COVID-19 Analysis and Forecasting using Harmony Search Optimization

A PROJECT REPORT

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

Electrical and Electronics Engineering

by

Biswajit Nanda 18BEE0084

Somdyuti Das Adhikary 18BEE0112

Manas Dixit 18BEE0117

Under the guidance of

Prof. M. N. Venkataraman

SCHOOL OF ELECTRICAL ENGINEERING

Vellore Institute Of Technology

VELLORE - 632014, Tamil Nadu, India

May 2022

DECLARATION

We hereby declare that the thesis entitled ***"COVID-19 Analysis and Forecasting using Harmony Search Optimization"*** submitted by us, for the award of the degree of *Bachelor of Technology* to VIT, Vellore is a record of bonafide work done under the supervision of Prof. M. N. Venkataraman.

We further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Biswajit Nanda

Somdyuti Das Adhikary

Manas Dixit

Place: Signature of the Candidates

Date:

CERTIFICATE

This is to certify that the thesis entitled ***"COVID-19 Analysis and Forecasting using Harmony Search Optimization"*** submitted by Biswajit Nanda (18BEE0084), Somdyuti Das Adhikary(18BEE0112), Manas Dixit (18BEE0117) School of Electrical Engineering, VIT, Vellore, for the award of the degree of Bachelor of Technology in Electrical and Electronics Engineering , is a record of bonafide work carried out by them under my supervision, as per the VIT code of academic and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The thesis fulfills the requirements and regulations of the University and in my opinion meets the necessary standards for submission.

Venkataraman M.V.

Place:Vellore
Date: 04/05/2022

Signature of the Guide

The thesis is satisfactory / unsatisfactory

Approved by

Head of the Department (EEE)

VELLORE INSTITUTE OF TECHNOLOGY
School of Electrical Engineering (SELECT)

N. Mathan Mithra

Dean

ACKNOWLEDGEMENTS

First and foremost we would like to thank our guide Prof. Venkataraman for his constant support and invaluable insight throughout the project. We've gained tremendous knowledge and experience and we are grateful to him. We would like to thank Prof. Venkataraman for his assistance with the MATLAB Software and his valuable inputs.

Last but not the least we extend our gratitude to the Chancellor Dr. G Viswanathan, Dean of SELECT Dr. Mathew M. Noel and the management at Vellore Institute Of Technology, Vellore for having given us this opportunity to study at a world-class University.

Biswajit Nanda
Somdyuti Das Adhikary
Manas Dixit

Executive Summary

The COVID-19 pandemic has resulted in a significant loss of human life around the world, and it poses an unprecedented threat to public health, food systems, and the workplace. The pandemic's economic and social effects are devastating: tens of millions of people are at risk of falling into extreme poverty, and the number of people who are undernourished, which is presently estimated to be around 690 million, might rise to 132 million by the end of the year.

The COVID-19 outbreak continues to harm communities such as those living in poverty, the elderly, individuals with disabilities, youth, and indigenous peoples. According to preliminary findings, poor individuals are bearing a disproportionate share of the virus's health and economic costs. Homeless persons, for example, are especially vulnerable to the virus because they may not be able to find a safe place to stay. People without access to running water, refugees, migrants, and displaced persons will be disproportionately affected by the epidemic and its aftermath whether as a result of restricted movement, fewer economic possibilities.

We, as Machine Learning enthusiasts aimed at doing this project to help tackle this pandemic in the best way we could. This pandemic motivated us to take up this project. Firstly, we collected the numerous datasets of COVID-19 cases, deaths, recoveries and vaccinations available in the internet. Detailed analysis of the datasets were done in Python using Jupyter Notebook. For the prediction of future COVID-19 cases and waves, we used a Deep Fuzzy Neural Network model with metaheuristic Harmony Search optimization technique to forecast the daily COVID-19 cases and deaths in the future from 100 upto 200 days.

Contents

List of Tables

List of Figures

List Of Abbreviations

SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
ML	Machine Learning
MSE	Mean Square Error
RMSE	Root Mean Square Error
HMS	Harmony Memory Size
HMCR	Harmony Memory Consideration Rate
PAR	Pitch Adjustment Rate
SVM	Support Vector Machine
KPI	Key Performance Indicator
JSON	Javascript Object Notation
SQL	Structured Query Language

Chapter 1

Introduction and Goals of the Project

1.1 Introduction

SARS-CoV-2 is the virus that causes the fatal sickness COVID-19. The World Health Organization first became aware of this virus on December 31, 2019, after receiving a report of numerous cases of 'viral pneumonia' from Wuhan, China. Since then, the virus has been designated a pandemic, with over 66,729,375 cases reported across 220 countries and 1,535,982 deaths were recorded.This study demonstrates the capability of ML models to forecast the number of upcoming patients affected by COVID-19 which is presently considered as a potential threat to mankind [1].

ML algorithms' learning is typically based on trial and error method quite opposite of conventional algorithms, which follows the programming instructions based on decision statements like if-else[2]. One of the most significant areas of ML is forecasting[3], numerous standard ML algorithms have been used in this area to guide the future course of actions needed in many application areas including weather forecasting, disease forecasting, stock market forecasting as well as disease prognosis. Various regression and neural network models have wide applicability in predicting the conditions of patients in the future with a specific disease. There are lots of studies performed for the prediction of different diseases using machine learning techniques such as coronary artery disease, cardiovascular disease prediction, and breast cancer prediction. These prediction systems can be very helpful in decision making to handle the present scenario to guide early interventions to manage these diseases very effectively and efficiently. Coronavirus (COVID-19) is a viral disease caused by severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2). The spread of COVID-19 seems to have a detrimental effect on the global economy and health[4].

1.2 Goals of the Project

In particular, the study[5] is focused on live forecasting of COVID-19 confirmed cases and study[6] is also focused on the forecast of COVID-19 outbreak and early response. These prediction systems can be very helpful in decision making to handle the present scenario to guide early interventions to manage these diseases very effectively. The COVID-19 waves have taught us a lot about the efficacy of different societal responses. To begin, having up-to-date vaccinations, including a recent booster, was found to be particularly useful in defending against COVID-19. Hospitalizations were significantly decoupled from cases in countries where a considerable fraction of patients at risk had received three doses of vaccine, including at least one dose of mRNA vaccine.

This project has done extensive research on analysis of the data and Covid cases from various sources.After the analysis, major focus was put on to the forecasting of the possible daily cases and deaths. We used a Metaheuristic Genetic Algorithm called Harmony Search Algorithm to develop our model.

First, we have done a time series analysis of the daily COVID-19 cases of different countries. A fuzzy based neural network model is used and then we have used the Genetic Algorithm HS to optimize the cost function using repeated iterations. The parameters of the Harmony Search: Maximum Iterations, HMS, HMCR, PAR and Termination criterion are initialized and tuned to obtain the best predictive curves. In every iteration, a new Harmony from Harmony Memory based on pitch adjustments and randomization is done and if it is better than previous HM, the new Harmony is updated in the Harmony Memory.

After successful optimization of the algorithm we also tried to predict the Covid-19 cases of countries that did not have the Fourth wave and countries that already had the 4th wave. The cases predicted were very accurate as the daily upcoming cases are very close to the one predicted by the algorithm.

In many countries, the six-month forecast is better than it has been in the previous two years. However, a number of concerns, beginning with the duration of immunity, could dampen enthusiasm. Both natural and vaccine-induced immunity, particularly against viruses, appears to diminish over time. The ultimate goal of this project is to predict the possible rise in Covid cases and eventually increased death cases in a particular country or area. The future forecast would help that country to be prepared and take precautionary measures to reduce the loss of human life at a maximum possible scale.

Chapter 2

Analysis of COVID-19 Data

Exploratory data analysis is a prerequisite to any machine learning task. It gives an insight about the nature of the data and which Machine Learning algorithm would be most appropriate. In this chapter we focus on a comparative study of the effect of coronavirus in India and other countries. We have focused our analysis on 4 countries viz. India, USA, Zimbabwe and South Africa. Specifically these countries are our point of interest because they showed different Covid-19 wave patterns. India and USA are countries with huge population and have experienced 3 waves of Covid-19. On the other hand Zimbabwe and South Africa are relatively small countries population wise and have experienced the 4th wave of Covid-19 already.

The datasets for the aforesaid analysis is taken from Git Hub [7],[8],[9] registry, supplied by Johns Hopkins University, Systems Science and Engineering Centre. These datasets are open source and are accessible to everyone. Our area of focus was increase in daily confirmed cases and deaths so these were extracted first. Since daily cases and deaths were producing highly jagged curves, it was best to take the 7-day rolling average as our input data for further analysis. To make the visualization easy and suitable for Machine Learning models to train, we numbered the days chronologically starting from 01/30/2020 instead of the usual date-time format.For better context, the worldwide cases and deaths is also shown along with the graphs of the specific countries.

As seen in Fig.2.7 the first wave of Covid-19 hit the country on 25 April 2020 and continued till 7 February 2021 making it one of the most prolonged wave that claimed numerous lives as seen in Fig.2.8. The second wave began on 4 March 2021 and carried on till 30 October 2021. Most recently the third wave hit on 26 December 2021 and carried on till 15 March 2022.

