



**NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF MEDICAL GENETICS
Ph.D. PROGRAM IN MEDICAL BIOLOGY
AND GENETICS**

**ARTIFICIAL INTELLIGENCE AND ARIMA BASED
MODELING OF DAILY AND CUMULATIVE COVID-19
CONFIRMED CASES IN AFRICA**

PhD. THESIS

Zurki IBRAHIM

Nicosia

February, 2023

ZURKI IBRAHIM

**ARTIFICIAL INTELLIGENCE AND ARIMA BASED MODELLING OF
DAILY AND COMMULATIVE COVID-19 CONFIRMED CASES IN AFRICA**

PHD THESIS 2023

**NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF MEDICAL GENETICS
Ph.D. PROGRAM IN MEDICAL BIOLOGY
AND GENETICS**

**ARTIFICIAL INTELLIGENCE AND ARIMA BASED
MODELING OF DAILY AND CUMULATIVE COVID-19
CONFIRMED CASES IN AFRICA**

PhD. THESIS








Zurki IBRAHIM

**Supervisors Prof. Dr. Pinar TULAY
Assist. Prof. Dr. Jazuli ABDALLAHI**

**Nicosia
February, 2023**

Approval

Thesis defence was held online. The Jury members declared their acceptance verbally which is recorded

Examining Committee	Name-Surname	Signature
Head of the Committee	Assoc. Prof. Dr. Nahat Rızaner	
Committee Member	Assoc. Prof. Dr. Mahmut Cerkez Ergoren	
Committee Member	Assoc. Prof. Dr. Buket Baddal	
Committee Member	Assoc. Prof. Dr. Ayşe Seyer	
Committee Member	Assoc. Prof. Dr. Emrah Güler	
Co-Supervisor Member	Assoc. Prof. Jazuli Abdullahi	
Supervisor	Prof. Dr. Pinar Tulay	

Approved by the Head of the Department

09/02/2023


Prof. Dr. Pinar Tulay
Head of Department

Approved by the Institute of Graduate Studies

...../...../2023.....

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute



Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Zurki Ibrahim

08/02/2023

Acknowledgments

In the name of Allah, the most merciful and the most beneficial. Praise be to Allah (SWT), lord of the world. Thanks, and Glory to the Almighty Allah for his endless blessings throughout my life. I will never relent by saying ‘Alhamdulillah’ for giving me this opportunity to complete my non-stop academic journey up-to the PhD level successfully. My PhD journey history can never be complete without mentioning the hardworking supervisor Prof. Dr. Pinar Tulay, who always strive hard to see that the journey was successful through her daily mentorship, enthusiasm, patience, assistance and motivation. I will equally like to thank my tutors that gave me their best during the PhD course work. I will not relent to thank my PhD. Jury members for their profound support and recommendations in order to enhance the quality of my research.

At this point, I will be glad to say a special prayer to my late father Alhaji Ibrahim Aliyu for doing his best to see that I make it in life, may your gentle soul keep resting in peace and may Allah make Jannah as your final abode. My special thanks go to my mom for doing her best to see that I make it, I think she deserve more credit than any person as far my academic career is concern through her encouragement and more notably her prayer. My special thanks also go to my family members and friends their kind support throughout this journey. I will always be grateful to my research teams as well as my academic network more especially; Dr. Jazuli Ibn Abdallah who mentored and collaborate with me to develop different artificial intelligence-based articles with high impact factor that can be beneficial to the academic circle. My special thanks also go to my academic peers and colleagues whom have always been source of inspiration to me through their emotional and social support. It’s equally a pleasure to show my appreciation to the management of Sule Lamido University, Kafin Hausa, as well as tertiary education trust fund (TETFund) for providing me with the scholarship.

Zurki Ibrahim

Abstract

Artificial Intelligence and Arima Based Modeling of Daily and Cumulative Covid-19 Confirmed Cases In Africa

Ibrahim Zurki

PhD, Department of Medical Genetics

PhD Program in Medical Biology and Genetics

Supervisors: Prof. Dr. Pinar TULAY and Assist. Prof. Dr. Jazuli ABDALLAHI

February 2023, 116 pages

Background: The global pandemic caused by COVID-19 has led to an appalling effect on the health of health. This work aims to show the performance of four ARIMA based models in COVID-19 cases forecast, also work was done to modeled the COVID-19 cases across African region using artificial intelligence-based models. **Methodology:** Phase one; Philips-perron and Augmented Dickey-Fuller were applied to carry out the unit root tests, Partial Auto correlation Function, Residual diagnostics and Schwarz Information Criterion were considered in selecting the fittest Autoregressive moving average model for cumulative COVID-19 incidences in West African nations. Phase two (methodology); AI based models because of accurate and nonlinearity estimate abilities were applied, comprising of ANN, ANFIS, SVM as well as conventional MLR models. Consequently, novel ensemble approaches including ANN-E and SVM-E were employed. **Results:** Phase one; It is shown that ARIMAML-ARIMAGLS provides better forecast accuracy than AUTOARIMA. Phase two (results); AI based outcomes, displayed that ANFIS has enhanced prediction capabilities followed by ANN for most of the countries in the stage of validation. The ensemble modeling outcomes has illustrated an elevated performance of the used models regarding prediction of the deadly virus across African regions with 0.0073, 0.0002, 0.0155 and 0.9616 for MAD, MSE, RMSE and R2 respectively. The ANN ensemble increased the individual model's abilities at the stage of validation up to 10, 14, 42, 6, 83, 10, 7, 5, 7, and 31 percent for Morocco, Republic of Sudan, Namibia, South Africa, Republic of Uganda, Rwanda, Republic of Nigeria, Senegal, Gabon as well as Cameroon correspondingly.

Conclusions: These findings can hand out as an allusion for modeling COVID-19 cases as well as determining hospitalization requirements.

Key words: COVID-19, ensemble, artificial intelligence, ARIMA, pandemic.

Table of Contents

Approval.....	i
Declaration	iv
Acknowledgments.....	v
Abstract	vi
Table of Contents	vii
List of Tables.....	ix
List of Figures	x
List of Abbreviations.....	xii

CHAPTER I

Introduction.....	1
1.1 Spread of the CCC across the West African states	3
1.2 Scope of the Study	4
1.3 Aim of the Study.....	4
1.4 Specific Objectives.....	4
1.5 Study Limitations	5

CHAPTER II

Literature Review.....	6
2.1 Structure of coronaviruses and their Classification	6
2.1.1 Genomic Categorization of SARS-CoV-2.....	6
2.1.2 Contagious Features of Spike Proteins of SARS-CoV-2.....	7
2.1.3 Attachment Motif within the S-protein of the SARS-CoV-2.....	7
2.1.4 Insertions Effect within the Sequence of S Protein on the Infectivity of SAR-CoV-2.....	8
2.2 Pathology and Pathogenesis of the SAR-CoV-2.....	9
2.3 History and Basis of the novel virus infection (SARS-CoV-2)	9
2.4 Transmission routes	11
2.5 Susceptibility of the Population	12
2.6 Mutations and Variants classification	12
2.6.1 Variant of Interest	13
2.6.2 Variants of Concern	13
2.6.3 Coronaviruses Variation of Elevated Consequence.....	14
2.7 Coronavirus Disease 2019 Variants	14
2.7.1 COVID-19 Variants in Africa.....	16
2.8 ANN Algorithm	17

2.9 ANFIS	17
2.10 Support vector machine (SVM)	17
2.11 multi-Linear regression	18
CHAPTER III	
Methodology	19
3.1 Phase one: Data and methodology for the Cumulative Cases using ARIMA	19
3.1.1 Variable definition	19
3.1.2 Model of the ARIMA.....	19
3.1.3 Evaluation criteria.....	21
3.2 Phase II: Data and Methodology for the Daily Cases using AI	22
3.2.1 Study area and data	22
3.2.2 Model Validation	25
3.2.3 Data normalization and performance criteria.....	27
3.2.4 Ensemble approaches (Modeling).....	28
CHAPTER IV	
Findings and Discussion	33
4.1 Phase I: Results and discussions for the Cumulative Cases Using ARIMA	33
4.2 Phase II: Results and Discussions for the Machine Learning Based Modeling	50
4.2.1 Results of the separate models:	50
4.2.2 Ensemble Approaches Results	59
4.2.3 Results of the Cumulative Confirmed Incidences:.....	65
CHAPTER V	
Discussion	72
CHAPTER VI	
Conclusion and Recommendations	77
6.1 Conclusion	77
6.2 Recommendations	78
REFERENCES.....	79
APPENDICIES	91
Turnitin Similarity Report.....	91
CURRICULUM VITAE	102

List of Tables

Table 1: Variables Descriptive statistics	20
Table 2. Statistical description of the study nations.....	24
Table 3. Cumulative cases, data partitioning and validation	26
Table 4: ADP and PP unit root test outcome	34
Table 5: Training sample Assessment outcome.....	40
Table 6: Testing sample Assessment outcome	43
Table 7. The Input variables employed.....	51
Table 8. Outcomes of the single models for Northern African states.....	52
Table 9. Findings of the employed models for Eastern African countries....	52
Table 10. Outcomes of the employed models for Western African countries	53
Table 11. Outcomes of the employed models for the Southern African countries	53
Table 12. Outcomes of the employed models for C. African countries.....	53
Table 13. Outcomes of the employed novel ensemble approach (ANN-E).	60
Table 14. SVM-E outcomes.....	60
Table 15: Outcomes of the engaged models in Northern Africa	65
Table 16: Outcomes of the engaged models in Eastern Africa.....	66
Table 17: Outcomes of the engaged models in West Africa.....	67
Table 18: Results of the applied models for South Africa	69
Table 19: Outcomes of the engaged models in Central Africa.....	70

List of Figures

Figure 1: Distribution of the CCC across the West African countries.....	4
Figure 2: ARIMA Models Algorithm	21
Figure 3. Map of the study area	23
Figure 4. Schemes for the time sequence of the daily-affirmed incidences..	25
Figure 5. K-fold validation employed in the finding	26
Figure 6. Main methodology used for the ML.....	28
Figure 7. The overall procedure of novel ensemble approach	29
Figure 8. Planned ANN-E procedure	31
Figure 9: The planned SVM-E procedure applied	32
Figure 10: Line graph for the CCC of West African countries.....	33
Figure 11: Autocorrelation Function of the primary variation of Cumulative COVID-19 Case for ECOWAS Countries	38
Figure 12: PACF of the initial variation of Cumulative incidences for West African nations	39
Figure 13: Forecast comparison graph for the training sample.....	46
Figure 14: Forecast comparison graph for the testing sample	47
Figure 15: Taylor illustration for the training sample.....	48
Figure 16: Taylor diagram for the testing sample.....	48
Figure 17: Fan plot for the training sample.....	49
Figure 18: Fan scheme of the RMSE for the testing sample.....	50
Figure 19. Comparison of individual model performance on R2 (1) ANN model, (2) ANFIS model, (3) SVM model (4) MLR model.....	56
Figure 20: Statistical performance of developed models versus observed values for (a) Northern African nations, (b) Southern African nations, (c) Eastern African nations, (d) Central African nations (e) Western African states	58
Figure 21. Matching between projected daily-affirmed cases of COVID-19 and observed incidences within the stage of validation for (a) broad set of the data (b) Zoom view	59
Figure 22. Matching of the models performance for (a) Northern African states, (b) Eastern African countries, (c) Western African countries, (d)	

