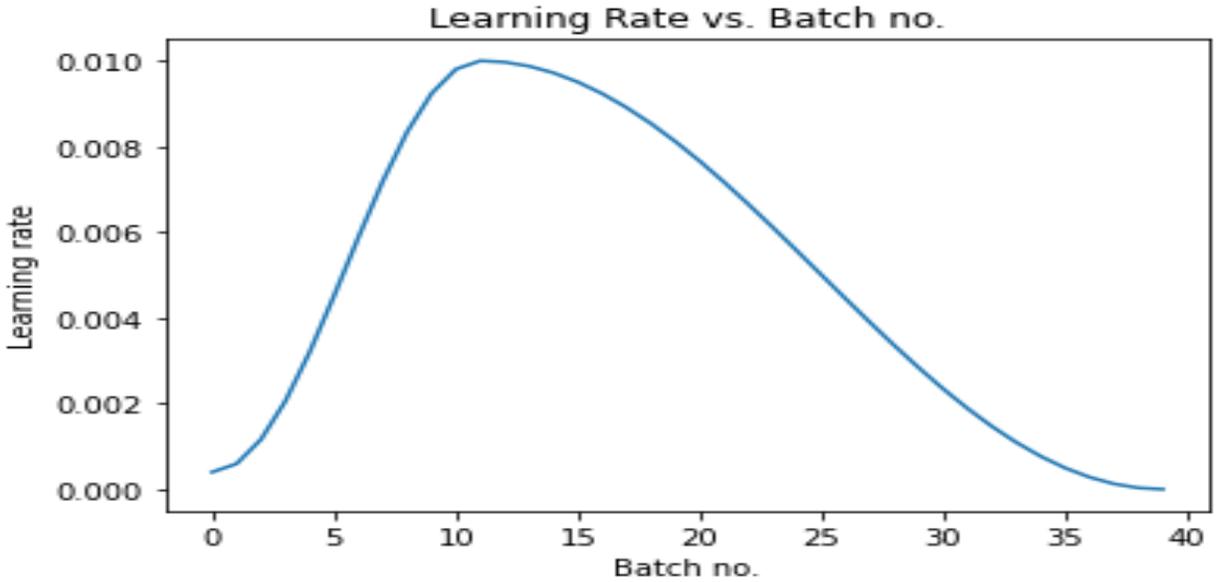


Figure 4.41: Learning rate of culture (indulgence) prediction with varying batch based on CNN with residual block.

estimate the true parameter, which is the next step in the study. Furthermore, the findings of this and future studies will aid in accessing and evaluating Nonpharmaceutical Interventions (NPIs) such as contact tracing, social distancing, masks, and quarantine as potential pandemic control strategies. Furthermore, we want to establish the CNN method and Algorithm 2 by allowing CNN to recognize a well-defined function as the two classes to identify, such as convex vs. concave, increasing vs. decreasing, Legendre vs. Chebychev, and oscillation vs. stable functions.