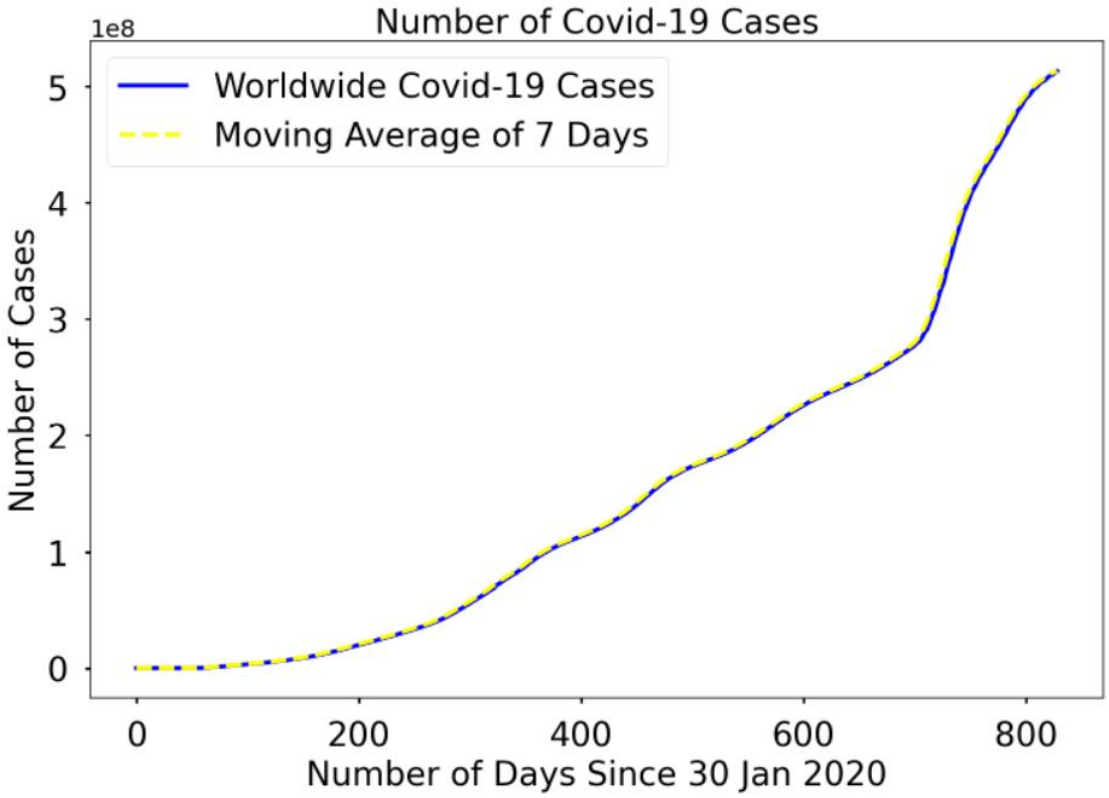

Figure 2.1: Worldwide COVID-19 Cases

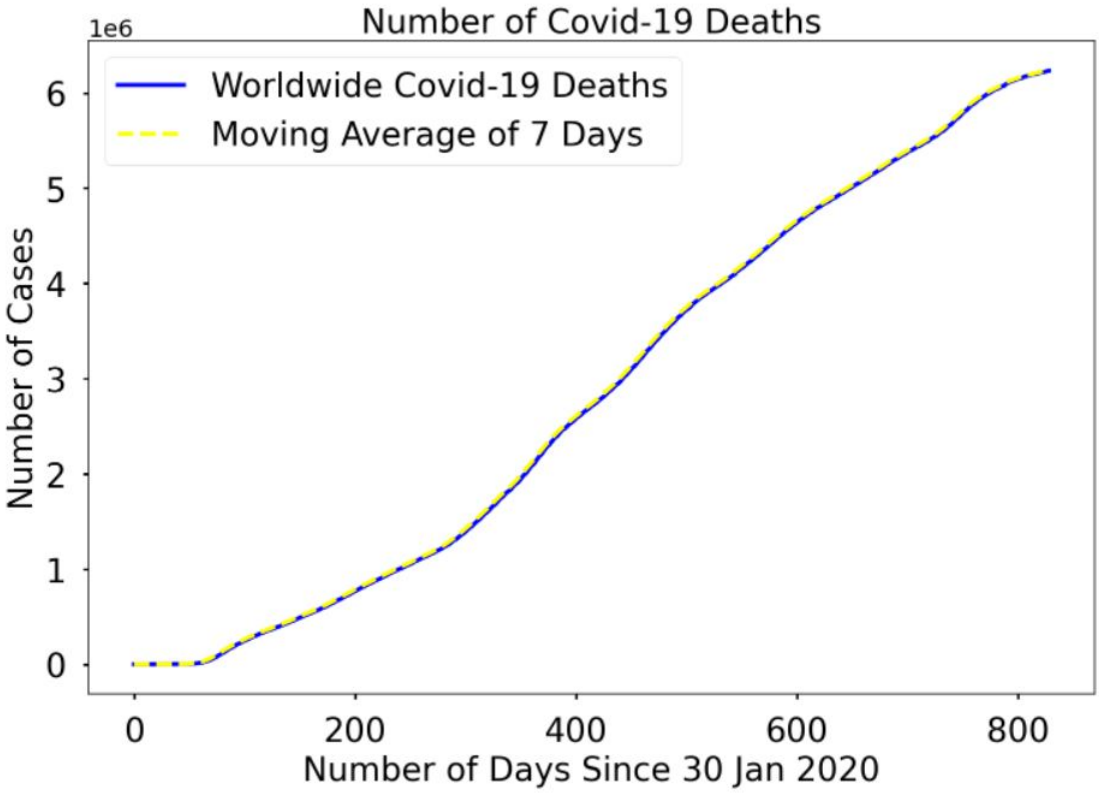

Figure 2.2: Worldwide COVID-19 Deaths

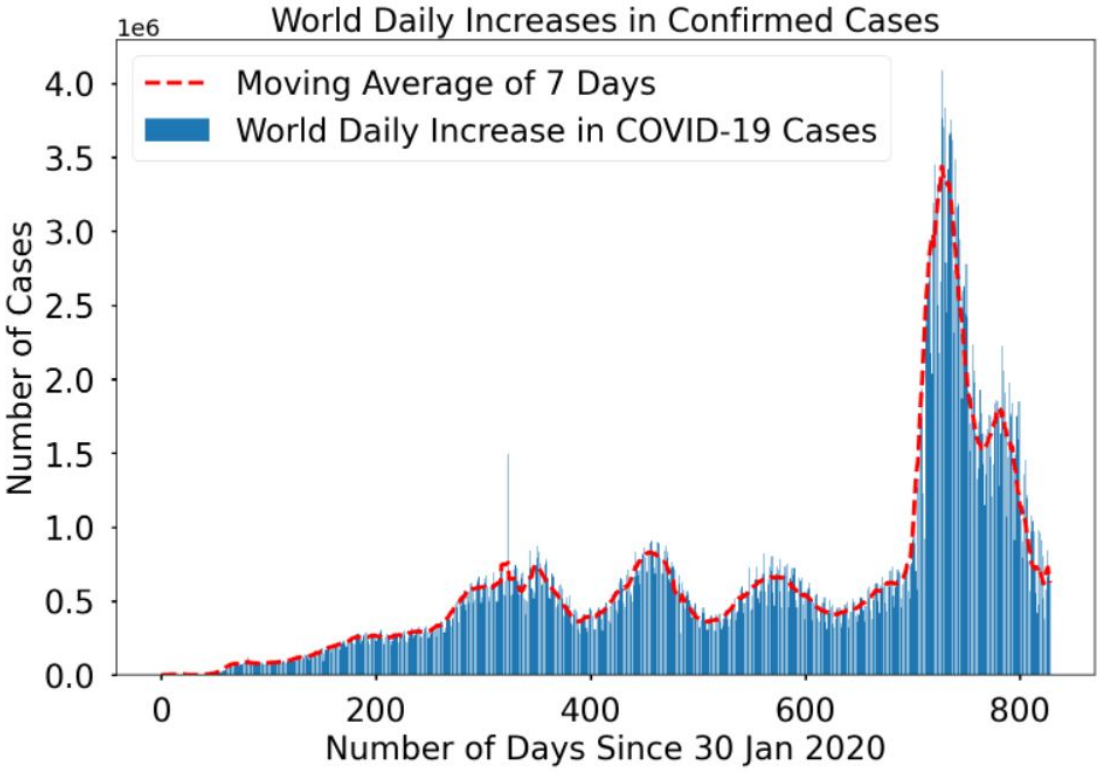

Figure 2.3: Worldwide Daily Cases

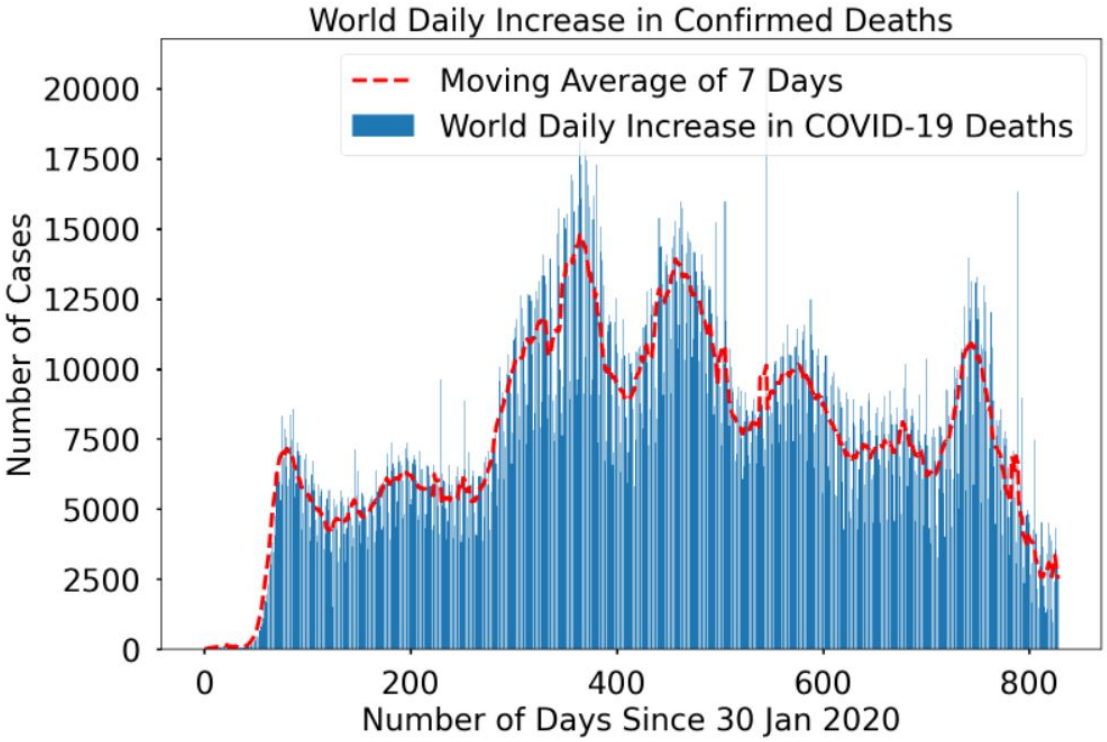

Figure 2.4: Worldwide Daily Deaths

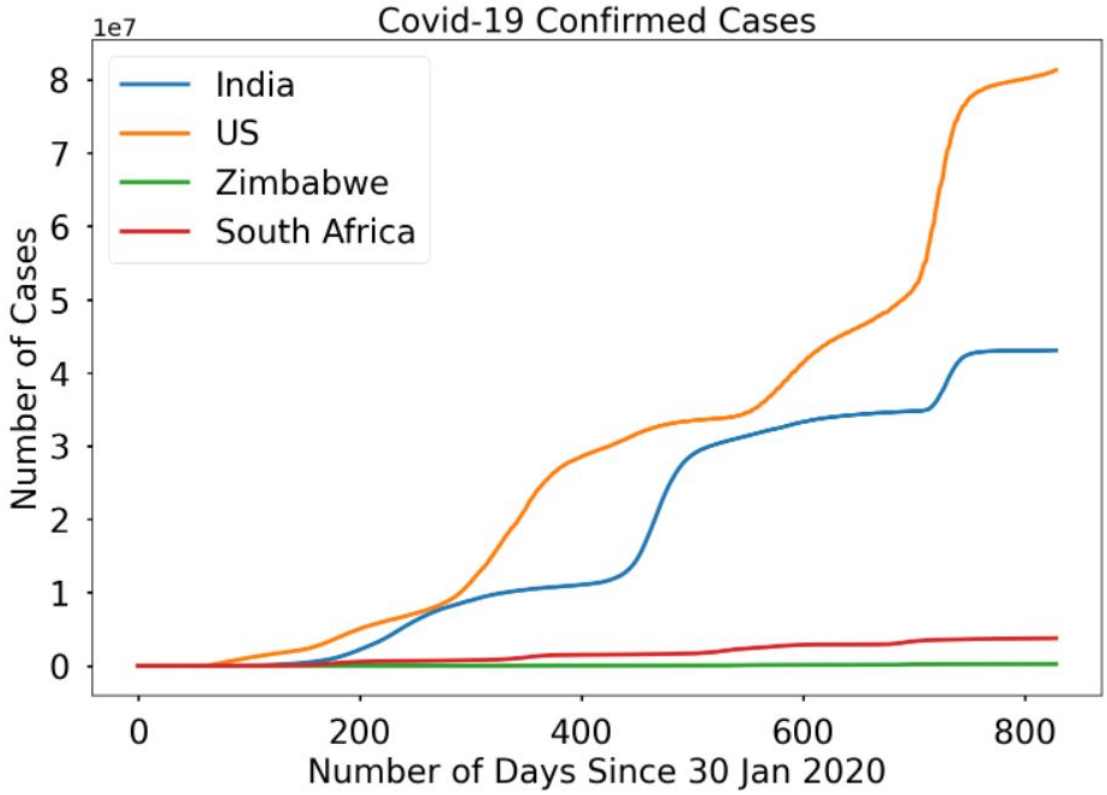

Figure 2.5: Countrywise Coronavirus Cases

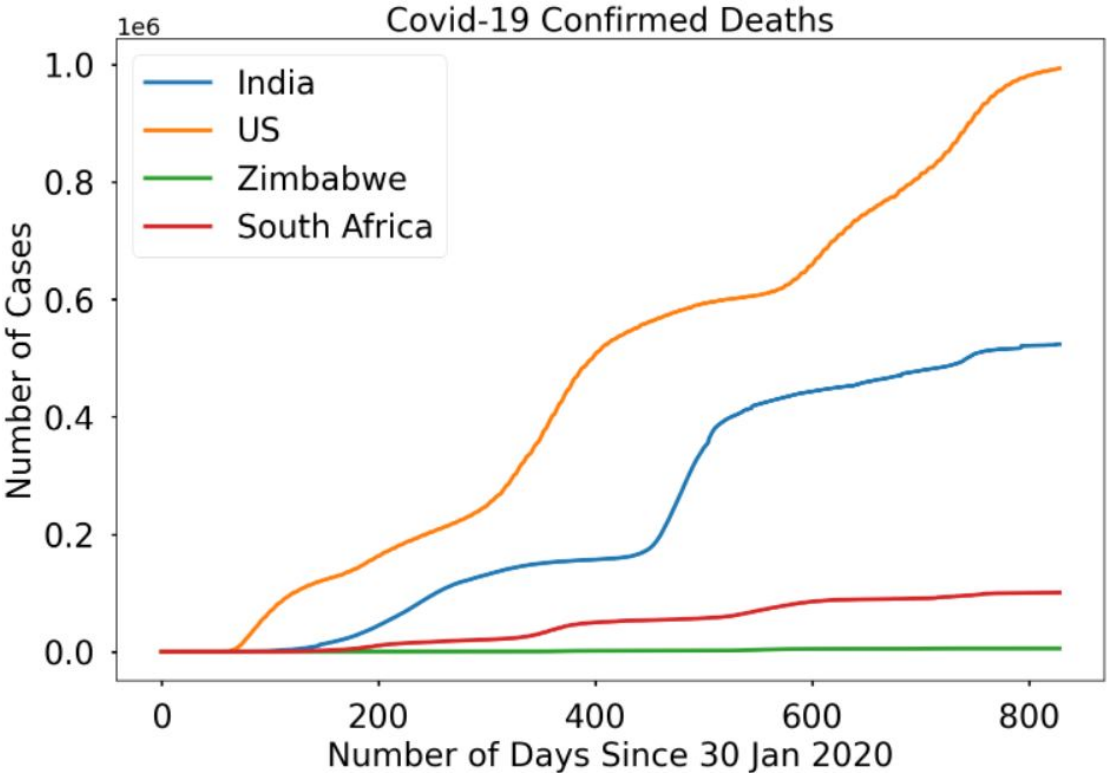

Figure 2.6: Countrywise Coronavirus Deaths

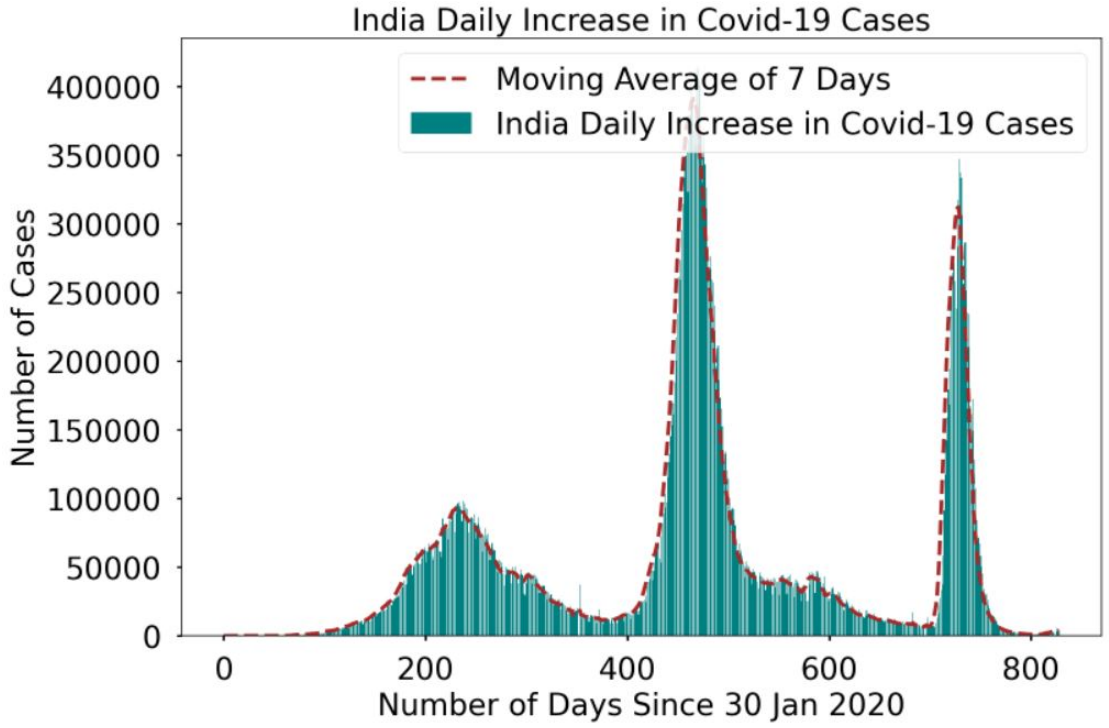

Figure 2.7: India COVID-19 Daily Cases

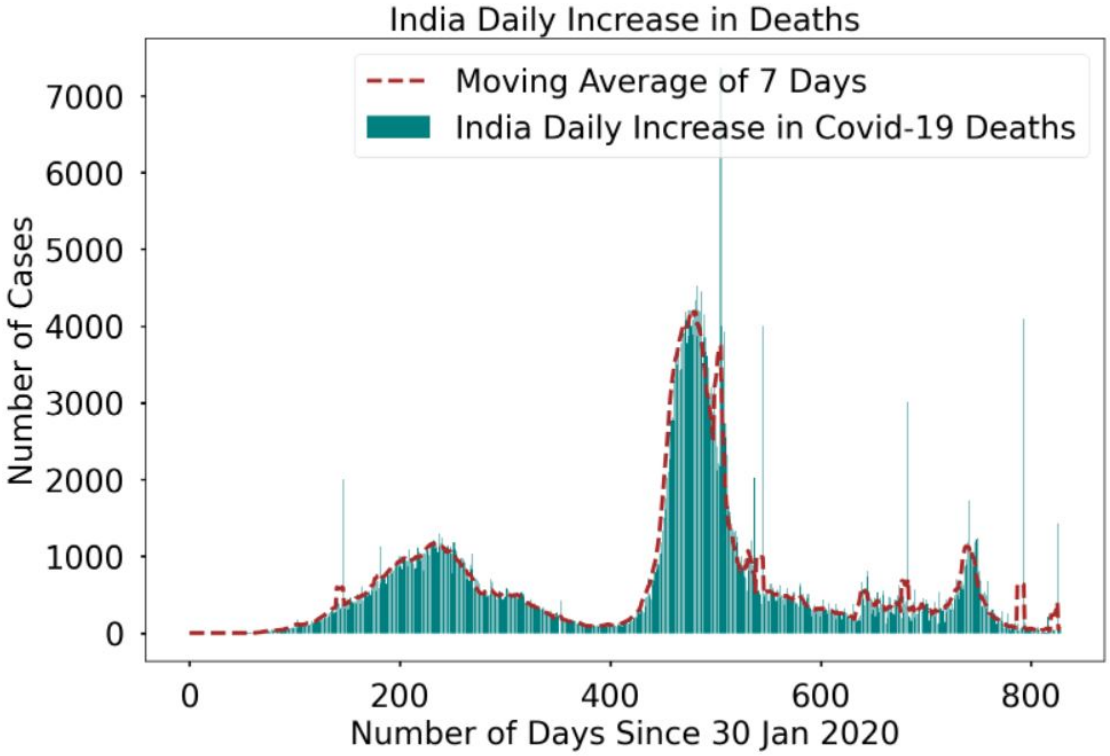

Figure 2.8: India COVID-19 Daily Deaths

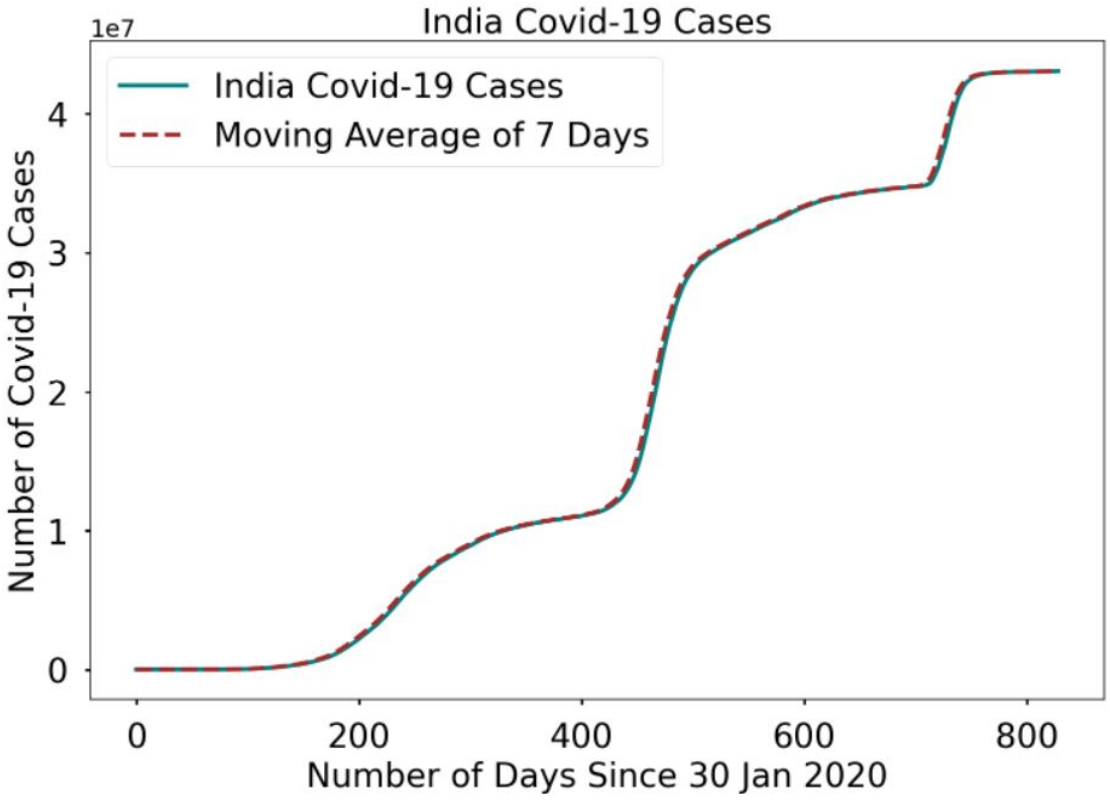

Figure 2.9: India Total COVID-19 Cases

Chapter 3

Prediction of COVID-19 parameters

3.1 Conventional Algorithms

Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Basically, supervised learning is when we teach or train the machine using data that is well labeled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labeled data [10].

The complete analysis is done of the COVID-19 situation in many countries all over the world. A comparative study of the accuracy of different ML algorithms is done. ML algorithms' learning is typically based on trial-and-error method quite opposite of conventional algorithms, which follows the programming instructions based on decision statements like if-else. One of the most significant areas of ML is forecasting, numerous standard ML algorithms have been used in this area to guide the future course of actions needed in many application areas including weather forecasting, disease forecasting, stock market forecasting as well as disease prognosis. Various regression and neural network models have wide applicability in predicting the conditions of patients in the future with a specific disease. There are lots of studies performed for the prediction of different diseases using machine learning techniques such as coronary artery disease, cardiovascular disease prediction, and breast cancer prediction. In particular, the study is focused on live forecasting of COVID-19 confirmed cases and study is also focused on the forecast of COVID-19 outbreak and early response. These prediction systems can be very helpful in decision making to handle the present scenario to guide early interventions to manage these diseases very effectively.

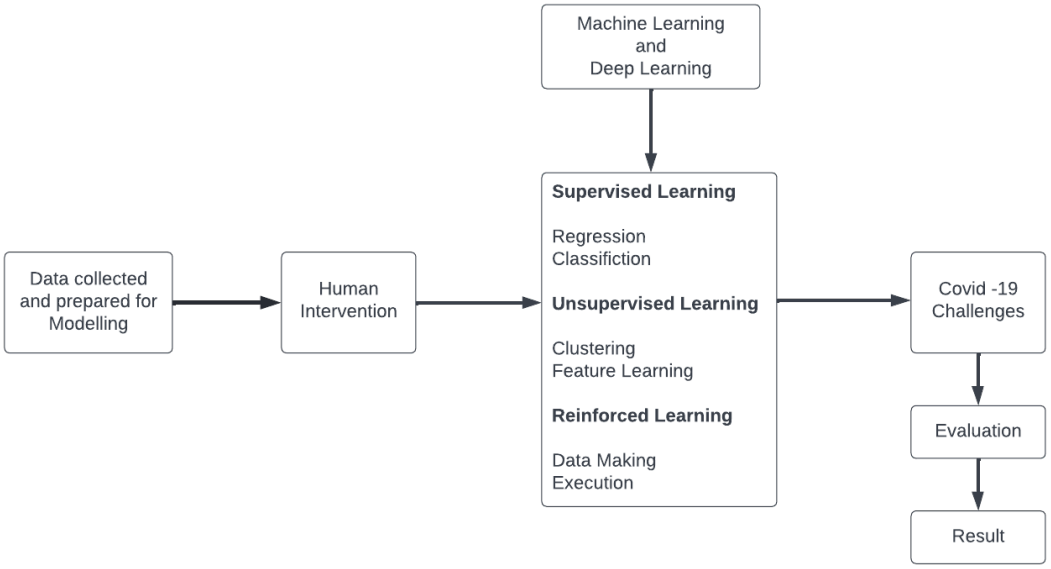

Figure 3.1: Block diagram of the dataflow model of prediction and analysis of COVID-19 using ML