South African countries and (e) Central Africa states.....	64
Figure 23: Time series plot for Northern African states.....	65
Figure 24: Bar chart illustrating the best performing models for North Africa	65
Figure 25: Time series for the cumulative incidences in East Africa	66
Figure 26: Bar chart illustrating the best performing models in East Africa	67
Figure 27: Time series plots for the cumulative incidences in West Africa .	68
Figure 28: Bar chart demonstrating the best performing models in West Africa	68
Figure 29: Time series plot for the cumulative incidences in South Africa..	69
Figure 30: Bar chart illustrating the best performing models in Southern Africa.....	69
Figure 31: Time series plot of the cumulative incidences for Central African countries..	70
Figure 32: Bar chart illustrating the best performing models in Central African nations.	71

List of Abbreviations

SARS:	Severe acute respiratory syndrome
MERS:	Middle East respiratory syndrome
COVID-19:	Coronavirus disease 2019
AI:	Artificial Intelligence
DNA:	Deoxyribonucleic acid
RNA:	Ribonucleic acid
ACE2:	Angiotensin converting enzyme 2
ANN:	Artificial neural network
SVM:	Support vector machine
MLR:	Multi linear regression
ANFIS:	Adaptive neuro-fuzzy inference system
ARIMA:	Auto-regressive integrated moving average
MAD:	Mean absolute deviation
WHO:	World health organization
ML:	Machine learning
SMAPE:	Symmetric mean absolute percentage error
CDC:	Center for disease control
FFNN:	Feed forward neural network
LM:	Levenberg Marquardt
LR:	Linear regression
MAPE:	Mean Absolute Percentage Error
LSSVM:	Least square-support vector machine
ARIMAML:	ARIMA with maximum likelihood
ARIMAGLS:	ARIMA with generalized least squares
SARS-CoV-2:	severe acute respiratory syndrome coronavirus 2
ML:	Mali
DC:	Daily cases
SAE:	Simple average ensemble
TMPRSS2:	Transmembrane serine protease 2
CCC:	Cumulative Covid-19 Cases

MSE:	Mean square error
RMSE:	Root mean square error
CC:	Correlation co-efficient
GLS:	Generalized Least Square
TIC:	Theil Inequality Coefficient
NG:	Nigeria
GM:	Gambia
GH:	Ghana
BJ:	Benin
R²:	Determination co-efficient value
BF:	Burkina Faso
CV:	Cape Verde
GN:	Guinea
GW:	Guinea Bissau
LR:	Liberia

CHAPTER I

Introduction

Coronavirus disease 2019 global outbreak was brought on by the new SARS-CoV-2. The unique disease was firstly recorded in Wuhan, Republic of China, in the month of December 2019, and it is highly contagious and was first identified in bats before spreading to dogs and raccoons (Morens et al., 2020). MERS, SARS, the most current SARS-CoV-2, and other Coronaviruses are the ones that led to cause various forms of diseases in birds as well as mammals, including cattle enteritis, pigs, chickens, and other mammals (Mahase, 2020; Islam et al. 2020). The Chinese CDTP had noted that after the discovery of the novel virus, the following methods could be used to identify Coronavirus disease 2019 in the presence of a negative result from the SARS-CoV-2 acid tests: a positive chest CT scan; b severe medical symptoms such as breath shortness (Roosa et al., 2020).

Around the world, the pandemic has impacted over 219 nations (Muhammad et al., 2021). The unique virus, which has been classified as an emergency health issue of global concern, spreads by direct and close contact with bodily fluids of the patients carrying the deadly virus (WHO). Additionally, asymptomatic incidents and inadequate diagnostic tools lead to late diagnoses, rendering visitors, healthcare officers and patients to the pathogenic. In the past twenty years, there were other coronaviruses that put the globe in peril (Zivkovic, et al., 2021). SARS virus outbreak in 2003 was followed by MERS pandemic in the year 2012. In addition to the recent Zika virus, there have been other disease outbreaks around the world during the past 20 years, including those caused by Swine flu, Ebola, H1N1, and others. Due to the pandemic, sophisticated and original models with great prediction precision were developed (Burki, 2020).

Coronavirus disease 2019 outbreak, however, has similarly shown a cluster of changes equated with the earlier viral occurrences, raising questions regarding the models' real ability to make precise forecasts and predictions (Ardabili et al., 2020). The newest virus plague still possessed a number of unidentified factors which are persuading the transmission of the novel virus, such as the diverse population's attitudes and complexity across different countries and territories, the various approaches taken by policy makers during adopting preventive methods to stop the blowout of the deadly virus (Ivanov, 2020).

These ambiguous indexes had significantly decreased the projection capability of the current models (Scarpino and Petri, 2019). The question if social isolation is sufficient to avert the infection was examined (Mirza et al., 2022). In an effort to diagnose and curtail the COVID-19 outbreak, ML, data mining technique, expert systems, as well as the rest of AI must significantly be applied, while offering the most accurate diagnostic and dependable procedures for COVID-19, the use of strategies which are not therapeutic possessed the possibility to lessen the enormous health care systems burdens (Lucas, Vahedi & Karimzadeh, 2022).

One of the most cutting-edge artificial intelligence (AI) concepts is machine learning (ML), which offers a strategic method for creating automated, intricate as well as algorithmic objective procedures for the mathematical data and multi-model processing (Sajda, 2006). Machine learning's algorithms have the capacity to modify its structures depending on the established identified information together with carried out adaptation through enhancing over a specific objective (Jabra et al., 2021). Models for the machine learning have already displayed forecasting and prediction abilities in numerous filed of research (Nourani et al., 2020) like, statistical downscaling (Ekiran et al., 2021), reference evapotranspiration (Nourani et al., 2020), absorptivity forecasting for hydrocarbon reservoirs (Talebkeikhah et al., 2021), soils suitability in the application of airfield (Sujatha et al., 2021).

Nevertheless, in regards to the outbreak forecast, the Machine Learning have displayed great prediction abilities. Important programs of Machine Learning algorithms for infections epidemic prediction, Oyster norovirus (Chenar & Deng, 2018). With respect to machine learning models applications for coronavirus disease 2019; numerous findings can be found within the literature. Applied hybrid machine learning technique for the projection of coronavirus disease 2019 was carried out in Hungary by Printer et al. (2020). Determining close to or ideal values for a model's parameters is one of the most difficult and problematic aspects of using ML to solve a particular problem. Unfortunately, there isn't a single acknowledged rule; therefore, each case requires a new set of parameter values (Zivkovic et al., 2021). However, both linear and nonlinear components are present in every natural process. The literature findings above are based on research using Machine Learning simulations for COVID-19 projection concentrated on using either

individual Machine Learning models or hybrid simulations, which are not linear procedures methods and ignored the detrimental effects of the system's linear method. As a result, inaccuracies brought on by the linear part of COVID-19 might produce predictions that are erroneous and less effective. Therefore, integrating the linear and nonlinear algorithms using ensemble techniques will improve the complicated nature associated with the COVID-19 and enhance projection. Additionally, each model has weakness and strength; the benefit of ensemble techniques is that they can compensate for a model's weakness with strength of another model, and vice versa. Some of these studies use ARIMA (Guleryuz D, 2021; Rostam-Tabar, 2021 & Alabudulrazak et al., 2012), Seasonal ARIMA (SARIMA) (Arunkumar et al., 2021), MLR model integrated using information from phone call (Chowdhury et al., 2020), Brown's Exponential Smoothing model (Guleryuz et al., 2021; Chowdhury et al., 2020), epidemiological models (Tang, 2020). Also, some studies focus on the policy intervention policy interventions like isolation, total shutdown of towns and physical distancing on coronavirus disease spread (Mati S, 2021).

This is the first study where by ARIMA model is tuned with generalized least squares to predict the CCC. This study also focuses on an important economic block, the ECOWAS. The study uses two-unit root tests to examine the properties of the time series of CCC before modeling, thus avoiding the danger of depending on a single unit root test. ARIMA algorithms evaluate using the ML techniques are compared against the ones estimated with the GLS. Even though Ayinde et al., (2020) employ ARIMA, none of these studies uses the GLS evaluation technique and greater than one unit root test.

1.1 Spread of the CCC across the West African states

The West African Countries include Burkina Faso (BF), Benin Republic, Cabo Verde (CV), Gambia (GM), Cote D'ivoire (CI), Guinea Bissau (GW), Guinea (GN), Ghana (GH), Mali, Liberia (LR), Nigeria (NG), Niger (NE), Togo (TG), Sierra Leone and Senegal. ECOWAS was created and birthed in 1975 with the aim of fostering interstate economic and political cooperation. People from West Africa have been described among the most mobile population in the world, even though much of the migration is said to be intra-regional. About 7.5 million of migrants in West Africa are residing in ECOWAS states rather than their own. Figure one illustrates the plot of the ECOWAS members.

Figure 1 presents the distribution of the CCC across the West African countries. It is clear that only Nigeria and Ghana have their CCC above one hundred thousand (100,000). Niger, Guinea Bissau, Gambia, Liberia and Sierra Leone have recorded less than ten thousand (10,000) cases of COVID-19.

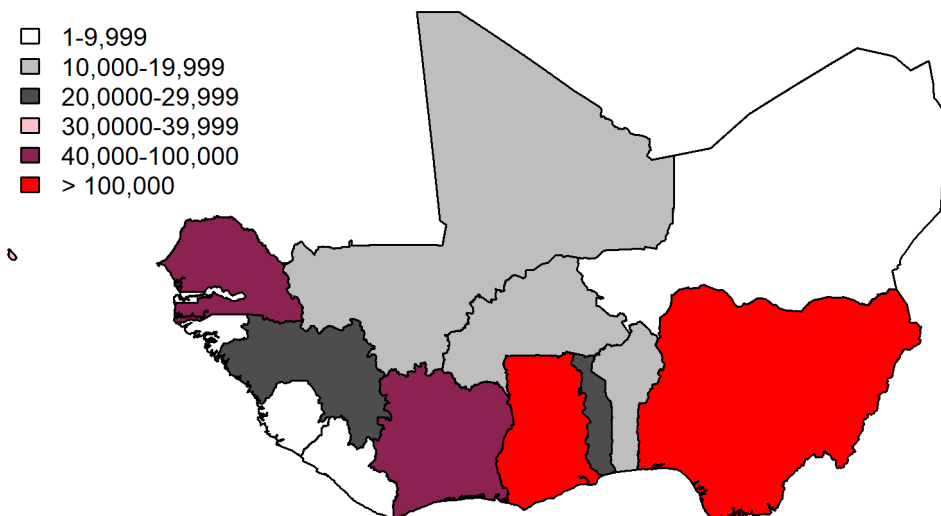


Figure 1: Distribution of the CCC across the West African countries

1.2 Scope of the Study

This study covers the machine learning models applications and ARIMA in forecasting and prediction of coronavirus disease 2019 incidences across African countries. As a result, the driving force behind this study as well as the fundamental research issue is as follows: is to promote performance of the machine learning models for COVID-19 prediction utilizing ensemble approaches.

1.3 Aim of the Study

The core aim of this work is modeling daily and cumulative affirmed incidences of COVID-19 in Africa.

1.4 Specific Objectives

- ✓ To apply the non-linear machine learning models (SVM, ANFIS and ANN) in predicting daily confirmed cases in African nations
- ✓ To apply the conventional MLR model in modeling confirmed cases of COVID-19
- ✓ To use the novel ensemble methods (SVM-E and ANN-E) in order to increase the individual model's predictive accuracy
- ✓ To assess the performance of ANN, SVM, ANFIS and MLR in modeling the cumulative confirmed incidences of COVID-19 in Africa.

- ✓ To apply and assess the capabilities of ARIMA algorithms in forecasting cases of COVID-19 within West African nations.