REFERENCES

- [1] *SEIR and SEIRS models*. <https://docs.idmod.org/projects/emod-generic/en/latest/model-SEIR.html>. Accessed: 2022-06-03.
- [2] *SIR and SIRS models*. <https://docs.idmod.org/projects/emod-generic/en/latest/model-sir.html>. Accessed: 2022-06-03.
- [3] *Modeling COVID-19 scenarios for the united states*, *Nature medicine*, 27 (2020), pp. 94–105.
- [4] D. ACEMOGLU, C. GARCÍA-JIMENO, AND J. A. ROBINSON, *State capacity and economic development: A network approach*, *American Economic Review*, 105 (2015), pp. 2364–2409.
- [5] L. ALLEN, *Stochastic population and epidemic models*, *Mathematical biosciences lecture series, stochastics in biological systems*, (2015).
- [6] L. J. ALLEN, *Introduction to mathematical biology*, Pearson/Prentice Hall, 2007.
- [7] A. D. BAI, X. X. LI, M. ALSALEM, S. KHAN, M. SMIEJA, D. MERTZ, AND Z. CHAGLA, *Utility of asymptomatic inpatient testing for COVID-19 in a low-prevalence setting: a multicenter point-prevalence study*, *Infection Control & Hospital Epidemiology*, 41 (2020), pp. 1233–1235.
- [8] J. J. V. BAVEL, K. BAICKER, P. S. BOGGIO, V. CAPRARO, A. CICHOCKA, M. CIKARA, M. J. CROCKETT, A. J. CRUM, K. M. DOUGLAS, J. N. DRUCKMAN, ET AL., *Using social and behavioural science to support COVID-19 pandemic response*, *Nature human behaviour*, 4 (2020), pp. 460–471.
- [9] P. H. BOERSCH-SUPAN, S. J. RYAN, AND L. R. JOHNSON, *Bayesian inference for a population growth model of the chytrid fungus*, See https://cran.r-project.org/web/packages/deBInfer/vignettes/chytrid_dede_example.pdf, (2016).
- [10] G. BONACCORSI, F. PIERRI, M. CINELLI, A. FLORI, A. GALEAZZI, F. PORCELLI, A. L. SCHMIDT, C. M. VALENSISE, A. SCALA, W. QUATTROCIOCCHI, ET AL., *Economic and social consequences of human mobility restrictions under COVID-19*, *Proceedings of the National Academy of Sciences*, 117 (2020), pp. 15530–15535.
- [11] C. BRADBURY-JONES AND L. ISHAM, *The pandemic paradox: The consequences of COVID-19 on domestic violence*, *Journal of clinical nursing*, (2020).
- [12] A. CANZIANI, A. PASZKE, AND E. CULURCIELLO, *An analysis of deep neural network models for practical applications*, arXiv preprint arXiv:1605.07678, (2016).

- [13] R. CLAIR, M. GORDON, M. KROON, AND C. REILLY, *The effects of social isolation on well-being and life satisfaction during pandemic*, Humanities and Social Sciences Communications, 8 (2021), pp. 1–6.
- [14] V. CLEMENS, P. DESCHAMPS, J. M. FEGERT, D. ANAGNOSTOPOULOS, S. BAILEY, M. DOYLE, S. ELIEZ, A. S. HANSEN, J. HEBEBRAND, M. HILLEGERS, ET AL., *Potential effects of “social” distancing measures and school lockdown on child and adolescent mental health*, 2020.
- [15] E. DUBÉ, C. LABERGE, M. GUAY, P. BRAMADAT, R. ROY, AND J. A. BETTINGER, *Vaccine hesitancy: an overview*, Human vaccines & immunotherapeutics, 9 (2013), pp. 1763–1773.
- [16] S. GARG, L. KIM, M. WHITAKER, A. O’HALLORAN, C. CUMMINGS, R. HOLSTEIN, M. PRILL, S. J. CHAI, P. D. KIRLEY, N. B. ALDEN, ET AL., *Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—covid-net, 14 states, march 1–30, 2020*, Morbidity and mortality weekly report, 69 (2020), p. 458.
- [17] I. GOODFELLOW, Y. BENGIO, AND A. COURVILLE, *Deep learning*, MIT press, 2016.
- [18] V. GRIMM, F. MENGEL, AND M. SCHMIDT, *Extensions of the SEIR model for the analysis of tailored social distancing and tracing approaches to cope with COVID-19*, Scientific Reports, 11 (2021), pp. 1–16.
- [19] C. GÜNSTER, R. BUSSE, M. SPODEN, T. ROMBEY, G. SCHILLINGER, W. HOFFMANN, S. WEBER-CARSTENS, A. SCHUPPERT, AND C. KARAGIANNIDIS, *6-month mortality and readmissions of hospitalized COVID-19 patients: a nationwide cohort study of 8,679 patients in germany*, PloS one, 16 (2021), p. e0255427.
- [20] R. GUPTA, S. GROVER, A. BASU, V. KRISHNAN, A. TRIPATHI, A. SUBRAMANYAM, A. NISCHAL, A. HUSSAIN, A. MEHRA, A. AMBEKAR, ET AL., *Changes in sleep pattern and sleep quality during COVID-19 lockdown*, Indian journal of psychiatry, 62 (2020), p. 370.
- [21] X. HAO, S. CHENG, D. WU, T. WU, X. LIN, AND C. WANG, *Reconstruction of the full transmission dynamics of COVID-19 in wuhan*, Nature, 584 (2020), pp. 420–424.
- [22] T. HASTIE, R. TIBSHIRANI, J. H. FRIEDMAN, AND J. H. FRIEDMAN, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, Springer, 2009.
- [23] K. HE, X. ZHANG, S. REN, AND J. SUN, *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [24] M. A. JOHANSSON, T. M. QUANDELACY, S. KADA, P. V. PRASAD, M. STEELE, J. T. BROOKS, R. B. SLAYTON, M. BIGGERSTAFF, AND J. C. BUTLER, *Sars-cov-2 transmission from people without COVID-19 symptoms*, JAMA network open, 4 (2021), pp. e2035057–e2035057.
- [25] Y.-J. KANG, *Mortality rate of infection with COVID-19 in korea from the perspective of underlying disease*, Disaster medicine and public health preparedness, 14 (2020), pp. 384–386.