After performing sufficient analysis on the dataset, it was our objective to predict the cases and deaths upto a reasonable number of days in the future. Since the Covid-19 cases are a time series data we performed the prediction using basic conventional algorithms like Support Vector Machines(SVM) for regression, Polynomial regression and Bayesian Ridge Regression.

3.1.1 Regression using SVM

Support Vector Machine algorithm is a popular Supervised learning algorithm finding its use more in machine learning problems based on Classification. The main objective of this algorithm is to find a hyperplane within an N-dimensional space (where N refers to the number of features) which is then able to classify data and sort new data points into their appropriate categories later on. A hyperplane is essentially a decision boundary. Furthermore, the algorithm chooses extreme points called vectors in order to create the required hyperplane and hence the name Support Vector Machine. The following Equations 3.1 & 3.2 govern the algorithm.

$$f(x) = x'\beta + b, \tag{3.1}$$

$$J(\beta) = \frac{1}{2}\beta'\beta \tag{3.2}$$

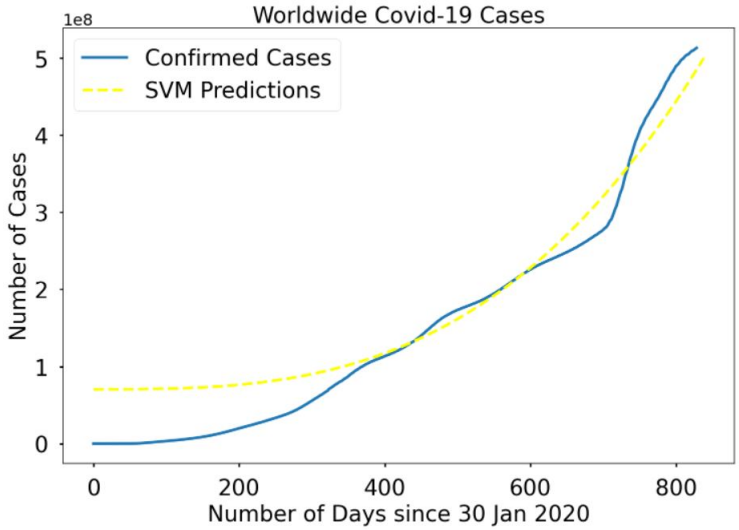

Figure 3.2: SVM Prediction of WorldWide Coronavirus Cases

As shown in Fig.3.2, the Support Vector Machine prediction model is erroneous in the beginning and then becomes more accurate as the number of days increase. Overall, it yields better accuracy than Polynomial Regression.

3.1.2 Polynomial Regression

Polynomial regression is a type of regression analysis where the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted $E(y|x)$. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the meaning that the regression function $E(y|x)$ is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a exceptional case of multiple linear regression.

An nth degree Polynomial regression is given by the Equation 3.3

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \beta_n x^n + \varepsilon \tag{3.3}$$

As shown in Fig.3.3, the Polynomial Regression model is erroneous in the beginning and then becomes more accurate as the number of days increase. Polynomial Regression will not be accurate in predicting the total number of COVID-19 cases as we cannot take into consideration all the parameters responsible for COVID-19 increase or outbreak in a region.

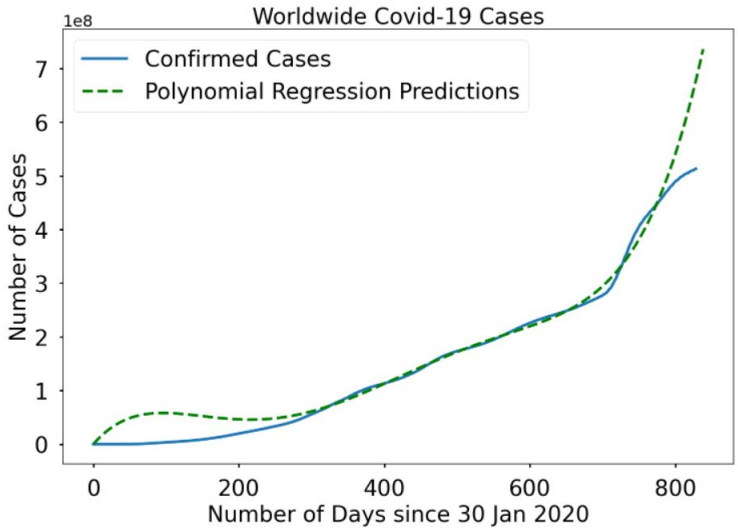

Figure 3.3: Polynomial Regression Prediction of Worldwide Coronavirus Cases

3.1.3 Bayesian Regression

Bayesian regression allows a natural mechanism to survive insufficient data or poorly distributed data by formulating linear regression using probability distributors rather than point estimates. The output or response y is assumed to drawn from a probability distribution rather than estimated as a single value.