1.5 Study Limitations

The key limitation in the SVM ensemble as well as ANN ensemble applications is that; in spite of merging both non-linear and linear models, which effectively aided in coopting both the nonlinear and linear compound nature of coronavirus disease 2019, the kernel functions of both SVM-E and ANN-E are nonlinear in nature. Another limitation is the inability to modeled the SARS-CoV-2 variants in Africa due to the poor systematic sequencing surveillance in the continent, and this led to the lack of accurate data regarding the novel SARS-CoV-2 variants.

CHAPTER II

Literature Review

2.1 Structure of coronaviruses and their Classification

The four key subgroups of coronaviruses are α , β , γ , and δ . The coronavirus family has six primary components, comprising of the human infections Cov-HKU1 and CoV-229E. MERS-CoV, CoV-OC43, and SARS-CoV are the three human pathogens that make up the coronavirus group (King et al., 2012).

Multi-shaped or round shape is typically the structure of the coronavirus virion particle. It measures between diameter of 120 to 160nm and comprised of petal- designed ledge containing a threefold spike P, which is an ordinary characteristic of coronavirus (Andrew, Tao & Penghua, 2020). During infection, the binding of the virus to the host's membrane is achieved or mediated by the spike protein. Additionally, the feature of spike P, the genome of coronaviruses commonly encodes 3 more structural protein, comprising of Envelope (E) protein, Membrane (M) protein as well as N-protein (Cody, Michael & Bing, 2022). The protein membrane of Coronavirus, 218-263 amino acids, possessed N-terminus amended by a hydrophilic C-terminal tail and *O*- or *N*-glycan. The 74 to 109 amino acids for Envelope protein, is said to be participated in virulence promotion; there are characteristically about twenty replicas of protein for each virion (Femendez, 2020). The 349-470 amino acids for the coronavirus Nucleocapsid protein, is a ribonucleic acid bound phosphorylated protein which mediates suitable genomic RNA folding to form the nucleocapsid (Cagliani, Forni, Clerica & Sironi, 2020).

2.1.1 Genomic Categorization of SARS-CoV-2

The virion of SARS-CoV-2 possessed about 60-140nm as well as positive sense, 29 891 base pairs single stranded RNA's genome content (Zhou, Zang & Wang, 2020). Genomic sequencing configuration devours shown a seventy nine percent sequence uniqueness amid SARS-CoV-2 and SARS-CoV and a significant 93.1% distinctiveness with RaTG12 viral sequence, secluded from *Rhinolophus affinis* (bat) living in the Area of Yunnan, China; these conclusion outcomes propose that Severe Acute Respiratory Syndrome Coronavirus 2 might be identified originally from the virus which is widespread in the bat types. Comparative genome investigation principally according to insert sequence secluded in the coronavirus spike protein insulated from the pangolin family indicated that

these animals were probably the hosts intermediate for the transmission of cross species (Wu et al., 2020).

Broad examination of the SARS-CoV-2 genetic content collectively with that of SARS-CoV indicated the existence of virtually 30 open reading frames as well as two newest insertions (Cui, Gao & Liu, 2020).

2.1.2 Contagious Features of Spike Proteins of SARS-CoV-2

Related to what was eventually identified for SARS-CoV-2, the attachment of the novel virus spike protein to the cell surface's receptor ACE2 instigate the entry of the virus into the class 2 pneumocytes within the human lungs (Zamorano, 2020). Per se, the spike protein contributes a lot toward the ongoing and initial spread of the newest virus. The spike protein of coronavirus comprises of 2 major areas; N-terminus's S1 domain of the protein intercedes attachment to ACE2 as well as the domain of the C-terminal S2 which accelerates viral fusion with the host cell's cellular membrane (Hofmann, Kleini & Kruger, 2020).

RBD is a sub domain of S1 which consists of 424 to 494 amino acids. The direct contact between motif and the extracellular binding location on angiotensin converting enzyme 2 is regarded as domain of peptidase. Spike protein has 2 cleavage positions, R797 and R667. The site for R667 is at the partition or division amid S2 and S1, the cleavage at the site for R797 lead in the concluding polypeptide. Several proteases (cellular) have the ability to slash the sequence of spike protein 2 locations, consisting of elastase, cathepsin L, trans membrane serine proteases, trypsin as well as factor Xa, to mention a few. Cleavage sites at both of the S protein is vital to accelerate the entrances to the host cell by the newest virus; the firstly it is essential for the attachment of S1 to ACE2 and secondly it is critical for the fusion of membrane.

2.1.3 Attachment Motif within the S-protein of the SARS-CoV-2

The S-protein amino acid of SAR-CoV-2 shares merely incomplete homology with SARS-CoV; the level of resemblance is small inside the S1 domain about 64 percent and is elevated inside S2 area (90%) when compared with the domain of S1 (Firth, 2020). N-terminal section is complete less preserved (51%) inside the S1 domain, while the RBD subdomain C-terminal has moderately high level of conservation (up to 74), thus allowing interactions with the receptor ACE2 of the cell surface (Jaimes, Andre & Millet, 2020). About 4 unique changes exist in sequence of the amino acid inside RBD of S1 of the novel

virus when comparing with that of SARS-CoV-2. F472, X442, N487 and C479 are the amino acids included within the S-protein sequence of SARS-CoV-2 (Zhou, Zang & Wang, 2020).

The variation inside a crucial motif also within the domain of S1 receptor binding could perhaps lead to persuade receptor mediated attachment and eventually the spread capacity of the newest virus (Jaimes, Andre & Millet, 2020). Numerous organizations had since investigated this case. For, instance, Wrapp (2020) established a finding that attachment of amount as small as 15nmol/L within the S1 domain of the novel virus could be identified at the angiotensin converting enzymes 2 by utilizing visual bio sensing through SPR. These findings revealed that the novel virus spike protein has about ten-to-twenty-fold elevated affinity than that of MERS and original SARS-CoV.

2.1.4 Insertions Effect within the Sequence of S Protein on the Infectivity of SARS-CoV-2

The novel virus is an extremely contagious coronavirus; assessment of transmission levels suggests or imply that this could be three to tenfold greater affinity than those of MERS and SARS-CoV, correspondingly (Pustake, 2022). The spread level of the deadly virus is unswervingly linked with spike protein sequence, which comprises one among the sequence insertion identified in the genomic content (Hui et al., 2020). About 4 residue insertion was established within the spike protein straightly nearby the cleavage location (Meng, Cao & Zhang, 2020).

The TMPRSSs has remained recognized as supporting element to MERS-CoV and SARS-CoV contagions respectively; TMRRSS as well as TMRSS11a have the ability to catalyze the spike protein cleavage to both S1 as well as S2 domain in the locations of both R797 and R667 deposits (Walls et al., 2020). The 4 amino acids within the insertion, P681 to P684 in conjunction R685 can develop an uncovered loop; and this lead in the increase of sensitivity to proteases. The cleavage location for the protease, furin, has been generated by the insert sequence. Findings by Meng et al. (2020) established that TMPRSS2 and TMPRSS1 could serve as the cellular proteases stimulating spike protein and afterward aiding to cell entry and attachment of SARS-CoV-2.

Moreover, Jaimes et al. (2020) and Walls et al. (2020) established that TMPRSS2 and TMPRSS1 could be protease stimulating spike protein afterward supporting to cell access and attachment of the novel virus.

2.2 Pathology and Pathogenesis of the SAR-CoV-2

A report of an autopsy from a fifty-year-old patient showed numerous facts of the lungs condition in patients battling from grave forms of coronavirus disease 2019. This patient lost his life as a result of ARDS with characteristics that comprised of pneumocytes peel off, formation of hyaline layer, interstitial swelling as well as inability to filtrate huge quantity of lymphocytes. Moreover, cytopathic-like changes of the virus, consisting of atypical enlarged pneumocytes and multiple nucleated syncytial cells were isolated in the places of alveolar (intra) (Zhou, Yang & Wang, 2020).

The novel virus pathogenesis is mostly not known, but to some extent it mimics SARS. The contagion of the newest virus is cytopathic to the air route of epithelia's cells as well as cells of the alveolar. Though, comparable to what was seen in reaction to SARS-CoV, immune stimulated injury might contribute immensely in COVID-19 pathogenesis, particularly within those who are chronically ill as a result of the disease severity. Virus-related contagion of pneumocytes stimulates local reactions of the inflammation and stimulates cytokines discharge such as TGF- β 1, and many chemokines that hand out to incorporate flowing leukocytes. In harsh nature of coronavirus disease 2019, the resulting inflammatory cascade could result to a cytokine squall, as seen from the latest finding which recognized an elevated serum cytokine degree, consisting of IL-2, IL-7, IL-10, MCP, granulocyte colony mediating factor as well as TNF- α . Both, extra-pulmonary failure of the organ and ARDS are believed to be derived by the cytokine storm (Li, 2016). The exterior lymphoma is usually seen, particularly in relationship along multiple chronic form of coronavirus disease. This reflection or statement could mirror efficient characterization as a result of the clear engagement of these cells into lung tissue infected by the virus somewhat than whichever particular virus stimulated suppression (Zhou et al., 2020). There is visibly an amplified amount of the activated HLADR+CD38+T cells within peripheral blood stream, in spite of reduction in the total number.

2.3 History and Basis of the novel virus infection (SARS-CoV-2)

Bases of SARS-CoV-2 infection are animals and human as well, possibly, early hosts of the novel virus are bats, whereas pangolins might serve as the transitional host (Ye et al.,

2020). Similarly, asymptomatic as well as symptomatic persons with the disease are recognized to be transmittable. Nevertheless, its non-apparent how extensive viral shedding perseveres and, in that way, its spread nature could be changed during the usual disease origination. The China's health agency has experimented samples from the environment as well as animal samples from the Seafood market Huanan city, China. The outcomes of the analysis showed that about 94% of SARS-CoV-2 nucleic specimens (acid-positive) (31/33 incidences) originated directly from the western section of the local seafood market, that consists of services which offer undomesticated animals for acquisition as well. The usual hosts of lots of identified coronaviruses are bats (Wit et al., 2016). As illustrated earlier, the novel virus is a β coronavirus; similarities of sequence involving β coronaviruses and SARS-CoV-2 isolated from species of bat can be as elevated as 89.0 percent (Zhou et al., 2020) to about ninety six percent, which signified that COVID-19 virus could be originated from ancestor coronavirus prevalent in the bat species.

Comparatively, findings showed that resemblance amid the genomes sequence of SARS-CoV and that of the novel virus reached about 79.5%, especially 73.8 to 74.9percent at the 2 receptor domains, all of them are said to share angiotensin converting enzyme 2 as a conjoint receptor (Zhou et al., 2020). Fascinatingly, the deadly viral genome as well as that of the bat species coronaviruses also varies, with around one thousand one hundred base pairs variance of nucleotide (Zhou et al., 2020). Additionally, it's vital to actually understand that the main plague appeared during the period of wintertime when the species of bats use to undergo hibernation. Intrinsically, the statistics recommend that there might be single or multiple transitional hosts that connect the coronaviruses of bat species to humans.

The research conducted by Guo et al. (2020), presented profound learning's procedures, revealed that minks' animals could be the possible transitional host, nevertheless no investigational proof was presented for the backing of this assumption. The pangolin specie is presently regarded to be the most probable within the intermediate host's candidate. A team of researchers from a University of Agriculture, China has recognized a strain of coronavirus from the specie of pangolins that possessed about 99% resemblance with that of the novel virus (Kong, 2020).