- [26] S. H. KHOSHNAW, R. H. SALIH, AND S. SULAIMANY, *Mathematical modelling for coronavirus disease (COVID-19) in predicting future behaviours and sensitivity analysis*, Mathematical Modelling of Natural Phenomena, 15 (2020), p. 33.
- [27] A. KRIZHEVSKY, I. SUTSKEVER, AND G. E. HINTON, *Imagenet classification with deep convolutional neural networks*, Advances in neural information processing systems, 25 (2012).
- [28] S. KUSHWAHA, S. BAHL, A. K. BAGHA, K. S. PARMAR, M. JAVAID, A. HALEEM, AND R. P. SINGH, *Significant applications of machine learning for COVID-19 pandemic*, Journal of Industrial Integration and Management, 5 (2020), pp. 453–479.
- [29] Y. LECUN, Y. BENGIO, AND G. HINTON, *Deep learning*, nature, 521 (2015), pp. 436–444.
- [30] A. MACEDO, N. GONÇALVES, AND C. FEBRA, *COVID-19 fatality rates in hospitalized patients: systematic review and meta-analysis*, Annals of epidemiology, (2021).
- [31] J. MACKOLIL AND J. MACKOLIL, *Addressing psychosocial problems associated with the COVID-19 lockdown*, Asian journal of psychiatry, 51 (2020), p. 102156.
- [32] V. MALHOTRA, S. BASU, N. SHARMA, S. KUMAR, S. GARG, K. DUSHYANT, AND A. BORLE, *Outcomes among 10,314 hospitalized COVID-19 patients at a tertiary care government hospital in delhi, india*, Journal of Medical Virology, 93 (2021), pp. 4553–4558.
- [33] R. MBUVHA AND T. MARWALA, *Bayesian inference of COVID-19 spreading rates in south africa*, PloS one, 15 (2020), p. e0237126.
- [34] J. S. MIGDAL, *Strong societies and weak states: state-society relations and state capabilities in the Third World*, Princeton University Press, 1988.
- [35] K. MIZUMOTO, K. KAGAYA, A. ZAREBSKI, AND G. CHOWELL, *Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020*, Eurosurveillance, 25 (2020), p. 2000180.
- [36] H. MURAYAMA, R. OKUBO, AND T. TABUCHI, *Increase in social isolation during the COVID-19 pandemic and its association with mental health: Findings from the jacsis 2020 study*, International journal of environmental research and public health, 18 (2021), p. 8238.
- [37] K. P. MURPHY, *Machine learning: a probabilistic perspective*, MIT press, 2012.
- [38] S. S. NADIM AND J. CHATTOPADHYAY, *Occurrence of backward bifurcation and prediction of disease transmission with imperfect lockdown: A case study on COVID-19*, Chaos, Solitons & Fractals, 140 (2020), p. 110163.
- [39] T. O’HAGAN, *Bayes factors*, Significance, 3 (2006), pp. 184–186.
- [40] T. ORABY, M. G. TYSHENKO, AND S. BHATTACHARYYA, *Human cultural dimensions and behavior during COVID-19 can lead to policy resistance and economic losses: a perspective from game theory analysis*, in Current Perspectives on Viral Disease Outbreaks-Epidemiology, Detection and Control, IntechOpen.