Mathematically, to obtain a fully probabilistic model the response y is assumed to be Gaussian distributed around X_w as shown in Equation 3.4

$$p(y|X, w, \alpha) = N(x, X_w, \alpha) \tag{3.4}$$

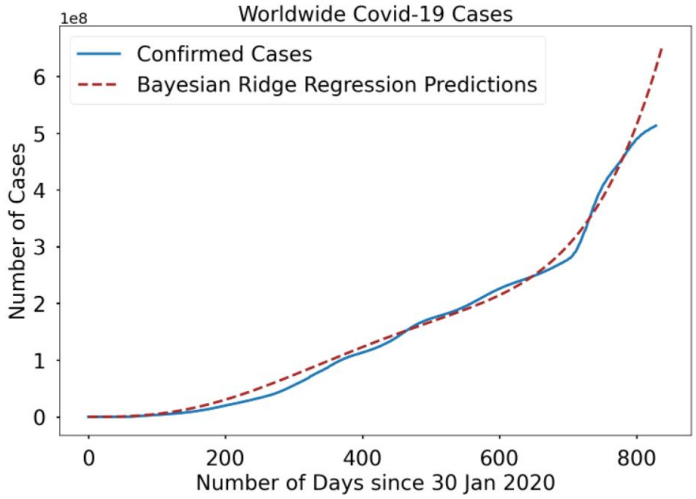

Figure 3.4: Bayesian Ridge Regression of Worldwide Coronavirus Cases

As shown in Fig.3.4, the Bayesian Ridge Regression model is the most accurate among the three regression models.

3.2 Neural Network with Harmony Search Optimization

Neural networks are a series of algorithms that are inspired by the biological neurons present in the human brain. These algorithms take in input and using target functions, they try to reach an output which is close to the target output. The algorithms endeavor to find a relationship between data and optimize the solution using different deep learning techniques. Usually, we use Gradient Search Algorithms to find the local minima using Gradient descent. However, there are certain limitations to this algorithm. The function should have only one minima for the Gradient Search Algorithm to work properly. The function should be continuous and differentiable otherwise the Gradient Search will not work on the entire workspace. To overcome these limitations, we use meta-heuristic algorithms which can work on non-differentiable functions and search the entire workspace for the global minima.

The proposed methodolgy is to use a Metaheuristic algorithm to optimize the neural network to find a better solution. Metaheuristic solutions are general purpose optimization techniques. It searches the entire workspace and finds the global optimum solution. These are nature inspired phenomena and they use stochastic optimization, which uses random numbers. Metaheuristic algorithms are a combination of two extremes namely, random search(extreme exploration) and hill climbing(extreme exploitation). Under metaheuristic algorithms fall the Harmony Search Algorithm which is a population based searching algorithm. Harmony Search was inspired by musicians, when they compose a harmony, they try out different combinations of the music pitches from memory. If the musician finds a better harmony in the vicinity of the previous harmony, they update the harmony in his memory. They search for the perfect harmony and in a similar fashion Harmony Search searches for the optimal solution [11].

As shown in Fig.3.5 , the first step in the harmony search is to initialize and tune the harmony search parameters, Harmony Memory Size(HMS), Harmony Memory Consideration Rate(HMCR) and Pitch Adjustment Rate(PAR) and Termination Criterion. We are required to initialize the parameters to best fit the data of COVID-19 daily cases and deaths of different countries such as India, USA, South Africa and Zimbabwe. In

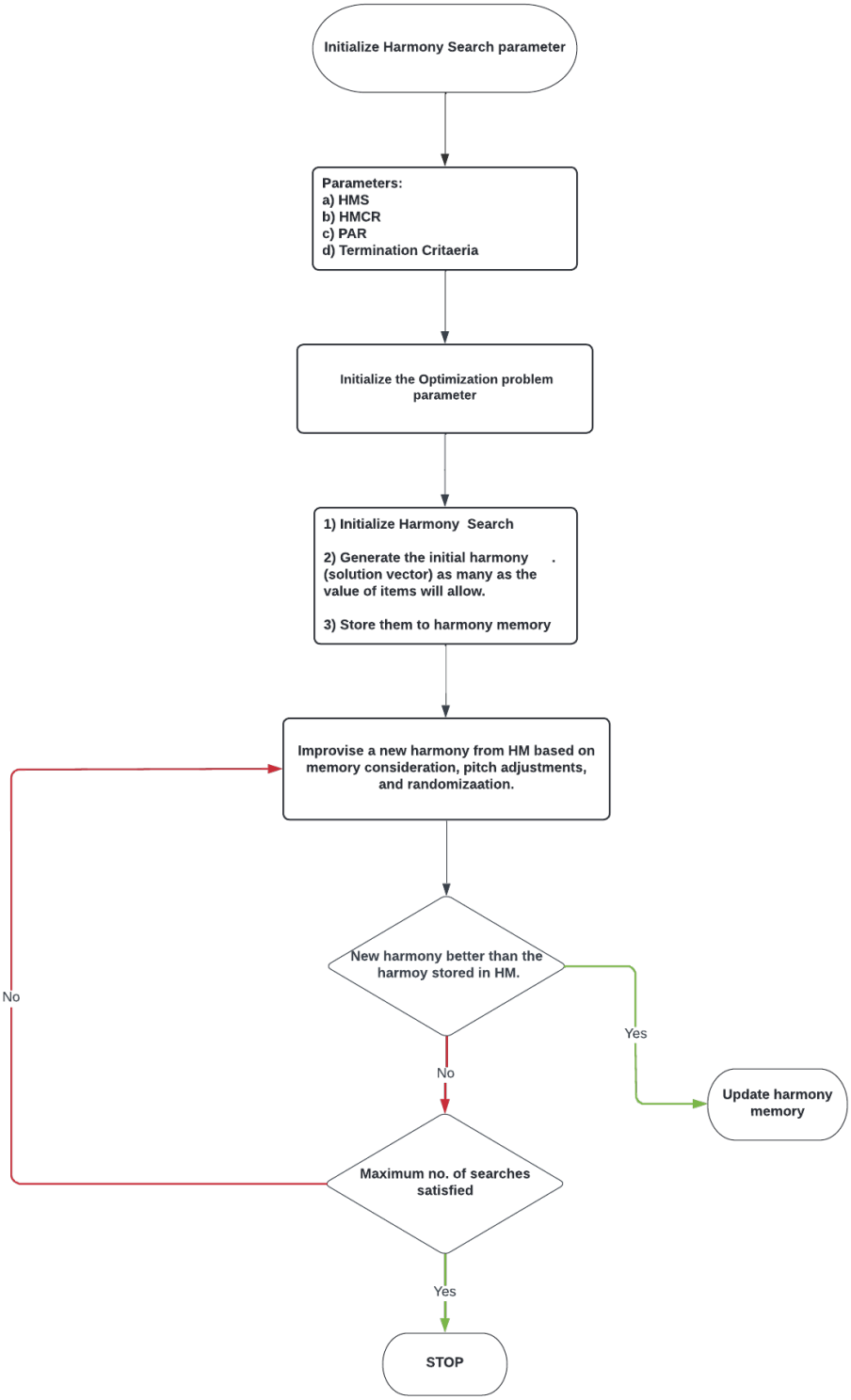

Figure 3.5: Dataflow Diagram of Harmony Search Algorithm

every iteration, the algorithm tries to improvise a new harmony based on memory consideration, pitch adjustments and randomization. Then it checks if the new harmony is better than the one stored in memory, and if yes, it updates the harmony memory. This process continues till the maximum number of iterations is reached.

A fuzzy based neural network model is used and the time series data of the daily cases of a country is fed into it. Then we use the harmony search optimization technique to optimize the output. To obtain the best fit data, we initialized the Maximum Iteration = 3000, HMS = 3, HMCR = 0.9 and PAR = 0.1, and we aimed at forecasting the COVID-19 daily cases and deaths for the next 100 days. We took the dataset of different countries for over 800 days starting from January 30,2020 to April 8,2022. The total dataset was split into a 80:20 ratio, where 80% of the training dataset was used for training and 20% was used for testing.

The data points for the training dataset were observed to be within the prediction bounds as shown in Fig.3.7. The red line is the best fit curve for the training set after training the dataset. The Harmony Search Cost Function decreases with every iteration and then becomes stable as shown in Fig.3.6.

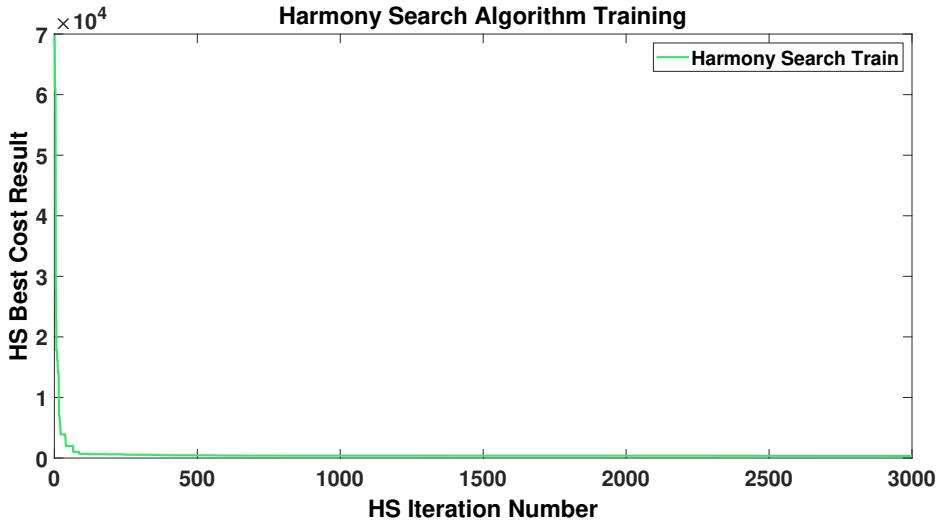

Figure 3.6: Harmony Search Best Cost Result vs Number of Iterations for Training Algorithm of India

The HS algorithm simulates music improvisation process and establishes a mathematical model to get sub-optimal solution or nearly optimal solution within the limited time. The improvisation of the HS algorithm is to update HS memory according to variation of random number, pitch-adjusted and random variation calculation. It is a global random search method.

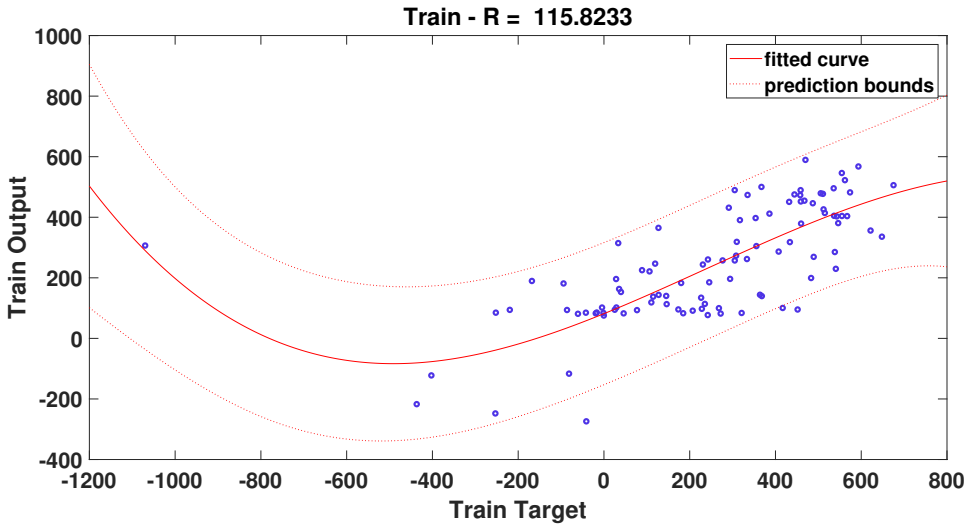

Figure 3.7: Prediction Bound of HS training of India

However, the HS algorithm and its variants also have some shortcomings such as the sensitivity on the control parameter, the blindness of searching, the low convergence rate and the tendency to stagnate. This prompted researchers to find new ideas to improve the deficiency of the HS algorithms. The research on HS algorithm is still active and important for applications[12].