The present prominent hypothesis or assumption is that the virus derived from bat evolved to pass on a disease to pangolins species; after a chain of mutations as well as events of recombination, it has been spread or transmitted to the human population. Though, the outcomes of this study are yet to be published. Temporarily, other findings have acknowledged and recognized a strain of coronavirus from the Malaysian pangolin specie with sum total of genetic makeup (genome) that possessed a similarity with that of the newest virus at 85.5 percent to 92.4 percent (Lan, Ge & Yu, 2020). Notably, the Malaysian pangolin's GD/P2S as well as GD/P1L coronaviruses are tightly connected with COVID-19 virus. These outcomes showed that the pangolins might be the enduring host of mentioned viruses. Moreover, more than four thousand eight hundred poultry as well as livestock samples has been tested by the Chinese center for animal health and epidemiology, and revealed the probability that the newest virus might have its roots in livestock or poultry (Lau et al., 2020)

2.4 Transmission routes

The key means of the novel virus transmission among people are respiratory droplets as a result of close contact, Other possible routes consist of fecal-oral and aerosol transmissions, though it is yet been affirmed (Huang et al., 2020). As have seen in the rest of respiratory viral diseases, droplets from respiration are believed to be the major way of transmission. When susceptible persons interact with body fluids containing virus (saliva, sputum, feces) from animals or humans, the newest virus could pass on via the oral and nasal's cavity as well as other peels of the mucous. Similarly, when predisposed persons contacted with items contaminated with fluids of the body, indirect spread of the deadly virus can also happen (Huang & Hermann, 2020).

Droplets encompassing pathogens are regarded as 'biological aerosol' that have been present within atmosphere in a specific amount of duration where its moisture is lost. The viruses and proteins that remains developed into droplet hub and have the ability of the traveling a distance with the current of air could result in the disease spread over long distance (Xiao, 2020).

During particular clinical involvements (non-invasive breathing, mask ventilation as well as tracheal canulation), patients with chronic SARS-CoV-2 sickness could shack the novel pathogen to an increased level, increasing the risk to the rest of the environment (Wax & Christian, 2020).

It was disclosed that the likelihood of mother to her child virus spread when, on February 6, 2020, in Wuhan Tongji Hospital, antenatal lady infected with COVID-19 delivered a baby carrying the disease thirty-six hours after giving birth. On February 8, 2020 in Zhejiang, China, a pregnant lady with chronic coronavirus disease 2019 gave birth to a baby who after laboratory diagnosis confirmed not be carrying the deadly infection on many subsequent nucleic acid virus examinations (Zheng et al., 2020).

Nevertheless, a new discovery revealed that an infection or contagion that affects the fetus may spread throughout the third trimester (Chen et al., 2020). These remarks might be related to cells found at the mother-fetal interface having low levels of ACE2. Through known methods of vertical or plumb transfer, there generally seems to be a little menace of fetal disease. Some experts have speculated that the novel virus might pass on via conjunctiva, though a latest finding invalidated this chance. In separate research by Zhou et al., (2020) revealed that only one patient had conjunctivitis out of the sixty-seven individuals with the newest virus; the viral nucleic acid diagnosis of discharges from conjunctiva's sac was negative.

2.5 Susceptibility of the Population

The general population is susceptible with no barrier of a given age or sex. Among all known individuals infected with the disease, those that are above the age of fifty accounted for about 53.6% of the documented cases and children less than ten years constitute for only 0.9%. Patients with multifactorial disorders such as cancer, hypertension, and the rest are more likely to undergo and succumb to severe stage of the infection and also possessed an elevated risk of forming difficulties (Lau et al., 2020). Members of a family belonging to COVID-19 patients and medical practitioners care are at high threat for infection as a result of more repeated contact and interaction with infected individuals. Healthcare providers accounted for about 29% of the total coronavirus disease 2019 patients registered to the Zhongnan's University Teaching Health center, China (Wang et al., 2020).

2.6 Mutations and Variants classification

A genetic sequence change is referred to as a mutation. Variants are considered to be genomes which vary from one another in the genetic sequence. Meanwhile, the variants are referred to as strains if a phenotypic variation or difference is discovered within them (Hoffmann, Kleine & Kruger, 2021). Genetic sequence of the coronavirus 2 linked with

severe acute respiratory syndrome was made available on GISAID in the month of January 2020. The next strain of the virus is the result of collaboration between researchers in Basel, Switzerland, and Seattle, USA, and it curates and studies the genetics of the innovative virus. The United States government interagency has classified SARS-CoV-2 mutations into three classifications (Chan, Kok & Zhu, 2020).

2.6.1 Variant of Interest

This particular division comprised of strains with definite indicators that are linked to variations to receptor attachment, neutralization of antibody decrease formed against earlier disease or vaccination, decreased of treatment efficacy, impact of potential diagnostic or forecast increase in the spread of severity of the disease, and so forth, to verify the disease's transmissibility as well as its severity, reinfection risk, and immunity to vaccination, this particular group also needs improved sequencing surveillance, laboratory classification, and epidemiological study (Roy, Dhillon, Habib & Pugazhadi, 2021).

B.1.525, B.1.526, and P.2 are the current versions that make up the variants that the United States of America listed. All of the aforementioned variations share the same mutation (D614G), and data suggests that these variants spread infection more quickly than versions without the mutation (Hodcroft, 2021).

The strain that was presently flowing was replaced with the D614G mutation-containing strain between early January and early February 2020. At this point in the spike protein, the 614 codon changes the amino acid from aspartic to glycine (Groves, Rowland & Angyal, 2020).

2.6.2 Variants of Concern

This includes those that established elevated infection spread ability, further severity of the disease including deaths and hospital admissions, significant decrease in neutralization of antibody, reduced effectiveness and efficacy of treatments as well as failure in diagnosis (Mallano, Ascione & Flego, 2022). This category of variants also needs increased interventions to curtail spread of the novel variants through development of testing kits, putting more efforts to determine the treatment and vaccine efficacy and effectiveness against the new variant (World health organization [WHO], 2021). The USA has currently listed the following variants: B.1.1.7, P.1, B.1.351, B.1.427, and B.1.429 this class of

variations share the D614G mutation with the variant of interest group, which spreads faster than variants with no changes in the nucleotide sequence (Bhattacharya et al., 2021).

2.6.3 Coronaviruses Variation of Elevated Consequence

This group was made up of variants that had evidence indicating the preventive methods were far less effective and efficient than previously used variations. There are no variations known for this group (World health organization [WHO], 2021).

2.7 Coronavirus Disease 2019 Variants

B.1.1.7, B.1.351, and P.1.B.1.1.7, also known as VOC202012/01, were found in the United Kingdom in the month of September 2020 (Center for disease control [CDC], 2021). The novel variant has about twenty-three genetic variations when matched to the original variant discovered in the city of Wuhan, China. Different variants of COVID-19 have been recorded in various countries, including B.1.1.7, B.1.351, and P.1.B.1.1.7. It was discovered that the spike protein included about eight of these mutations. The deletion of 69/70 and the mutations P681H and N501Y are the most notable ones. The N501Y mutation seems to enable the spike protein to form a strong bond with the ACE receptor. It is 40 to 80 percent more infectious (Mallano, Ascione & Flego, 2020).

Almost five thousand out of seventeen thousand four hundred and fifty-two coronavirus disease 2019 deaths all through the months of September to February were as a result of this new variant. Furthermore, it has been assessed that the deaths were nearly 55% greater when matched to other variants. Equally, the clinical reports in the month of January, 2021 showed that there existed amplified frequency of passing away due to this newest variant (WHO). There were twelve thousand five hundred and five cases across 51 jurisdictions, as of April 1, 2021 (Center for disease control [CDC]), 2021), it has also been recognized in 82 countries and territories.

It is proposed to have amplified spread level and often observed in younger individuals with no any primary ailment (CDC, USA). This variant mutation E484K facilitates the antibody leakage is the key purpose for the decreased vaccines sensitivity (Vasireddy et al., 2021). The messenger RNA vaccines (Moderna and Pfizer) were certified in the USA prior to the isolation of the novel strain in the nation (Jan, Venkataraman, Wechsler & Peppas, 2021). According to the newest findings, these 2 vaccines stimulated lesser neutralizing antibodies compared with the previous strains. Janssen, Novavax as well as

Astra-Zeneca carried out experimentally in S. Africa that partake leading B.1.351 altered strains (Mwenda et al., 2020).

P.1 novel strain also regarded as B.1.1.28.1 was initially discovered in the Northern Brazil city (Manaus) in the month of December 2020. It was also isolated outside the city of Tokyo in January 2021, during a routine screening when travelers from Brazil landed in the city's airport. There were two hundred and twenty-four reported incidences as of April 1, 2021 in twenty-two jurisdictions in the USA.

The variant in the US Midwest also called 20C-US or COH.20G/501 was identified in the city of Ohio accompanied by additional variant in the month of December 2020 and month of January 2021 respectively (Tada et al., 2021), The variant possessed a mutation on the membrane protein (A85S0), N protein (D377Y) and spike protein (Q677H). An additional variant has been discovered with the with nucleotide sequence change S N501Y, a genetic indicator of the B.1.1.7, and also without any related change in the nucleotide sequence with that novel strain, presently there is not any proof of an elevated virulence or transmissibility for this newest variant (Tada et al., 2021).

CAL.20C variant was detected in Southern California in July 2020, and isolated again within populace samples of the identical territory in the month of October 2020 (Zhang et al., 2021), the main important changes in the nucleotide sequence of it are ORF1b: D1183Y, ORF1a: D1183Y, W152C, S: S13I, as well as L452R. The attachment of the spike protein might be completed with ease by the last 3 nucleotide sequence changes B.1.525 (20A/S: 484K) as well as B.1.526 (20A/S: 484K variants were initially discovered in the city of New York, United State of America. S477N and E484K are the famous mutations within this variant. S477 elevated the binding process while E484K decreases the response of the antibody.

B.1.617 known as double mutant variant was firstly existed and detected in India (Pascarella et al., 2021), in this newest variant 2 changes in the nucleotide sequence are observed (in the same virus). The first case in the US that involve this double variant was detected in the city of Francisco on the first week of April 2021.this variant has the capacity to be transmitted very quickly and it resist vaccination. The most notable mutations are L452R and E484Q. Bharat Biotech's COVAXIN vaccine is the most appropriate and effective vaccine against this virus (Pascarella et al., 2021).

The novel triple variant was identified in late April 2021 in addition to L425R and E484Q; it is categorized by the presence of deletion in 2 amino acids, Y145del and H146del within spike protein (Alem, Akbar & Slenker, 2022). One thousand one hundred and eighty-nine samples were tested positive as of April 21 2021 in the city of Maharashtra, India. Just like other variants, mutant triple variant has greater transmissibility.

Another variant known as 20A.EU1 possess non-terminal (NTD) changes in the nucleotides sequence which did not contribute a direct part in the attachment to the receptor of the membrane (Harvey, Carabelli & Jackson, 2021), this newest variant was originally isolated on June, 2020 in the city of Madrid, Spain but quickly spread all over European and other countries.

Meanwhile, 20A.EU2 strain was recognized in the month of June, 2020 in France and has turn out to be the second most prevailing novel variant in European countries, the remarkable mutations in the variant includes E484K, S477N as well as N501Y, which established slight rise in the angiotensin converting enzyme 2 binding, convalescent sera and resistance to numerous antibodies (Hodcroft, 2021).

The most recent discovery is that there is another unusual variety termed N440K with a mutation in the spike protein, which is what caused the dramatic increase in cases among the people of Andhra Pradesh, India. According to the Cellular and Molecular Biology Center, this variation is ten to one thousand times more infectious, has better affinity to angiotensin converting enzyme 2 receptors, and is resistant to the class three monoclonal antibodies C135 and REGN10987. Numerous documented cases of COVID-19 reinfection with anti-severe acute respiratory syndrome coronavirus 2 antibodies present suggest a possible decline in the antibodies' ability to neutralize the vaccine-induced pathogens (Rothe et al., 2021).