- [41] J. REASON, *Human error: models and management*. *bmj (clin res ed)* 320 (7237): 768–770, 2000.
- [42] H. RITCHIE, E. MATHIEU, L. RODÉS-GUIRAO, C. APPEL, C. GIATTINO, E. ORTIZ-OSPINA, J. HASELL, B. MACDONALD, D. BELTEKIAN, AND M. ROSER, *Coronavirus pandemic (COVID-19)*, Our world in data, (2020).
- [43] S. J. RUSSELL, *Artificial intelligence a modern approach*, Pearson Education, Inc., 2010.
- [44] A. RĂDULESCU, C. WILLIAMS, AND K. CAVANAGH, *Management strategies in a SEIR-type model of covid 19 community spread*, Scientific reports, 10 (2020), pp. 1–16.
- [45] P. SAMUI, J. MONDAL, AND S. KHAJANCHI, *A mathematical model for COVID-19 transmission dynamics with a case study of india*, Chaos, Solitons & Fractals, 140 (2020), p. 110173.
- [46] H. SEALE, C. E. DYER, I. ABDI, K. M. RAHMAN, Y. SUN, M. O. QURESHI, A. DOWELL-DAY, J. SWARD, AND M. S. ISLAM, *Improving the impact of non-pharmaceutical interventions during COVID-19: examining the factors that influence engagement and the impact on individuals*, BMC Infectious Diseases, 20 (2020), pp. 1–13.
- [47] K. SIMONYAN AND A. ZISSERMAN, *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556, (2014).
- [48] R. SUBRAMANIAN, Q. HE, AND M. PASCUAL, *Quantifying asymptomatic infection and transmission of COVID-19 in new york city using observed cases, serology, and testing capacity*, Proceedings of the National Academy of Sciences, 118 (2021).
- [49] P. SUN, X. LU, C. XU, W. SUN, AND B. PAN, *Understanding of COVID-19 based on current evidence*, Journal of medical virology, 92 (2020), pp. 548–551.
- [50] R. SURESH, J. JAMES, AND B. RSJ, *Migrant workers at crossroads—the COVID-19 pandemic and the migrant experience in india*, Social Work in Public Health, 35 (2020), pp. 633–643.
- [51] C. SZEGEDY, W. LIU, Y. JIA, P. SERMANET, S. REED, D. ANGUELOV, D. ERHAN, V. VANHOUCKE, AND A. RABINOVICH, *Going deeper with convolutions*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
- [52] B. TANG, N. L. BRAGAZZI, Q. LI, S. TANG, Y. XIAO, AND J. WU, *An updated estimation of the risk of transmission of the novel coronavirus (2019-ncov)*, Infectious disease modelling, 5 (2020), pp. 248–255.
- [53] B. TEH, K. OLSEN, J. BLACK, A. C. CHENG, C. ABOLTINS, K. BULL, P. D. JOHNSON, M. L. GRAYSON, AND J. TORRESI, *Impact of swine influenza and quarantine measures on patients and households during the h1n1/09 pandemic*, Scandinavian journal of infectious diseases, 44 (2012), pp. 289–296.
- [54] L. YING AND T. XIAOQING, *COVID-19: Is it safe now? study of asymptomatic infection spread and quantity risk based on sair model*, Chaos, Solitons & Fractals: X, (2021), p. 100060.

- [55] S. ZHAO, L. STONE, D. GAO, S. S. MUSA, M. K. CHONG, D. HE, AND M. H. WANG, *Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in wuhan, china from 2019 to 2020*, *Annals of Translational Medicine*, 8 (2020).

BIOGRAPHICAL SKETCH

Salman Rahman was born in Bangladesh. Before moving to North America, he completed his undergraduate in engineering in August 2018 at Bangladesh. He completed his master's in applied statistics and data science in August 2022 at UTRGV, where he was awarded a presidential graduate research assistant-ship. He published several research in a reputed journal (Q1) on varying topics, including satellite remote sensing, applied machine learning, and computational sustainability. During his graduate study, he won several national and local competitions, including UTRGV big idea competition, Blackstone LaunchPad Ideas Competition, and Rafael Munguia Business Plan Competition. He is joining the New York University as a Computer Science Ph.D. student, where he aims to research deep learning and healthcare. Salman wants to empower the health care system with artificial intelligence and machine learning. He can be reached at salmanrahman350@gmail.com