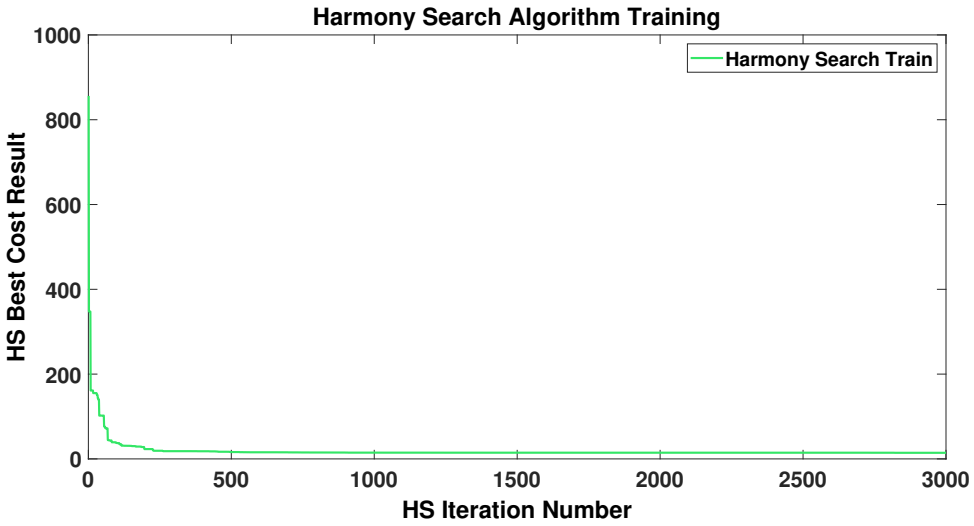

Figure 3.8: Harmony Search Best Cost Result for Zimbabwe

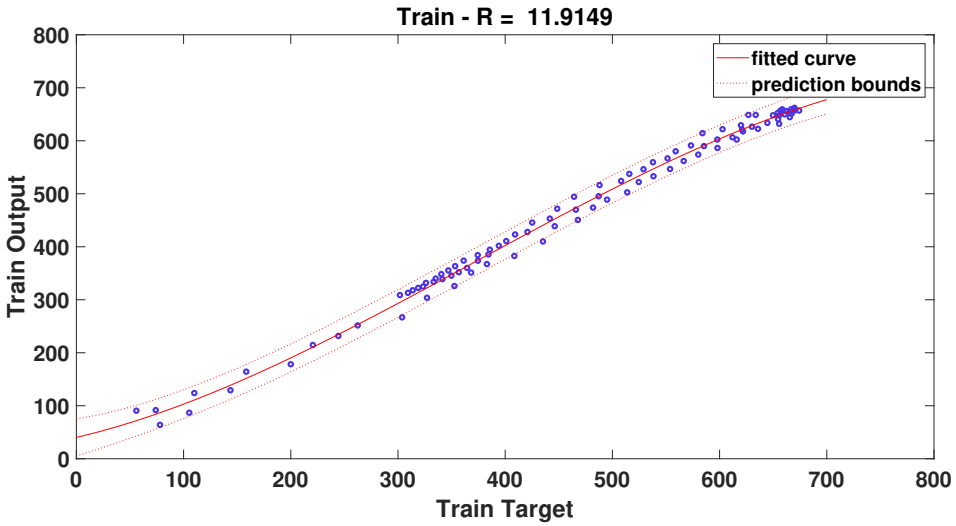

Figure 3.9: Prediction Bound of HS training of Zimbabwe

Chapter 4

Results and Discussion

4.1 Forecasting of Covid-19 Cases & Deaths

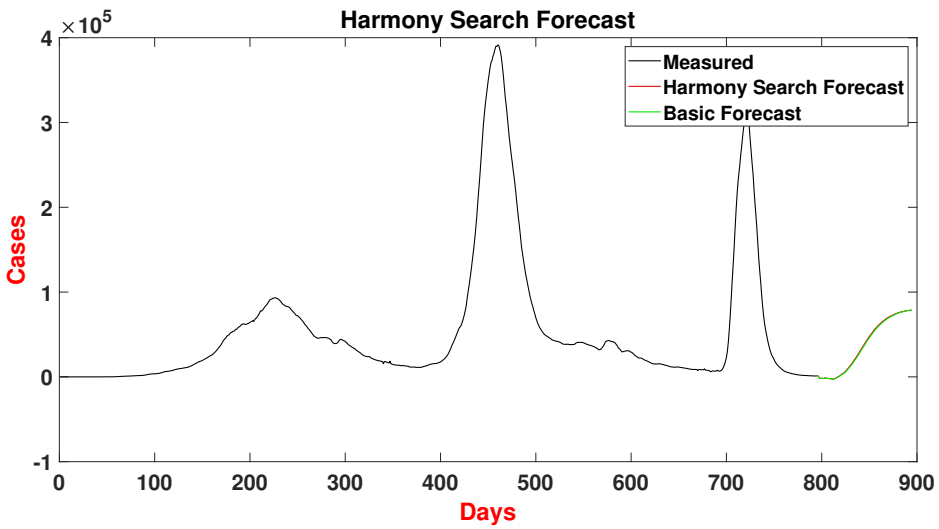

Figure 4.1: India COVID-19 Daily Cases Forecasting

We can observe from the graphs of the different countries that there is a possibility of a new COVID-19 wave emerging in the near future. The forecasting of the next 100 days in India suggests that there will be a fourth wave starting from June and continuing till August. The wave will reach its peak around the middle of July. The severity of the fourth wave will be less than that of the first, second and third waves. USA follows a similar pattern as India and might observe a fourth wave in the near future. In other countries like Zimbabwe and South Africa which already witnessed their fourth waves, there is a possibility of fifth wave.

The forecasting of the next 100 days for the 800 days dataset seemed to be quite ac-

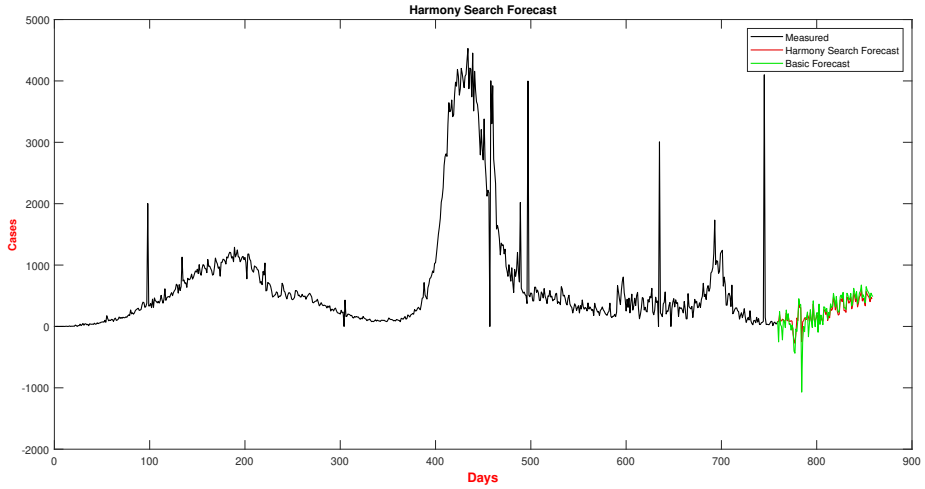

Figure 4.2: India COVID-19 Death Forecasting

curate with very low Mean Squared Error(MSE) and Root Mean Squared Error(RMSE). When we try to increase the forecasting days to 180 days, we are able to predict the daily COVID-19 cases and deaths with an increased MSE and RMSE errors. A good accuracy prediction can be done upto 25% of the days of the dataset using the Harmony Search algorithm. Increasing the dataset leads to increase in accuracy of prediction.

4.1.1 Mean Square Error (MSE)

The mean absolute error is the average magnitude of the errors in the set of model predictions [13],[14]. This is an average on test data between the model predictions and actual data where all individual differences have equal weight. Its matrix value range is from 0 to infinity and fewer score values show the goodness of learning models that's the reason it's also called negatively-oriented scores [15]. Mean square error is another way to measure the performance of regression models [13]. MSE takes the distance of data points from the regression line and squaring them. Squaring is necessary because it removes the negative sign from the value and gives more weight to larger differences. The smaller mean squared error shows the closer you are to finding the line of best fit. MSE can be calculated as given by the following Equation 4.1.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{4.1}$$

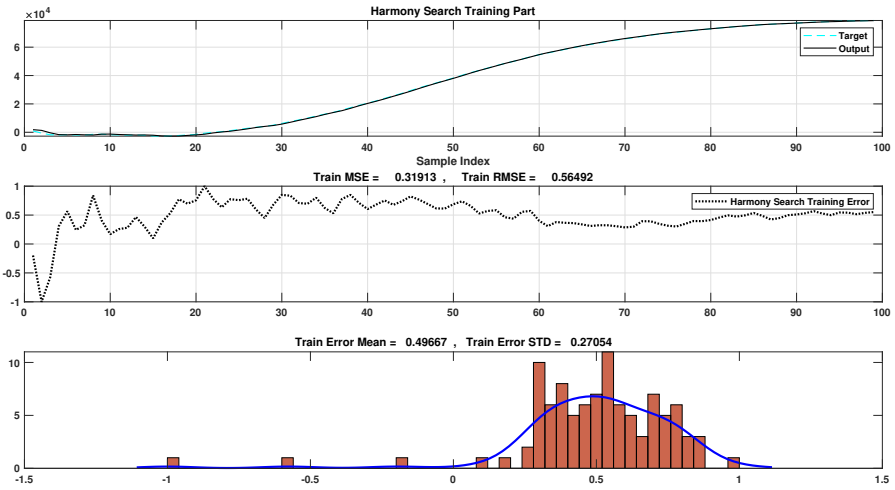

Figure 4.3: MSE and RMSE of Training dataset of COVID-19 cases in India

4.1.2 Root Mean Square Error (RMSE)

Root mean square error can be defined as the standard deviation of the prediction errors. Prediction errors also known as residuals is the distance from the best fit line and actual data points. RMSE is thus a measure of how concentrated the actual data points are around the best fit line. It is the error rate given by the square root of MSE as shown in Equation 4.2.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \tag{4.2}$$

Performance Comparison of Conventional Algorithms		
Algorithm	Mean Square Error	Root Mean Square Error
Support Vector Machine	17329914.5	4162.4
Polynomial Regression	81395714.3	9021.9
Bayesian Ridge Regression	1780551.9	1334.3

Table 4.1: Training performance of Conventional Algorithms

Table 4.1 illustrates the KPI's of the Conventional algorithms like SVM, Polynomial Regression and Bayesian Ridge Regression during training. Among them SVM and Bayesian Ridge Regression performed nearly similarly while polynomial regression did not yield good results.

Table 4.2 compares the training KPI's of the 4 countries India, USA, Zimbabwe and South Africa. India's cases and deaths were analyzed separately. India and USA had

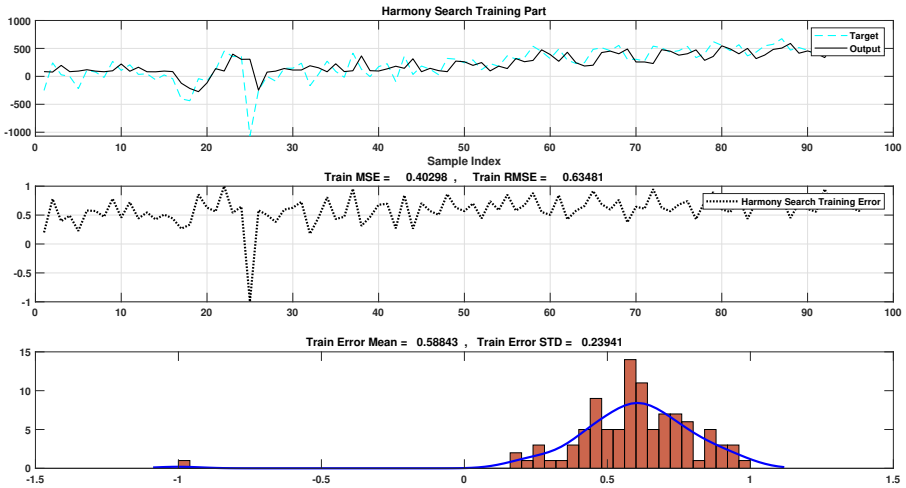

Figure 4.4: MSE and RMSE of Training dataset of COVID-19 deaths in India

their third wave while Zimbabwe and South Africa had experienced the 4th wave of Covid-19.

Country Wise Comparison Of Harmony Search Algorithm				
Country	Train MSE	Train RMSE	Train Mean	Train STD
India Cases	0.31913	0.56492	0.49667	0.27054
India Deaths	0.4298	0.63481	0.58843	0.23941
United States of America	0.1916	0.43772	0.19203	0.39535
Zimbabwe	0.16117	0.40146	-0.0205	0.40297
South Africa	0.19154	0.43766	0.15408	0.41172

Table 4.2: Table showing the different training KPIs of 4 countries

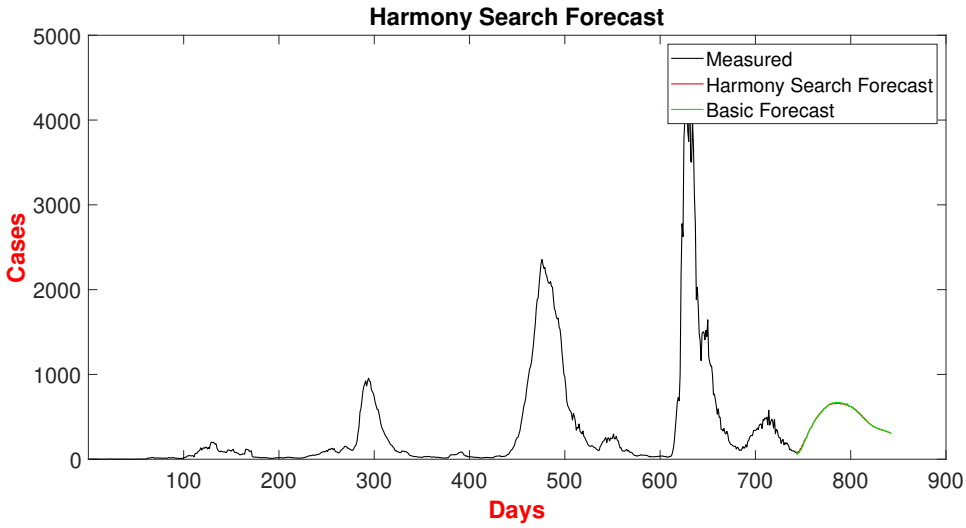

Figure 4.5: Zimbabwe COVID-19 Daily Cases Forecasting

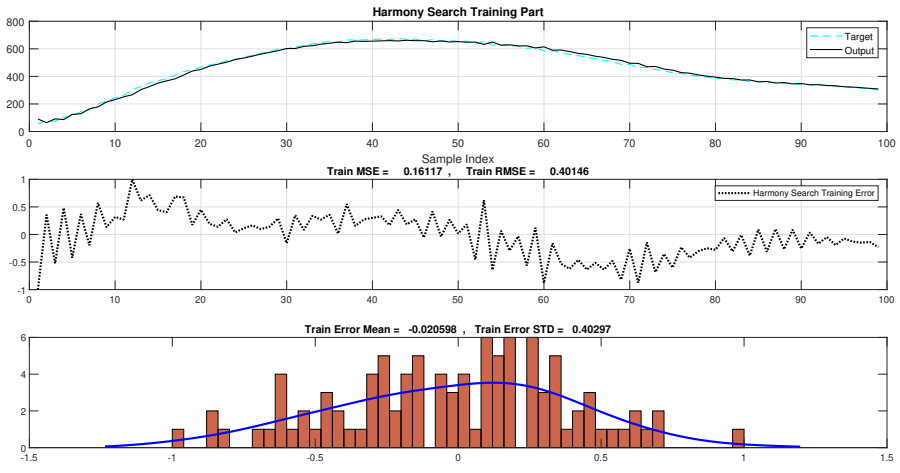

Figure 4.6: MSE and RMSE of Training dataset of COVID-19 cases in Zimbabwe

Chapter 5

Conclusion and Possible Future Work

The COVID-19 pandemic has taught us a lot about the efficacy of different societal responses. To begin, having up-to-date vaccinations, including a recent booster, was found to be particularly useful in defending against newer variations of COVID-19. Hospitalizations were significantly decoupled from cases in countries where a considerable fraction of patients at risk had received three doses of vaccine, including at least one dose of mRNA vaccine. In many countries, the six-month forecast is better than it has been in the previous two years. However, a number of concerns, beginning with the duration of immunity, could dampen enthusiasm. Both natural and vaccine-induced immunity, particularly against viruses, appears to diminish over time. The ultimate goal of this project is to predict the possible rise in COVID-19 cases and in-turn death cases in a particular country or area. The future forecast would help that country be prepared and take precautionary measures to reduce the loss of human life at a maximum possible scale.