2.7.1 COVID-19 Variants in Africa

The variant B.1.351, likewise referred as 501Y.V2 was originally isolated in the first week of October 2020 in a South African location known as Nelson Mandela Bay (Tang, Toovey, Harvey & Hui, 2021), the novel variant was also identified in Zambia in December 2020. There are Twenty-three mutations together with amino acids variations but the remarkable nucleotide sequence changes within the variant are E484K, K417N, as well as N501 situated in the spike protein. It is hard to distinguish the sum total figures of variants circulating within African nations because there is no efficient sequencing

surveillance. As of 22nd January 2021, there were only five thousand visibly existing sequences from twenty four of the forty-seven nations (World health organization [WHO], 2021).

2.8 ANN Algorithm

ANN main structure contains hidden, output and input layers correspondingly: the three respective layers usually handles the operation input layer, hidden layer as well as output layer; typically handle the operations for training data (Ekhmaj, 2012). Input and output layers encapsulated neurons normally linked via input and output vectors. In the meantime, neurons contained within the hidden layer were connected to neurons within the three layers, and they essentially helped change the input information into the corresponding output information. Furthermore, a transfer function was used to transfer the subjective sum of the input figures (Abdullahi, Ekiran & Nourani, 2017). Artificial neural networks typically permit links between the neurons encased in all the layers and the layers above and below it but disallow links between levels. Last but not least, desirable outcomes are usually achieved if the ANN is well trained before the data flow through the network stops and an association with the essential precision is made (Nourani & Fard, 2016).

2.9 ANFIS

The literature on the system of fuzzy inference or NN references several learning algorithms for fuzzy modeling through the use of neuro-fuzzy simulation (Akrami et al., 2014). Adaptive Neuro Fuzzy Inference System was developed by Jang in 1997 and utilizing the learning's simulation of neural network, is a novel approach to the creation of neuro-fuzzy. The algorithm can compress the bunch of proficiency to any stage for every factual constant function as an overall approximator.

ANFIS is functionally equivalent to a FIS (Jang, 1997). In this case, the ANFIS's interest is similar to that of the primary order Sugeno fuzzy algorithm.

2.10 Support vector machine (SVM)

In 1998, Vapnik put up the idea of a support vector machine. It utilizes non-linear mapping to a greater dimensional pit depending on the planned minimization instruction, that includes regression, regularization, complexity of the model, and kernel function. Numerous findings highlighted the realization and success of support vector machines in

anticipating various outcomes. In regards to the choice of parameter, support vector machine do not have direction theoretically. It is complicated since the support vector is calculated using quadratic-based programming (Nagana et al., 2019).

Compared to quadratic simulations, it has higher algorithmic complexity and demands a lot of memory (Kocadagli et al., 2022). Additionally, choosing the right kernel combination is crucial for improved model performance. However, picking the right kernel function might be challenging.

2.11 multi-Linear regression

Commonly, regarding multi linear regression, the regressor (n) variables as well as the reliant on variable y may be connected by:

$$y=b_0+b_1x_1+b_2x_2+b_3x_3+\dots+b_ix_i+\zeta \dots\dots\dots(12)$$

b_0 stand as the regression continuous, x_i stand as the symbol of the i th predictor, and b_i stand as the i th forecaster coefficient, similarly ζ is the error term.

CHAPTER III

Methodology

3.1 Phase one: Data and methodology for the Cumulative Cases using ARIMA

The set of figures (data) in this work were extracted in the month December 2020 from the dashboard of W.H.O. The size of the sample for each nation is illustrated in table one.

3.1.1 Variable definition

The variable utilized in this investigation has the following definition. The cumulative sum of each day's test results for COVID-19 patients is referred to as the CCC.

3.1.2 Model of the ARIMA

ARIMA refers to the statistical based model analysis which utilized data of time series to forecast future trends or to better understand the set of data. A statistical based model is regarded as autoregressive if it forecast values of the future based on the previous values. The Box, G., and Jenkins team proposed the ARIMA model Equation is a representation of the ARIMA (p, d, q) model (1).

- (a) Auto-regressive: is regarded to a model that display a changing variable which regresses on prior or its own lagged and values
- (b) Integrated: demonstrate the differencing of raw observations to permit the series of time to become stationary (where values of the data are usually changed by the variation between the information values and the past values)
- (c) The moving average: co-opt the reliance amid the reflection and a remaining error from a moving average archetypal employed to lagged observation.

Therefore, the terms "autoregressive order," "order of integration," and "order of moving average" are denoted by the letters "p," "d," and "q," correspondingly.

$$\Psi Y_t = \alpha_0 + \Gamma \varepsilon_t \quad (1)$$

ARIMA with Maximum Likelihood:

ML estimation is a technique that examine values of the model parameters. The values of the parameters usually exist in such a way that expand the likelihood where approach described by the model resulting in the data set that were genuinely observed.

Auto-ARIMA:

A time series library called Auto ARIMA automates the creation of ARIMA models. In modeling and forecasting, Auto ARIMA uses the ARIMA ideas. Auto ARIMA automatically finds the best parameters of an ARIMA model (Hyndman et al., 2008).

Unit Root Tests:

(a) Augmented Dickey Fuller:

Finding the figure of differencing required in order to achieve the stationarity of the series which is the first stage in ARIMA time series forecasting. When performing time series analysis, it must be proficient in the ADF test.

The ADF test is primarily a statistical significance test, which is another important thing to keep in mind. This indicates that a null and alternate hypothesis are used in the hypothesis test, and as a result, a test statistic is produced and p-values are presented.

(b) Philips Pheron:

PP unit root test is non-parametric, that is, it doesn't need to incorporate the degree of correlation serially as well.

Table one demonstrates statistics description of the CCC for the West African countries. The number of observations (N) represents each country sample size. The larger the observation number of indicates that the country records the case of coronavirus disease 2019 earlier. Therefore, the initial COVID-19 case in West African states was documented in Nigeria, then Senegal, Togo and so on. The mean (\bar{Y}) represents the average number of the CCC for each country, the median (Y_{med}) provides the value of the CCC at the middle of the sample set for each and every nation, while standard deviation (σ) measures variability of the cumulative incidences from the mean value for every nation within the study area. The minimum (Y_{min}) give the first number of cases recorded for each country. For example, Burkina Faso's first record of affirmed case of the novel virus is 6, whereas Niger's is one. (Y_{max}) stands as the maximum value and it show the sum total or cumulative incidences in the early month of September, 2021

Table 1: Variables Descriptive statistics

	\bar{Y}	Y_{med}	Y_{max}	Y_{min}	σ	N
BF	6,521.37	3,123.50	13,777	6	5,602.20	540
BJ	4,132.20	3,081.50	13,366	1	3,234.60	534
CI	25,829.73	21,441.00	55,669	5	17,151.41	539
CV	13,265.08	10,626.00	35,283	1	11,750.78	531
GH	56,485.34	52,274.00	119,436	5	34,905.33	537
GM	3,633.67	3,776.00	9,698	2	2,634.99	533
GN	13,473.15	13,233.00	29,400	4	8,188.29	537

	\bar{Y}	Y_{med}	Y_{max}	Y_{min}	σ	N
GW	2,631.43	2,444.00	5,799	2	1,264.65	525
LR	1,867.63	1,676.00	5,594	3	1,462.34	534
ML	6,926.06	5,721.00	14,874	2	5,185.64	525
NE	2,938.35	1,856.00	5,857	2	2,053.94	532
NG	90,012.80	67,371.00	191,805	5	65,646.34	552
SL	2,967.82	2,451.00	6,367	1	1,735.67	519
SN	23,630.40	16,089.00	72,805	5	19,163.29	549
TG	5,787.13	3,014.00	21,261	5	5,931.69	544

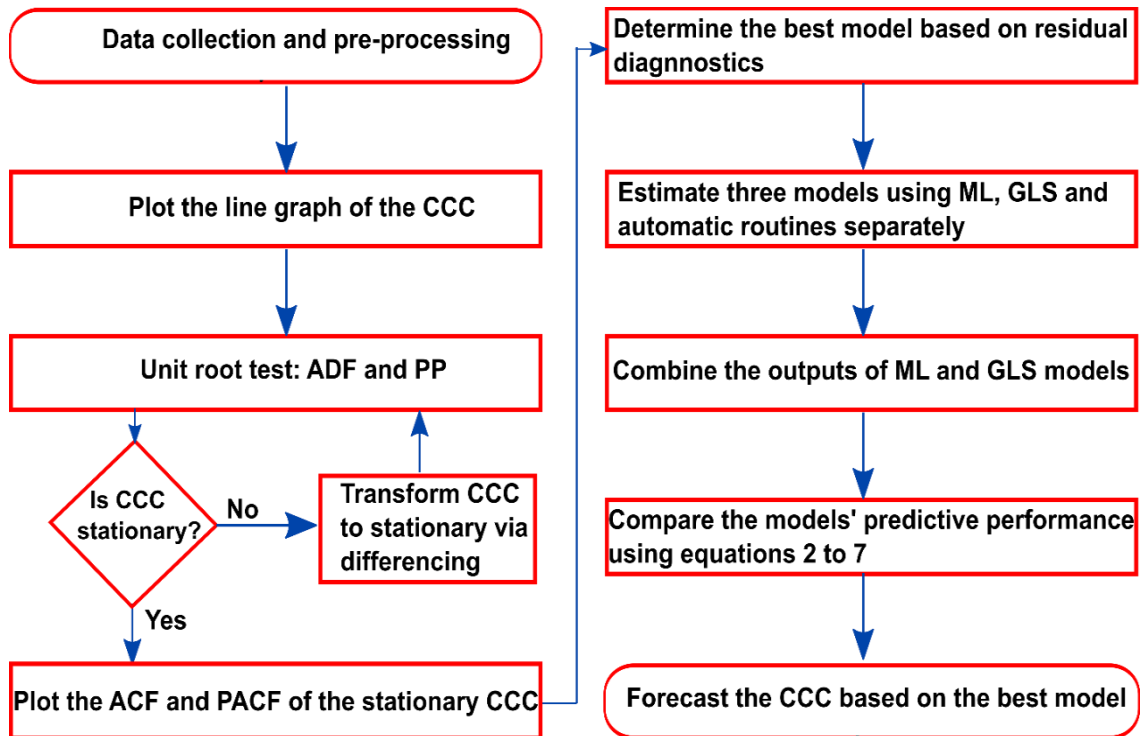


Figure 2: ARIMA Models Algorithm (Source; Author)

3.1.3 Evaluation criteria

This study employs the following estimation criteria: MAE, RMSE, SMAPE, Theil Inequality Coefficient and MAPE. The estimation criteria are given in Mathematical Equations as: (2), (3), (4), (5), (6) and (7) in that order.

$$RMSE = \sqrt{\frac{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2}{h}} \quad (2)$$

$$MAE = \frac{1}{h} \sum_{t=T+1}^{T+h} |\hat{y}_t - y_t| \quad (3)$$

$$MAPE = \frac{1}{h} \left[\sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{y_t} \right| \right] \times 100 \quad (4)$$

$$SMAPE = \frac{1}{h} \left[\sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{\hat{y}_t + y_t} \right| \times 2 \right] \times 100 \quad (5)$$

$$U_1 = \sqrt{\frac{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2}{h}} \cdot \left(\sqrt{\frac{\sum_{t=T+1}^{T+h} \hat{y}_t^2}{h}} + \sqrt{\frac{\sum_{t=T+1}^{T+h} y_t^2}{h}} \right)^{-1} \quad (6)$$

The notations in Equations two to seven comprise the definite value y_t , the estimate value \hat{y}_t , the estimate horizon h and the training or testing sample T .