In this project, we are successfully able to implement a Neural Network and optimize it using HS, which will be able to predict the future waves. This will alert countries of possible future outbreaks so that they will be fully equipped to tackle the outbreak with sufficient medical equipment and frontline workers. Predicting how and where the next viral outbreak will occur and having a response team will positively impact our society and will lead to less people getting infected and eventual death.

Appendices

Appendix A

Analysis

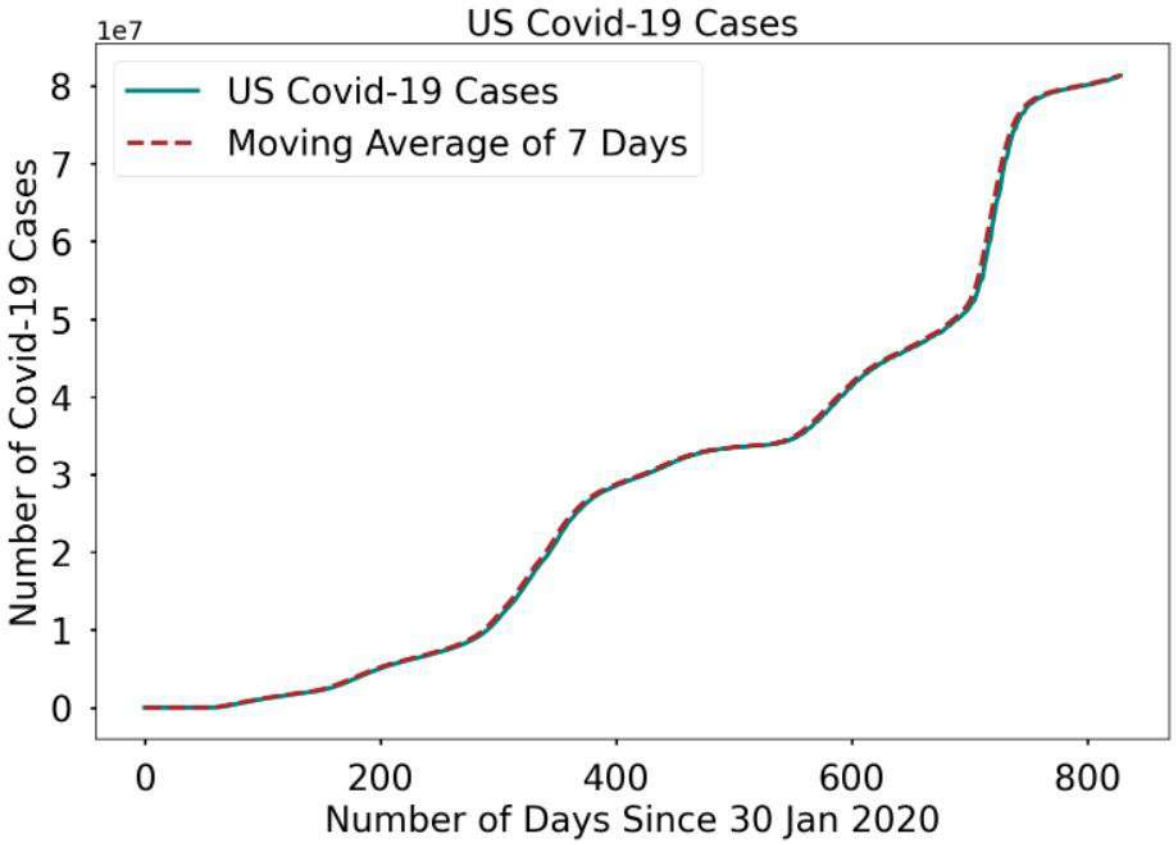

Figure A.1: USA Total COVID-19 cases

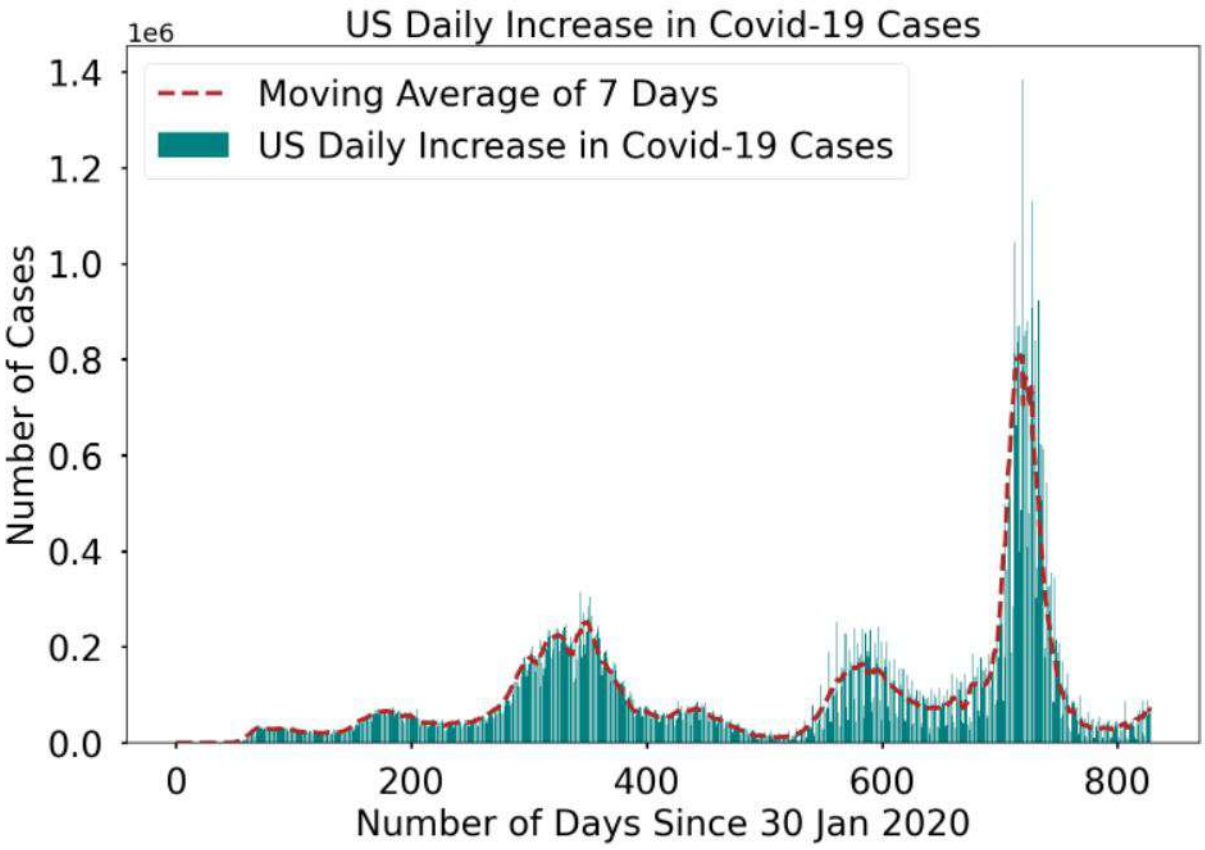

Figure A.2: USA Daily COVID-19 cases

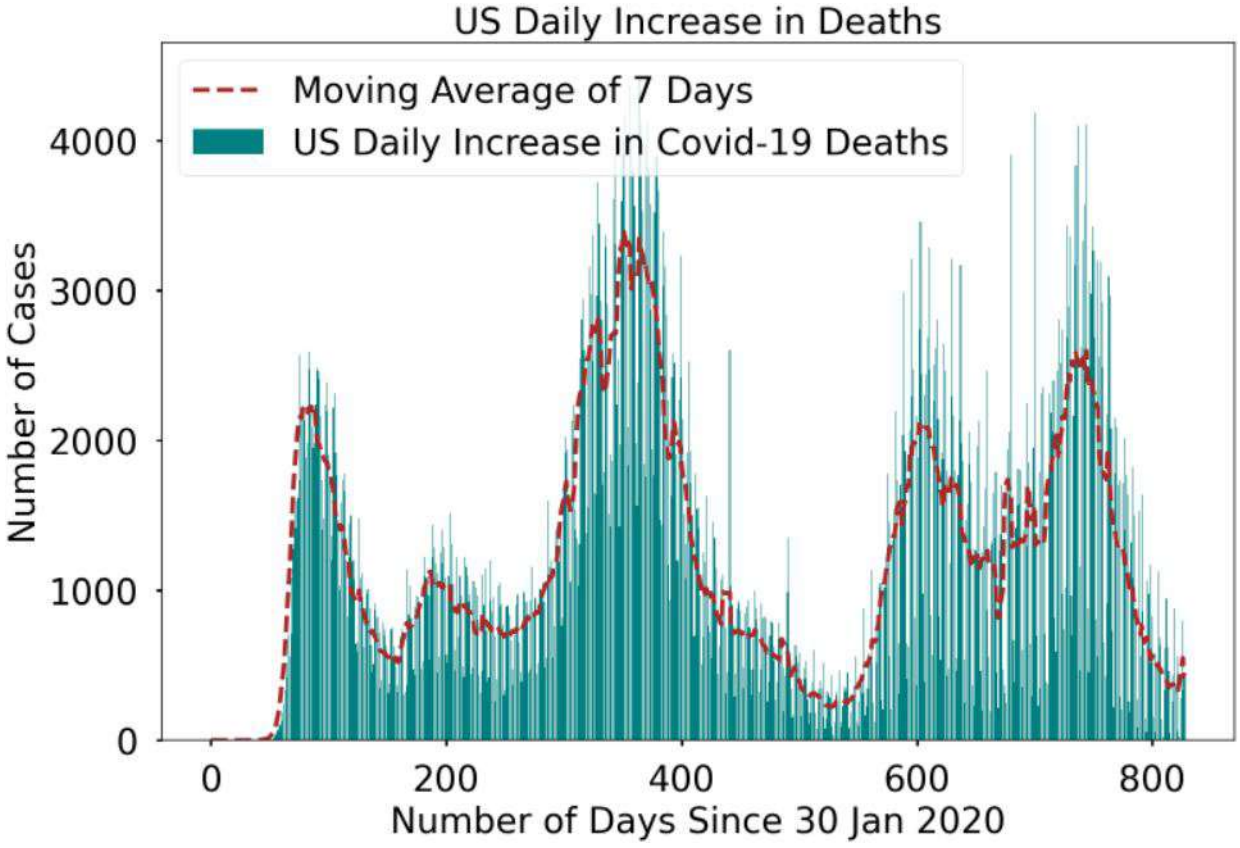

Figure A.3: USA Daily COVID-19 Deaths

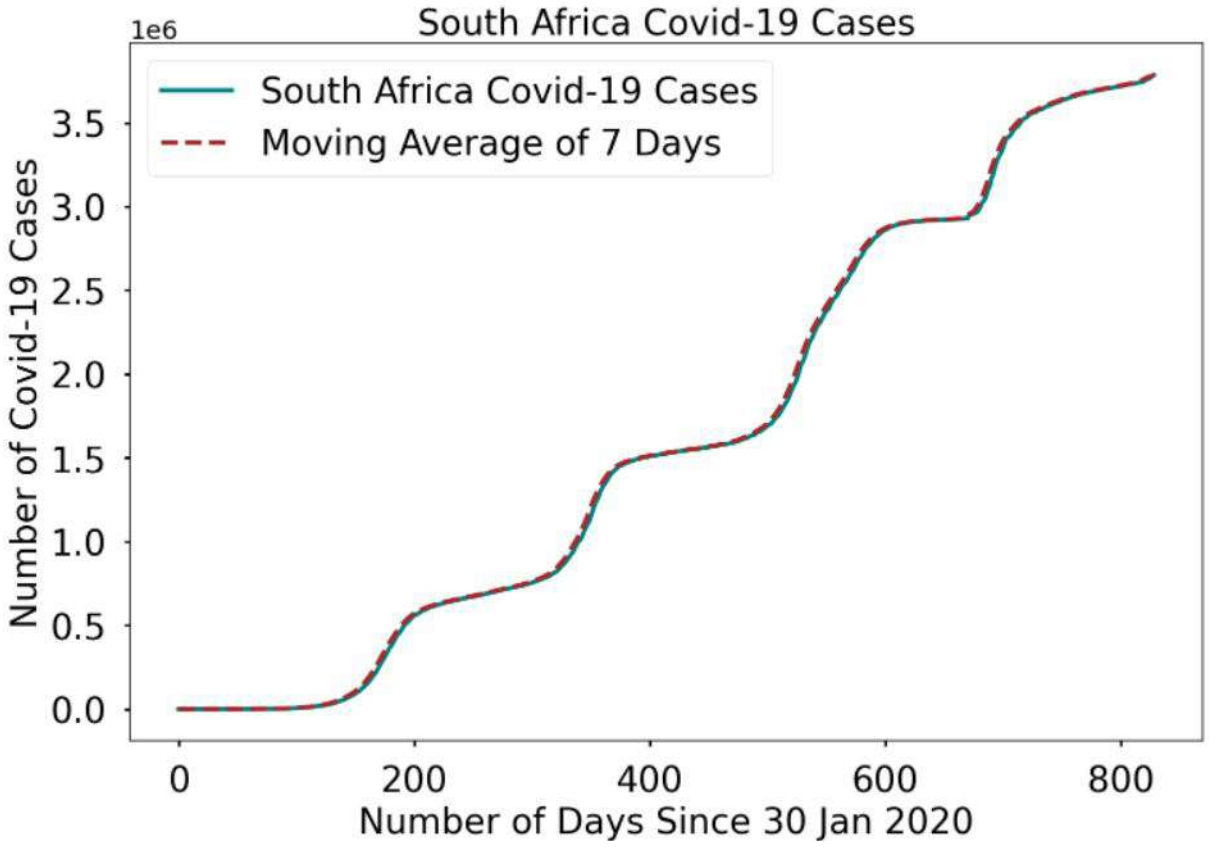

Figure A.4: South Africa Total COVID-19 cases

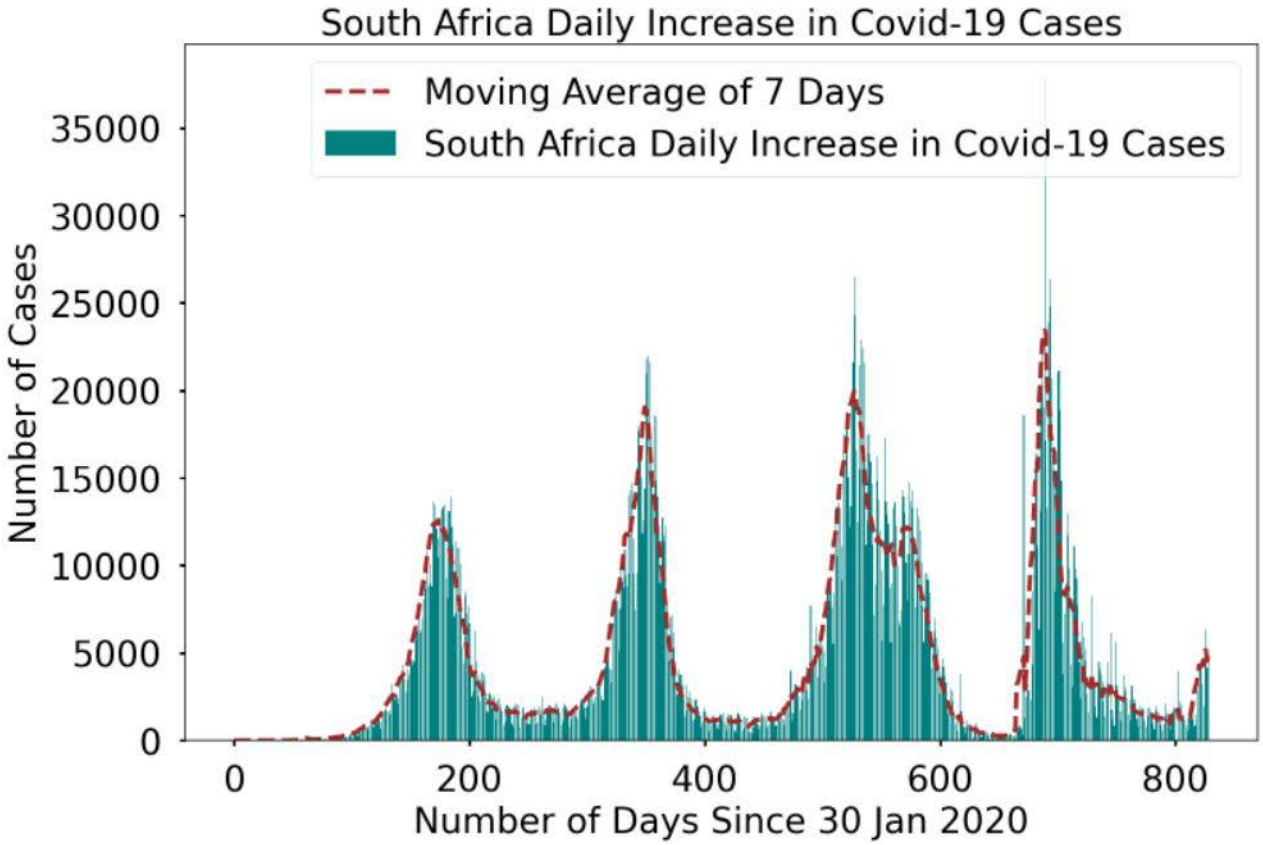

Figure A.5: South Africa Daily COVID-19 cases

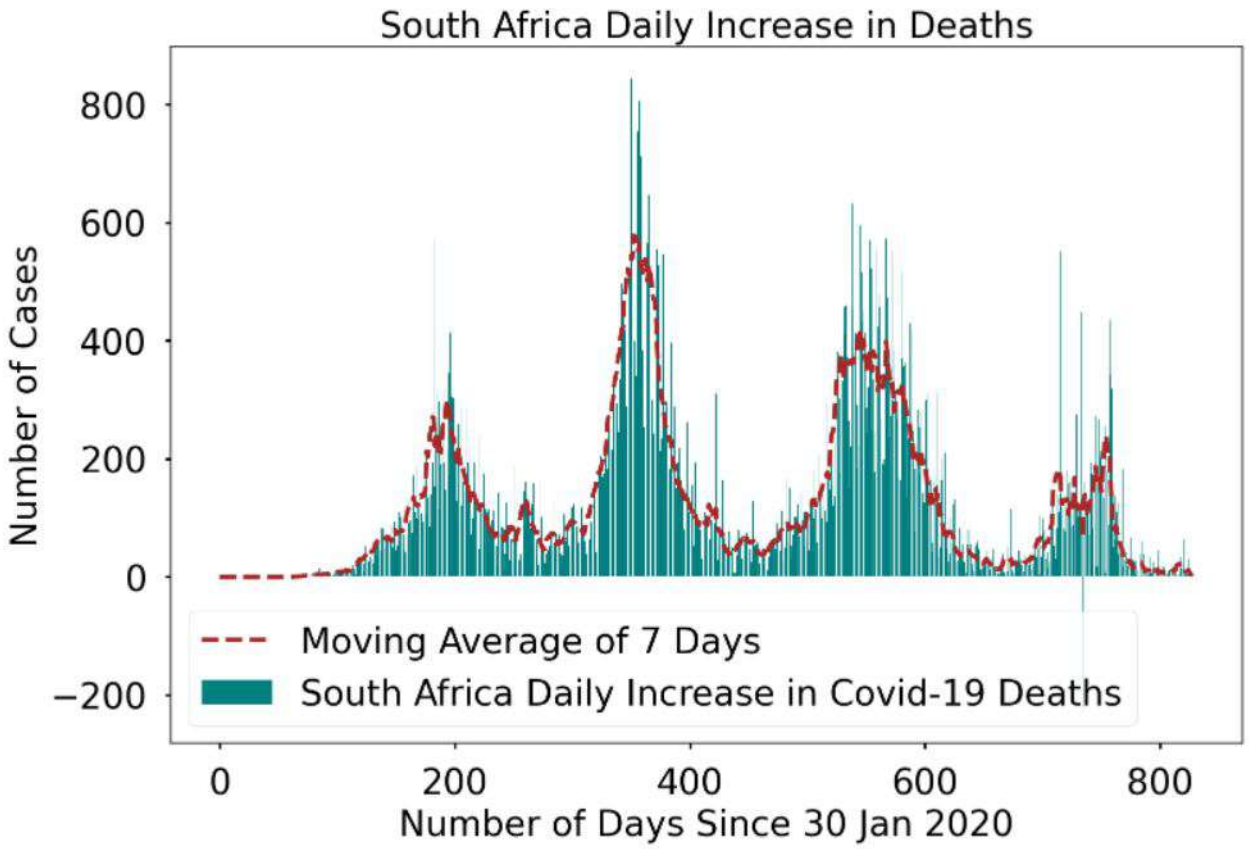

Figure A.6: South Africa Daily COVID-19 Deaths

Appendix B

Harmony Search Time Series Forecasting

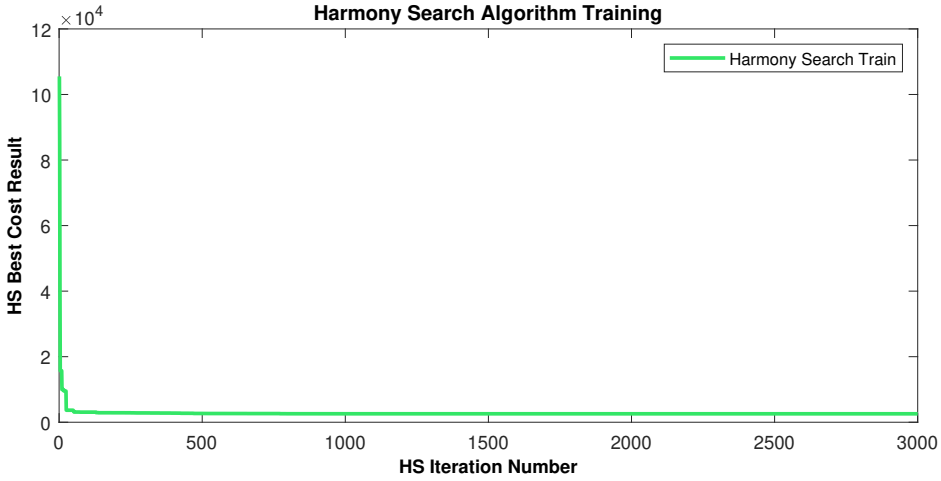

Figure B.1: Harmony Search Best Cost Result for USA

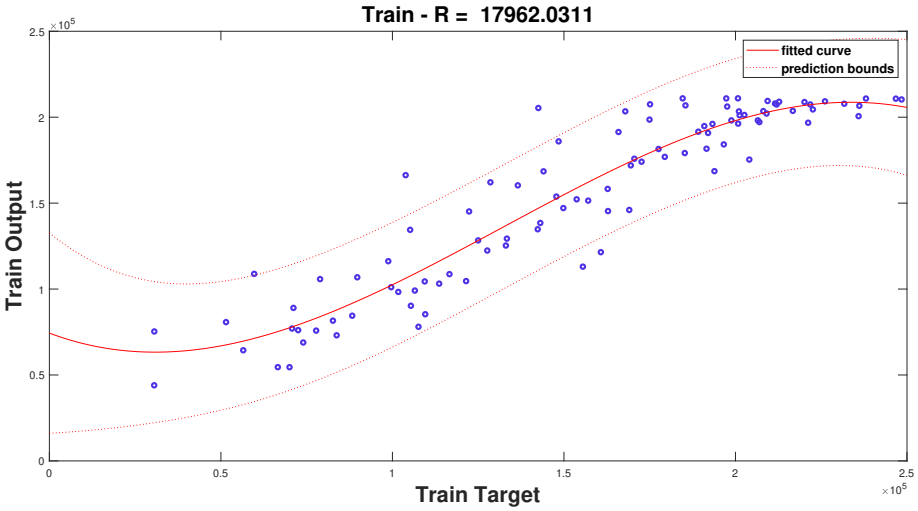

Figure B.2: Prediction Bound of HS training of USA

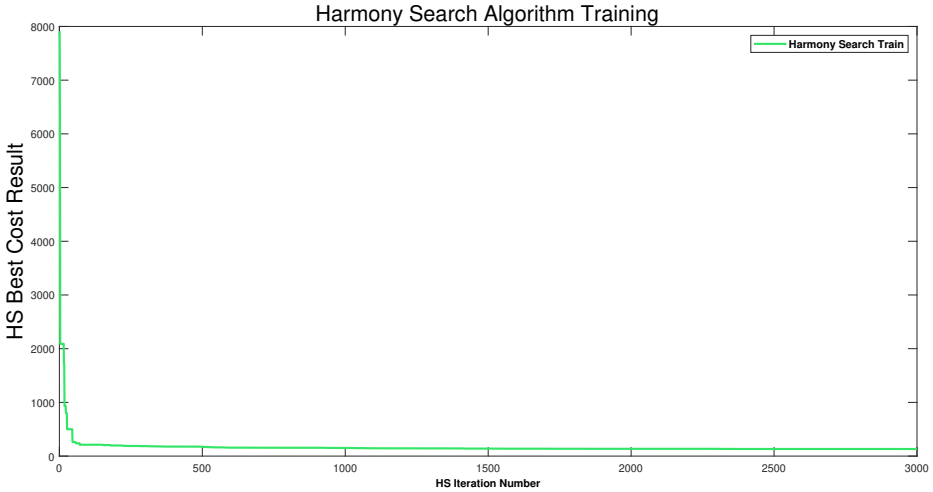

Figure B.3: Harmony Search Best Cost Result for South Africa

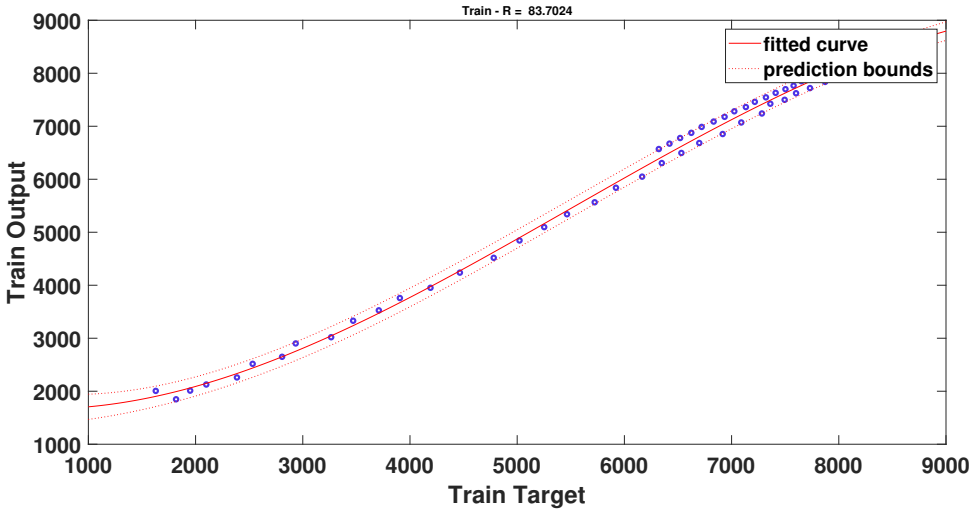

Figure B.4: Prediction Bound of HS training of South Africa

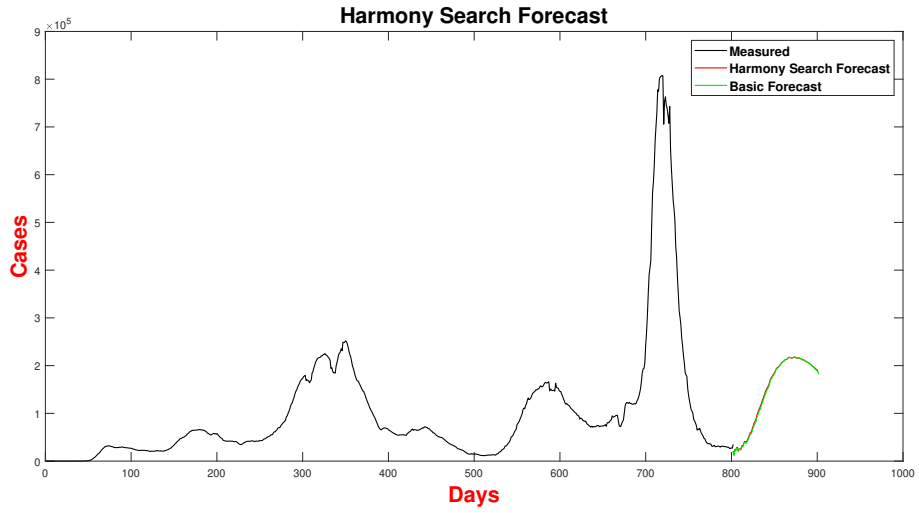

Figure B.5: USA COVID-19 Daily Cases Forecasting

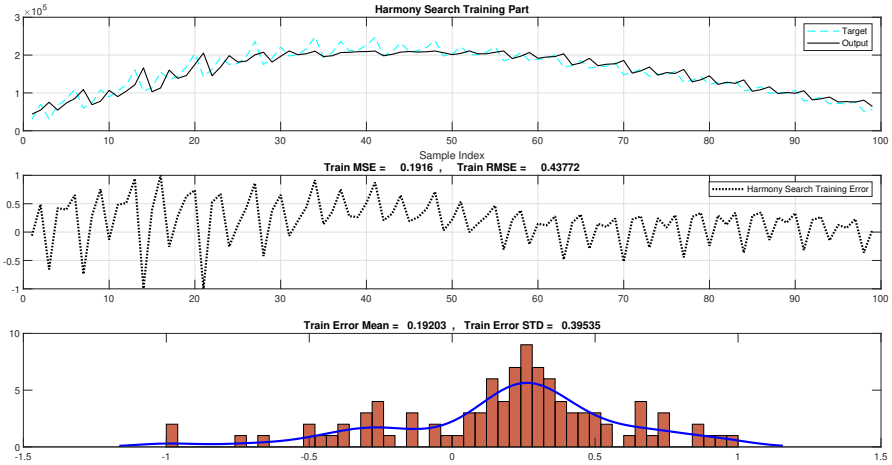

Figure B.6: MSE and RMSE of Training dataset of COVID-19 cases in USA

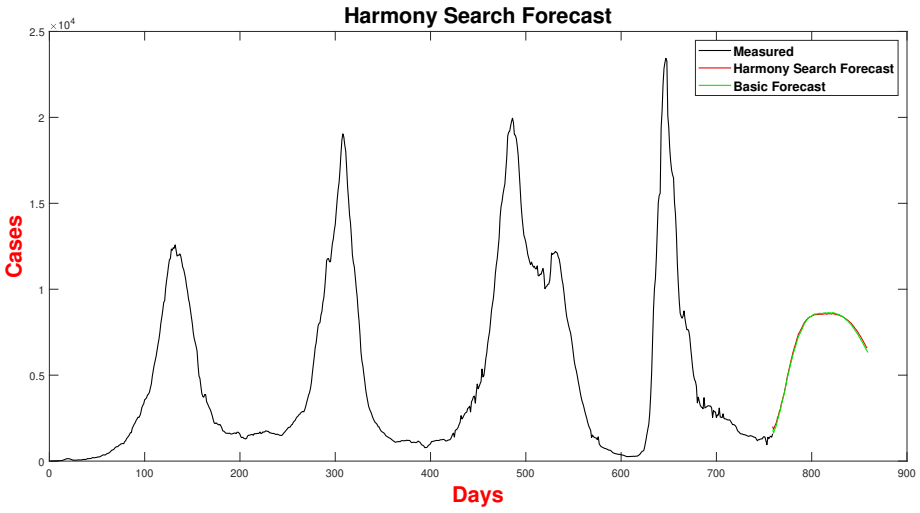

Figure B.7: South Africa COVID-19 Daily Cases Forecasting

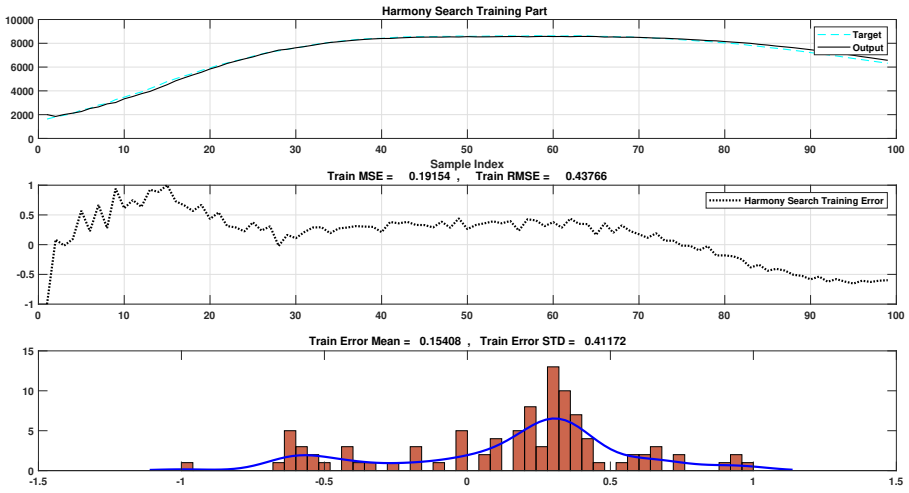

Figure B.8: MSE and RMSE of Training dataset of COVID-19 cases in South Africa

Appendix C

Python Implementation

All the codes for analysis of Covid-19 waves and Prediction through Conventional ML algorithms was done using Google Colab which uses Python 3.0. Various libraries were used to import ML models and Data handling functions. Detailed explanation of the imported libraries is given below:

$Pandas$: Pandas is primarily used to analyse data. Pandas supports data import from a variety of file formats, including comma-separated values, JSON, SQL, and Microsoft Excel. Pandas supports a wide range of data manipulation operations, including mixing, reshaping, and selecting, as well as data cleaning and wrangling.

$NumPy$: NumPy is a Python library that adds support for huge, multi-dimensional matrices and arrays, as well as a large number of high-level arithmetic functions to work on these arrays.

$Matplotlib$: Matplotlib is a data visualization and graphical plotting library for Python and its numerical extension NumPy that runs on many platforms.

$Sklearn$: Classification, regression, clustering, and dimensionality reduction are just a few of the useful capabilities in the sklearn toolkit for machine learning and statistical modelling. It is employed in the creation of machine learning models.

Appendix D

MATLAB Implementation

All the codes for forecast of Harmony Search have been implemented in MATLAB 2022A. Machine Learning Toolbox was used to implement few inbuilt functions which are given as follows:
$xlsread$ - Read Microsoft Excel spreadsheet file that contain data of various countries.
$nlarx$ - Estimates a nonlinear model to fit the given estimation data using the specified ARX model orders and the default wavelet network output function.