3.2 Phase II: Data and Methodology for the Daily Cases using AI

3.2.1 Study area and data

In this phase, the daily and cumulative affirmed cases of COVID-19 were modeled by means of machine learning models comprising of SVM, ANN, ANFIS as well as conventional MLR. Ensemble modeling (ANN-E and SVM-E) was also carried out to increase the predictive accuracy of the single models in modeling daily cases. The complete confirmed cases of COVID-19 and cumulative mortality figures due to the infection by the novel virus in Sudan, Morocco, Rwanda, Uganda, Gabon, Cameroon, Namibia, South Africa, Senegal and Nigeria were considered in this research. These respective nations have been selected crosswise diverse African provinces to symbolize multiplicity. Additionally, their Statistics of affirmed cases are order of extents disparities, which make available sufficient opportunity to assess the planned simulations for the territories with both low and high amounts of confirmed cases. Additionally, a small number of these states has documented the confirmed cases and the rate of mortality moderately longer time than several other nations, which is one more evidence for selecting them. Figure 4 illustrate the map of the African states selected in this research.

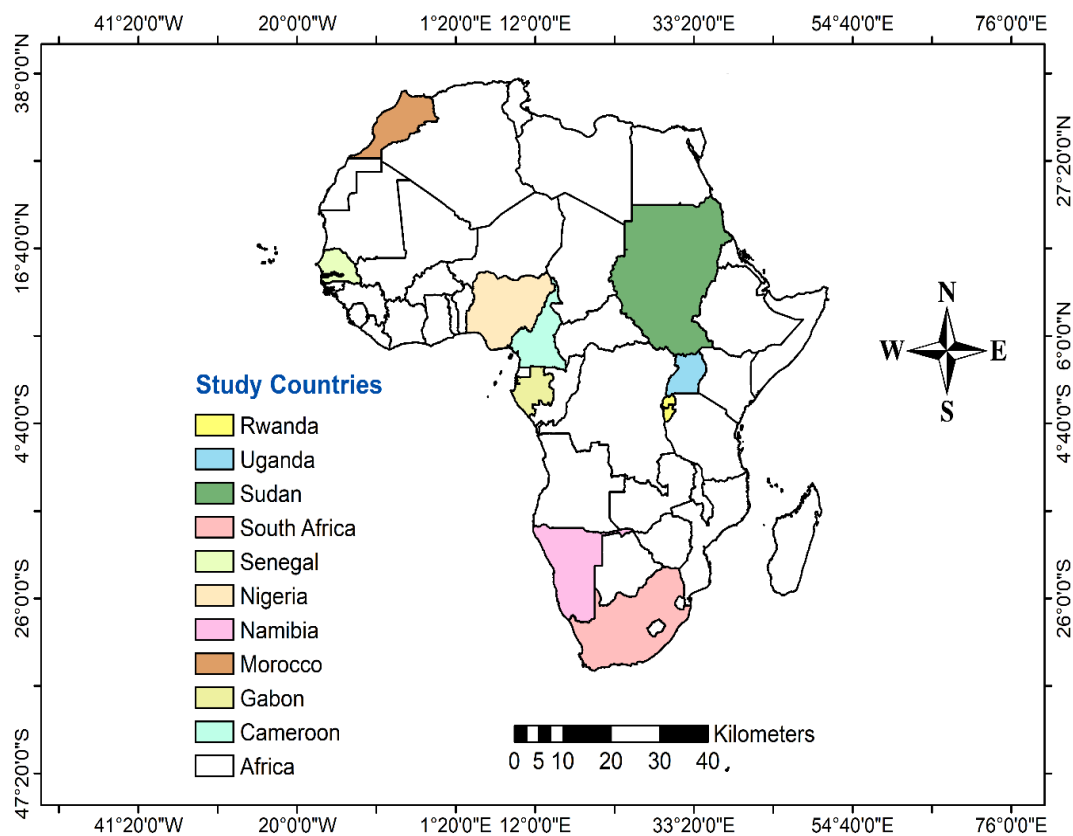


Figure 3. Map of the study area

Data employed for this research were categorized into two parts, involving seventy five percent and twenty five percent respectively. The 75% was employed for the training, whereas the 25% was used for the validation reasons as well. Therefore, the predictable affirmed incidences and cumulative mortality belonging to the validation dataset were matched with the observed information. The affirmed (daily) cases and cumulative figures sequential data were generated from W.H.O dashboard. Table 2 demonstrates the nations, period of the total information as well as descriptive statistics of the data.

Table 2. Statistical description of the study nations

Nation	Duration	Min.	Max.	Mean	St. Deviation
Morocco	1/3/2020 - 16/12/2021	0	12039	1453	2095
Sudan	1/3/2020 - 16/12/2021	0	1215	69	101
Uganda	1/3/2020 - 16/12/2021	0	20692	196	852
Rwanda	1/3/2020 - 16/12/2021	0	3141	154	325
Nigeria	1/3/2020 - 16/12/2021	0	3402	336	413
Senegal	1/3/2020 - 16/12/2021	0	1722	113	178
Namibia	1/3/2020 - 16/12/2021	0	3937	205	413
South					
Africa	1/3/2020 - 16/12/2021	0	37875	4925	5677
Gabon	1/3/2020 - 16/12/2021	0	640	57	117
Cameroon	1/3/2020 - 16/12/2021	0	8681	164	689

3

As coronavirus disease 2019 incidences in Africa begins to be affirmed in the month of March, 2020, this work assumes 1st March, 2020 as the initial day of data collection up to 16th December, 2021. From Table 2, all the selected states possessed a least value of zero incidence, that shows the duration of the coronavirus disease 2019 incidences, also, it can be observed from descriptive statistics table that three countries which are Morocco, Uganda as well as S. Africa possessed the biggest figures of the daily-affirmed cases of COVID-19 with 12039, 20692 as well as 37875, correspondingly. In this research three machine learning models which comprised of ANN, ANFIS, SVM and conventional linear MLR were used. In this work, ANN was trained using MATLAB and a Levenberg Marquardt optimization algorithm along feed-forward back propagation network. Common features of the artificial neural network were built up in accordance with those used in prior studies.

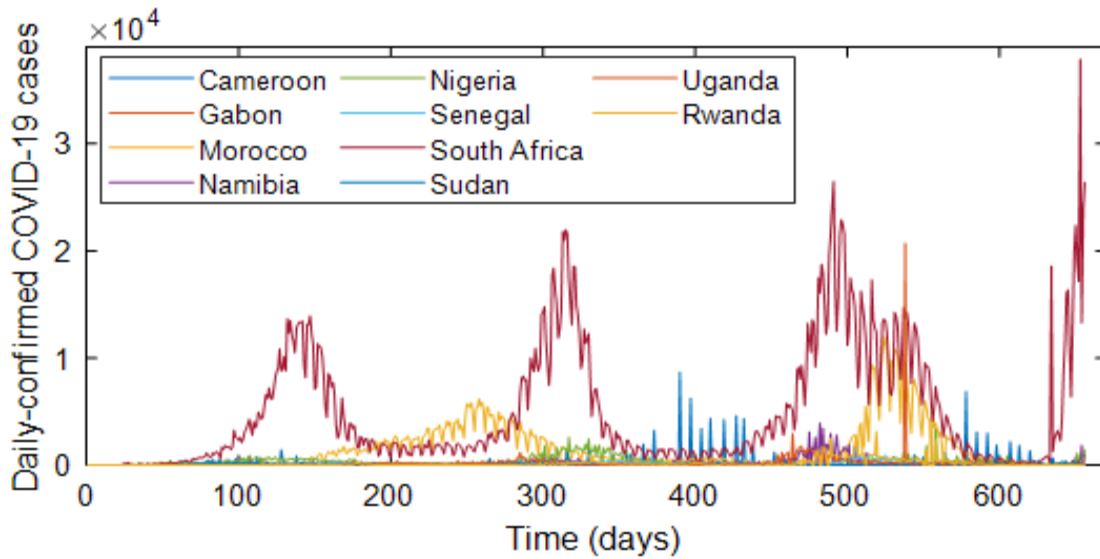


Figure 4. Schemes for the time sequence of the daily-affirmed incidences

3.2.2 Model Validation

To make sure suitable outcomes are attained for this work, k-fold crossing validation was used. In ten countries, the sample of the dataset was arbitrarily separated into four folds in this work as could be illustrated from Figure 6. In this manner, three folds was employed for the training and the residual fold sub-sample was applied for validation purpose. For distinct 4 to 1 training sub-samples and a particular validation subsample, the procedure repeats up to k (four) times. The k outcomes from the folds were at that time be close to produce the final distinct outcome. K-fold validation has the benefit of using the whole set of observations for both validation and training. Figure 5 displays the k-fold employed, Table 3 displays the total amount of observations employed for both training and validation.

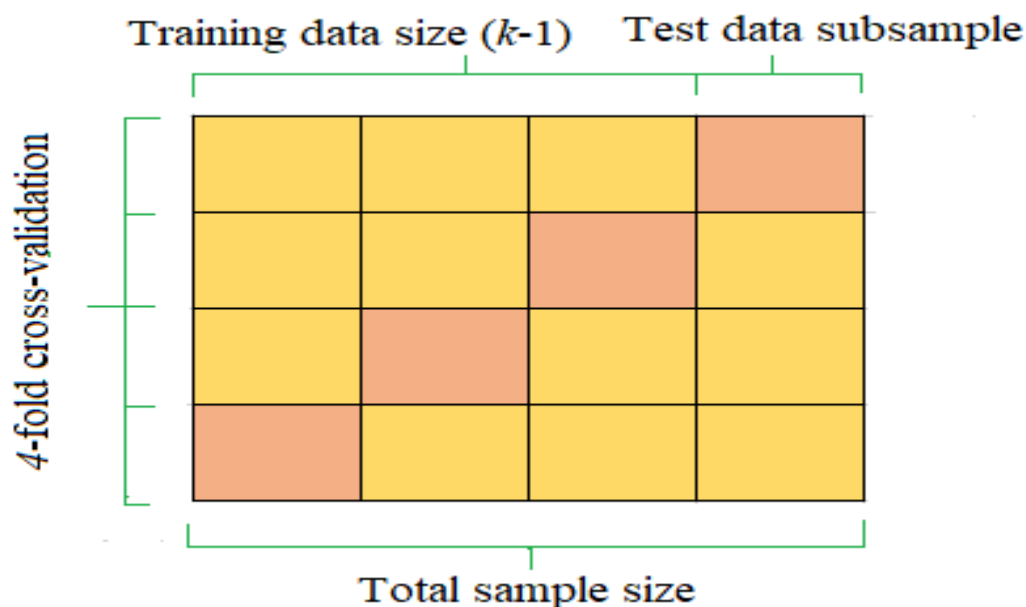


Figure 5. K-fold validation employed in the finding

Table 3. Cumulative cases, data partitioning and validation

Nation/Country	Cumulative incidences	No. of observation	Training		Validation sample (25%)	Validation type
			of data set (75%)			
Morocco	951763	656	492		164	4-fold
Sudan	45112	656	492		164	4-fold
Uganda	128369	656	492		164	4-fold
Rwanda	100978	656	492		164	4-fold
Nigeria	220109	656	492		164	4-fold
Senegal	74105	656	492		164	4-fold
Namibia	134160	656	492		164	4-fold
South Africa	3231039	656	492		164	4-fold
Gabon	37681	656	492		164	4-fold
Cameroon	107662	656	492		164	4-fold

3.2.3 Data normalization and performance criteria

Normalization of dataset is generally carried out for artificial intelligence modeling in order to make sure all the variables partake equivalent consideration and to remove the dimensional inconsistency in them. Regarding the data normalization in this finding, the annotations were scaled amid or between zero and one. The general formula is given as:

$$DC_n = \frac{DC_i - DC_{min}}{DC_{max} - DC_{min}} \quad (7)$$

Where DC_n , DC_{max} , DC_{min} and DC_i symbolize the normalized, minimum, i th, maximum value respectively of daily-verified COVID-19 cases.

To assess the precision of the simulations used for the modeling of COVI-19 outbreak within ten African nations, four universal statistical indices were employed comprising of mean square error (MSE) (Bhagat et al., 2020), mean absolute deviation, determination coefficient (R^2) (Abdullahi et al. 2019c) and RMSE, and given by;

$$MAD = \frac{1}{N} \sum_{i=1}^n |p_i - a_i| \quad (8)$$

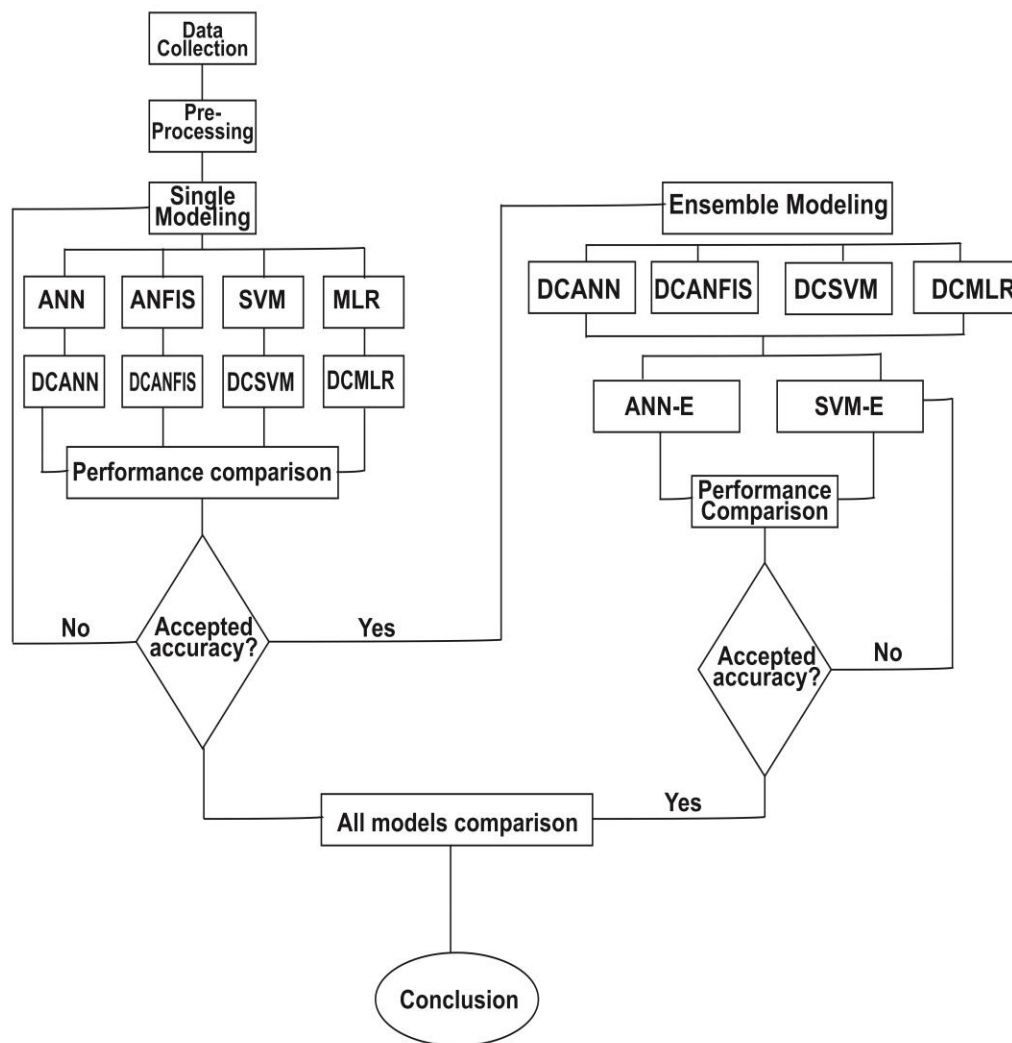
$$MSE = \frac{1}{N} \sum_{i=1}^n |p_i - a_i| \quad (9)$$

$$DC = 1 - \frac{\sum_{i=1}^N (a_i - p_i)^2}{\sum_{i=1}^N (a_i - \bar{a})^2} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (a_i - p_i)^2}{N}} \quad (11)$$

Where a_i , p_i , \bar{a} and N are the definite values, projected values, mean of the definite values as well as sum of observations correspondingly.

Figure 6. Main methodology used for the ML



3.2.4 Ensemble approaches (Modeling)

In trying to maintain simplification and likewise to take advantage from the impacts of all techniques, ensemble model is designed that make use of the distinct output of every procedure with certain preference degree allocated to each of them with the support of an intermediary to offer the output.

In this study, the ensemble modeling was carried out via two non-linear ensemble approaches SVM-E and ANN-E. Though, further processes like adaptive neuro-fuzzy inference system can be applied for the ensemble modeling (non-linear), the selection of the stated techniques are based on (a) ANN ensemble is the utmost broadly non-linear ensemble approach used, it is easier to apply and results to effective prediction efficacy

while (b) support vector machine ensemble has at no time been applied before by any researcher. The overall technique of the ensemble based modeling is presented in Figure 9.

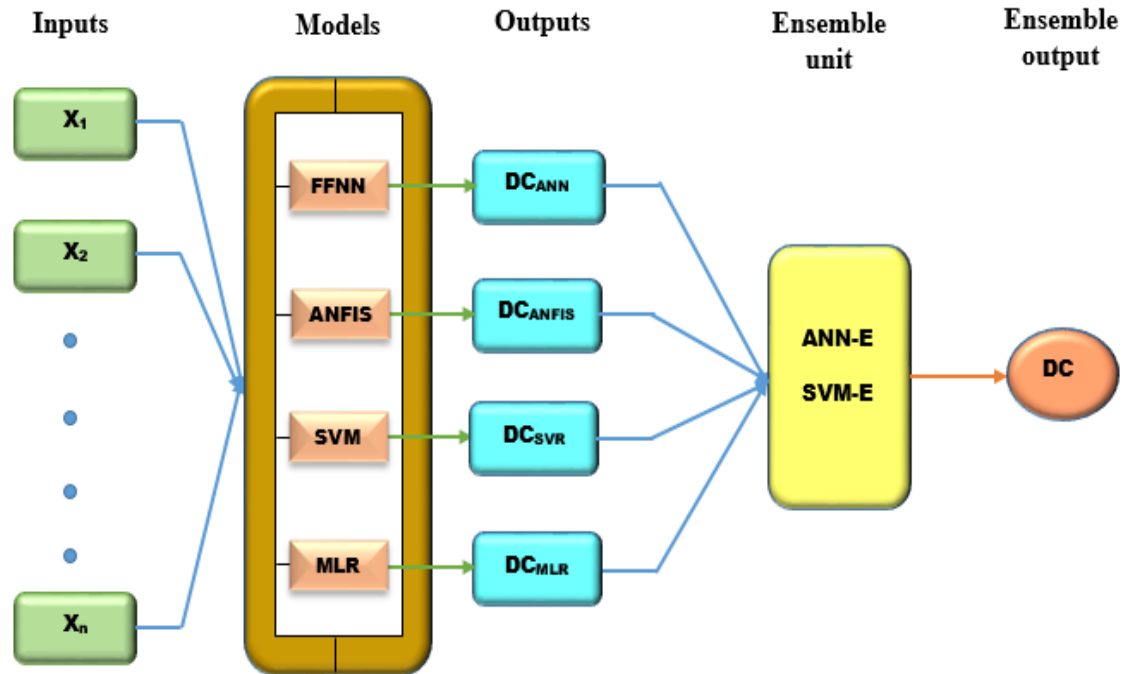


Figure 7. The overall procedure of novel ensemble approach (Source; author)

The possibility of applying ensemble idea to supplementary increase COVID-19 prediction precision was studied in regards to daily and cumulative cases in this work. Furthermore, the cumulative deaths as a result of the COVID-19 pandemic were modeled. Primarily, machine learning models comprising of SVM, ANFIS, ANN as well as traditional MLR model were employed for daily-confirmed coronavirus disease 2019 cases and cumulative deaths prediction respectively across ten African nations including Sudan, Morocco, South Africa, Namibia, Rwanda, Uganda, Senegal, Nigeria, Cameroon and Gabon. Subsequently, 2 ensemble techniques were used to advance the coronavirus disease 2019 prediction of the daily cases.

The major benefits of employing ensemble techniques includes: (a) Considering if the fundamental procedure for a specific issue is persuaded by a non-linear and linear aspect is challenging work to achieve in real circumstances desirable technique to be selected between the rest. Consequently, for a distinctive task, selecting a fitting process has developed to be a problematic assignment before investigators. Consequently, the

challenge of choosing the greatest and suitable algorithms might be taking care by ensemble methods. (b) Actual global procedures could comprise both the linear features and non-linear characteristics. Therefore, in that situation, non-linear machine learning models (ANFIS, SVM, and ANN,) and the linear multi regressions will either be adequate for the series of time prediction because multiple linear regressions might not go along with the nonlinear associations and machine learning simulations may amplify faults of a linear form. Therefore, by linking the machine learning and multiple linear regression algorithms, the procedure's composite way might be taken more precisely. (c) No typical approach that can faultlessly identify the unique forms of time sequence as a result of the compound feature of the actual global delinquent. The two ensemble procedures are as follows:

ANN Ensemble

Regarding the artificial neural network ensemble, the daily-verified coronavirus disease 2019 incidences were applied as outputs function of the individual representations based on artificial neural network model;

$$DC_{ANN-E} = f(DC_{ANN}, DC_{ANFIS}, DC_{SVM}, DC_{MLR})$$

(13)

Where DC_{ANN-E} stands for the daily-affirmed values by ANN ensemble, DC_{ANN} , DC_{ANFIS} , DC_{ANFIS} , DC_{SVM} as well as DC_{MLR} are the respective outputs of the daily-affirmed cases of the respective states created by SVM, ANFIS, ANN as well as MLR, correspondingly. Figure 10 demonstrates the overall nonlinear ensemble method in regards to ANN-E.