$linspace$ - Generates n points. The spacing between the points is $(x2 - x1)/(n - 1)$. linspace is similar to the colon operator, :, but gives direct control over the number of points and always includes the endpoints. lin in the name linspace refers to generating linearly spaced values as opposed to the sibling function logspace, which generates logarithmically spaced values.
$forecastOptions$('InitialCondition','z') - Create an option set for forecast using zero initial conditions.

References

[1] F. Rustam et al., ”COVID-19 Future Forecasting Using Supervised Machine Learning Models,” in IEEE Access, vol. 8, pp. 101489-101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

[2] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and machine learning forecasting methods: Concerns and ways forward,” PLoS ONE, vol. 13, no. 3, Mar. 2018, Art. no. e0194889.

[3] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, “Machine learning strategies for time series forecasting,” in Proc. Eur. Bus. Intell. Summer School. Berlin, Germany: Springer, 2012, pp. 677.

[4] WHO. Naming the Coronavirus Disease (Covid-19) and the Virus That Causes it. Accessed: Apr. 1, 2020. [Online]. Available:https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technic%alguidance/naming-the-coronavirus-disease-(covid-2019)-andthe-virusthat-cau%ses-it

[5] F. Petropoulos and S. Makridakis, “Forecasting the novel coronavirus COVID-19,” PLoS ONE, vol. 15, no. 3, Mar. 2020, Art. no. e0231236.

[6] G. Grasselli, A. Pesenti, and M. Cecconi, “Critical care utilization for the COVID-19 outbreak in Lombardy, italy: Early experience and forecast during an emergency response,” JAMA, vol. 323, no. 16, p. 1545, Apr. 2020.

[7] ’https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv

[8] https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv

[9] https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_reports/04-09-2022.csv

[10] Mojjada RK, Yadav A, Prabhu AV, Natarajan Y. Machine Learning Models for covid-19 future forecasting [published online ahead of print, 2020 Dec 9]. Mater Today Proc. 2020;10.1016/j.matpr.2020.10.962. doi:10.1016/j.matpr.2020.10.962

[11] Mahima Dubey, Vijay Kumar, Manjit Kaur, Thanh-Phong Dao, "A Systematic Review on Harmony Search Algorithm: Theory, Literature, and Applications", Mathematical Problems in Engineering, vol. 2021, Article ID 5594267, 22 pages, 2021. https://doi.org/10.1155/2021/5594267

[12] J. Zhang and P. Zhang, "A study on harmony search algorithm and applications," 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 736-739, doi: 10.1109/CCDC.2018.8407228.

[13] C. Willmott and K. Matsuura, "Advantages of the mean absolute error(MAE) over the root mean square error (RMSE) in assessing average model performance," Climate Res., vol. 30, no. 1, pp. 7982, 2005.

[14] R. Kaundal, A. S. Kapoor, and G. P. Raghava, "Machine learning techniques in disease forecasting: A case study on rice blast prediction," BMC Bioinf., vol. 7, no. 1, p. 485, 2006.

[15] S. Baran and D. Nemoda, "Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting,"Environmetrics, vol. 27, no. 5, pp. 280292, Aug. 2016.

Curriculum Vitae

Name:	Biswajit Nanda
Father's Name:	Jaya Krushna Nanda
Date of Birth:	01-12-1999
Nationality:	Indian
Sex:	Male
Company Placed:	Adobe India Private Limited, Bangalore
Permanent Address:	Flat A-301 Prateeti Apartment,165A Prantik Pally,Kolkata-700042
Phone number:	9051334400
Mobile:	9051334400
E-mail ID:	biswajit.nanda1@gmail.com

CGPA:8.73

Examinations Taken:

1. GATE 2021, GATE 2022

Placement Details:Adobe India Private Limited, Bangalore

Curriculum Vitae

Name:	Manas Dixit
Father's Name:	M K Dixit
Date of Birth:	12-06-2002
Nationality:	Indian
Sex:	Male
Company Placed:	WIPRO-
Permanent Address:	H 128 Hindalco Colony, Renukoot,Sonbhadra ,(U.P.)
Phone number:	NIL
Mobile:	8948920856
E-mail ID:	manas.dixit2018@vitstudent.ac.in

CGPA:7.9

Examinations Taken: NIL

Placement Details: WIPRO , Banglore

Curriculum Vitae

Name:	Somdyuti Das Adhikary
Father's Name:	Sankar Das Adhikary
Date of Birth:	22-12-1999
Nationality:	Indian
Sex:	Male
Company Placed:	Accenture India Private Limited, Bangalore
Permanent Address:	Krishna 3/4, Mohan Garden, Kamalgazi, Kolkata-700103, P.O. Narendrapur
Phone number:	7980873470
Mobile:	7980873470
E-mail ID:	somdyuti.newton159@gmail.com

CGPA: 8.80

Examinations Taken:

1. CAT 2021

Placement Details: Accenture India Private Limited, Bangalore

Capstone Project

Project title:	COVID-19 Analysis and Forecasting using Harmony Search Optimization
Team Members:	Biswajit Nanda, Manas Dixit, Somdyuti Das Adhikary
Faculty Guide:	Dr. M.N. Venkataraman
Semester/Year:	8th Semester/ 4th Year
Project Abstract:(Not more than 200 words)	In this Capstone project, the aim was to design and implement a model solution which would help the larger society tackle the COVID-19 pandemic and other virus outbreaks in the future. Firstly, we collected the numerous datasets of COVID-19 cases, deaths, recoveries and vaccinations available in the internet. Detailed analysis of the datasets were done in Python using Jupyter Notebook. For the prediction of future COVID-19 cases and waves, we used a Deep Fuzzy Neural Network model with metaheuristic Harmony Search optimization technique to forecast the daily COVID-19 cases and deaths in the future from 100 upto 200 days. The countries on which we performed the predictions are India, USA, South Africa and Zimbabwe.
List **codes** and **standards** that significantly affect your project (Must)	1.COVID19 Prediction using SVM, Linear Regression and Bayesian Ridge.ipynb 2. Harmony_Search_Time_Series.m, 3. HarmonySearchFCN.m, 4. GettingFuzzyParameters.m, 5.GenerateFuzzy.m, 6. FuzzyParameters.m, 7. FuzzyCost.m, 8. MakeTheTimeSeries.m

List at least two significant realistic design constraints that are applied to your project.(Must)	In the prediction of COVID-19 cases and deaths, we used a metaheuristic algorithm namely, Harmony Search Algorithm. A constraint of heuristic algorithms is that it is a trial and error method. It is a approximate method based on common sense and educated guesses. Therefore, there is no need to prove the correctness of the algorithm. It is a problem specific approach and does not give a general purpose solution. On the advantage side, metaheuristic algorithms work on both differential and non differential functions and also multimodal functions. It finds the global minima and is not restricted to finding only local minima.
Briefly explain two significant trade-offs considered in your design,including options considered and the solution chosen (Must)	In Harmony Search Algorithm, one factor where there was no compromisation done was regarding the optimization of the Cost Function. To find the optimal solution by minimizing the cost function, we maximised the number of iterations. The tradeoff was that the compilation and runtime of the algorithm was huge.
Describe the computing aspects, if any, of your project. Specifically identifying hardware-software trade-offs, interfaces, and/or interactions	The project is a fully software based project and no hardware were used in its implementation. The softwares that were used for the implementation of this project are - Jupyter Notebook for the analysis of COVID-19 of different countries. Harmony Search time series forecasting was done in MATLAB version 2022a. The Harmony search algorithm was run for 3000 iterations to find the optimal solution by reducing the cost function.