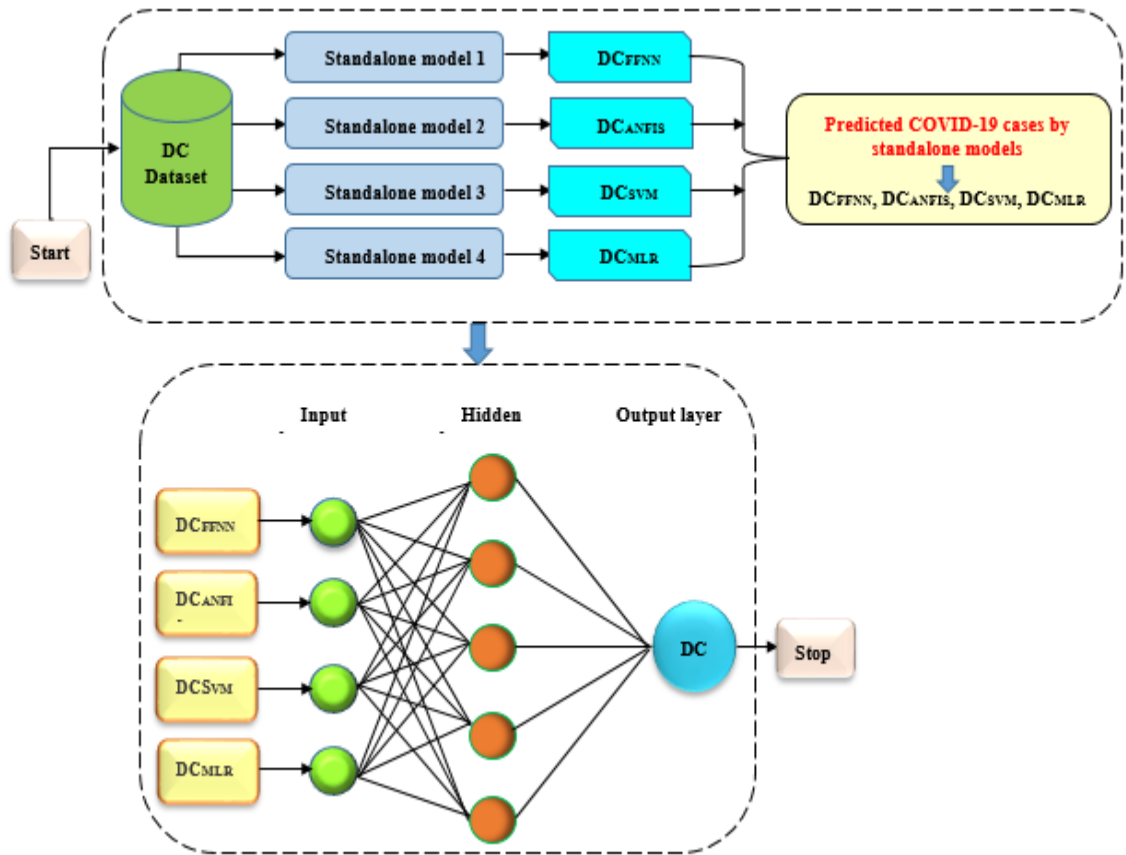


Figure 8. Planned ANN-E procedure (Source; author)

As observed from Figure 10, the coronavirus disease 2019 dataset acquired after moving over preprocessing data, SVM, ANFIS, ANN and MLR algorithms have been used as individual models. ANN ensemble modeling of the coronavirus disease 2019 was later carried out by applying artificial neural network in the form of ensemble (kernel). In this manner, the single models' outputs were applied to replace the neurons in the layer of input, which consist of input, hidden as well as output layers configuration. Through its capacity to check the lowest essential error, FFNN with the algorithm of back propagation was used. The Levenberg Marquardt (LM) has been engaged as the training procedure as well. Another procedure used was trial and error in order to determine the hidden layer neurons of the optimum figure. To have adequate repetitions for increase precision, the epoch amount was established through trial and error to drop amid one hundred and two hundred accordingly.

SVM-E

The support vector machine-based ensemble modeling was done using the SVM kernel to incorporate or combine the single models' outputs, given as;

$$DC_{SVM-E} = f(DC_{ANN}, DC_{ANFIS}, DC_{SVM}, DC_{MLR}) \quad (14)$$

Where DC_{SVM-E} represent daily-confirmed coronavirus disease 2019 values by SVM-E for both countries. The overall methodology and approach proposed by this research is given in Figure 11.

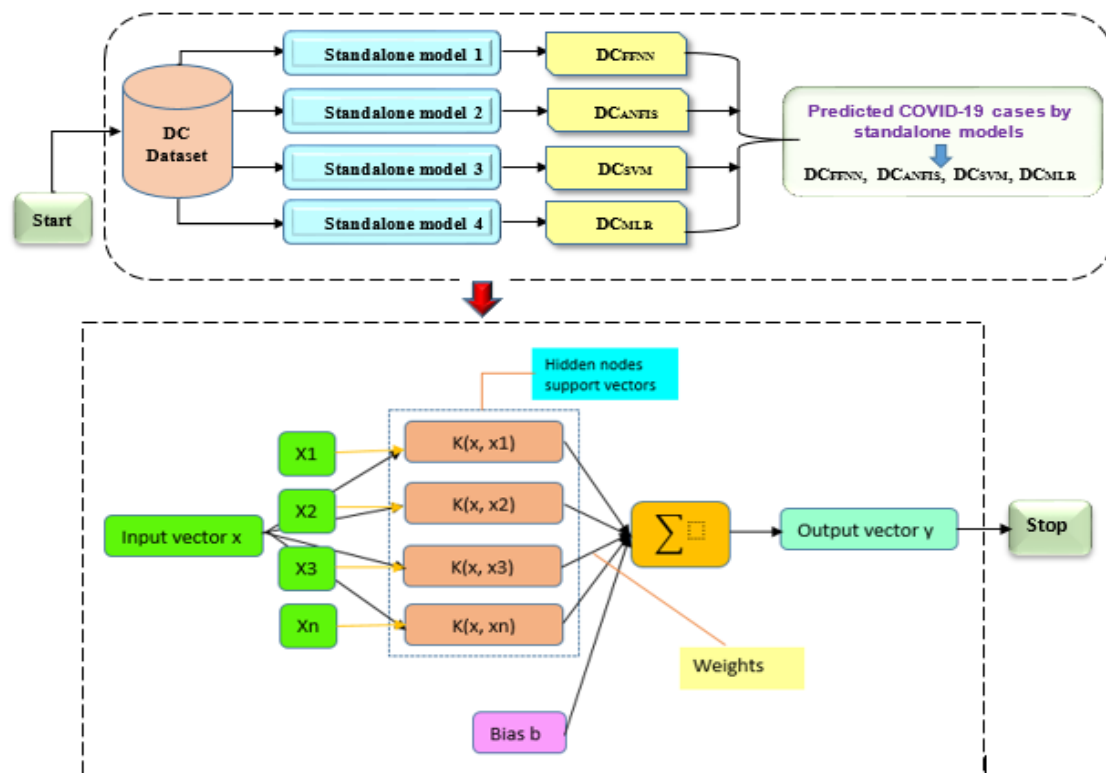


Figure 9: The planned SVM-E procedure applied

For support vector machine ensemble (SVM-E) modeling was accomplished using outputs of the single models (SVM, ANFIS, ANN and MLR). The individual outputs were engaged to substitute the variable of the input layer as illustrated in Figure 11. For a complex nonlinear procedure, the Gaussian kernel function is the most appropriate (Ghorbani et al. 2016). Consequently, Gaussian function (kernel) function was selected for the support vector machine ensemble to deal with the undefined and complicated coronavirus disease 2019 outbreak.

CHAPTER V

Findings and Discussion

4.1 Phase I: Results and discussions for the Cumulative Cases Using ARIMA

Figure 13 illustrates the time series graph for the CCC within the West African states. Observation of the figure indicates that the CCC for each country does not have constant mean and variance. Therefore, Philips-Perron and ADF was the formal unit root tests employed in this work. The outcomes of the tests are illustrated in table two. ADF test reported in the table displays that the COVID-19 cumulative incidences for each nation is non-stationary at intensity, but each becomes stationary during the first disparity except the CCC for BF, NG, SN and TG which becomes stationary after the second difference. In other words, the ADF unit root test shows that the CCC for BF, NG, SN and TG is I (2), while the CCC for the rest of West African nations is I (1). However, the PP unit root test reveals that the CCC for all the West African countries members is I (1). The CCC for each country can be modeled as ARIMA (p,1,q) processes, as each is integrated of order one. The moving average (p) and autoregressive order (p) is selected based on the combination that gives minimum Schwarz Information Criteria (SIC), along using white-noise errors. Since the CCC for each country is one (1), the first discrepancy of apiece cumulative COVID-19 case is taken before estimation.

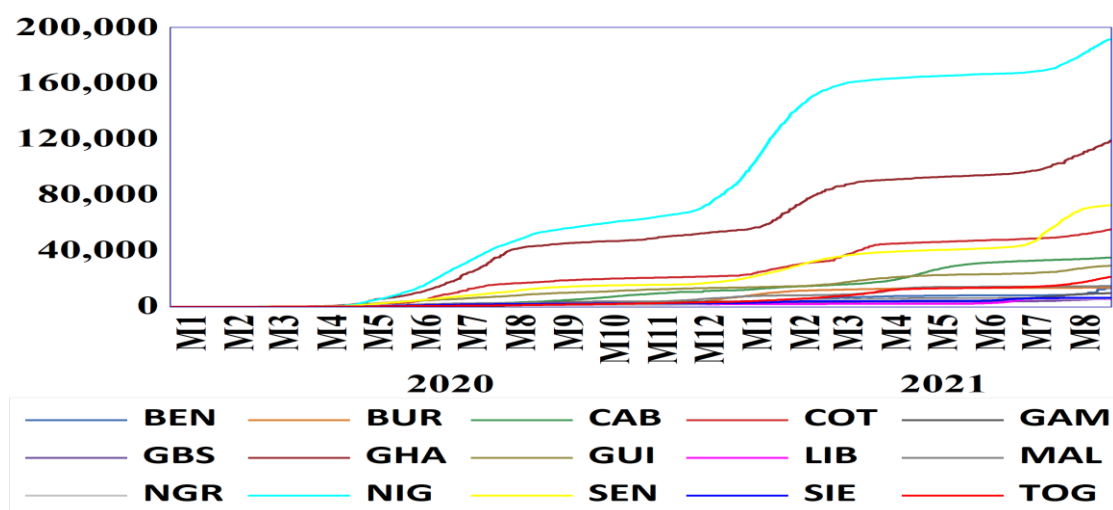


Figure 10: Line graph for the CCC of West African countries

Table 4: ADP and PP unit root test outcome

Variab les	No ne	invaria ble	invariable and inclination/t rend	None	Invaria ble	invariable and inclination/t rend	conclus ion
ADF Test							
BF	0.7 21	-0.676	-1.733	-2.283	-2.711	-2.693	I(2)
BJ	5.5 3	3.156	0.955	- 7.548* **	- 23.653* **	-24.048****	I(1)
CI	2.7 47	0.297	-1.925	- 2.523* **	- 3.912** *	-3.947****	I(1)
CV	2.9 1	1.387	-1.996	- 1.907* **	- 2.906** *	-3.053****	I(1)
GH	3.0 86	0.263	-1.756	- 2.525* **	- 4.014** *	-4.035****	I(1)
GM	3.1 17	1.23	-1.139	- 3.700* **	- 4.524** *	-4.712****	I(1)
GN	3.2 76	0.701	-1.554	- 2.406* **	- 4.611** *	-4.658****	I(1)
GW	3.8 28	0.555	-0.577	- 3.706* **	- 4.815** *	-4.845****	I(1)

Variab les	No ne	invaria ble	invariable and inclination/t rend	None	Invaria ble	invariable and inclination/t rend	conclus ion
LR	1.8 95	0.529	-1.109	- 2.622* **	- 2.934** *	-3.005***	I(1)
ML	1.4 39	-0.381	-2.105	- 3.030* **	- 3.881** *	-3.870***	I(1)
NE	3.0 68	-0.239	-0.669	- 5.197* **	- 6.099** *	-6.094***	I(1)
NG	0.6 06	-0.824	-3.081	-1.611	-2.37	-2.338	I(2)
SL	1.6 06	-0.542	-2.66	- 3.199* **	- 4.027** *	-4.011***	I(1)
SN	2.1 02	0.944	-1.596	-2.133	-2.753	-2.935	I(2)
TG	3.6 94	2.87	0.233	-1.693	-2.543	-3.536	I(2)
PP Test							
BF	2.4 62	-0.099	-1.244	- 16.038 ***	- 19.557* **	-19.582***	I(1)
BJ	5.2 05	2.935	0.721	- 24.078 ***	- 24.124* **	-24.214***	I(1)

Variab les	No ne	invaria ble	invariable and inclination/t rend	None	Invaria ble	invariable and inclination/t rend	conclus ion
CI	4.6 62	0.442	-1.719	- 17.937 ***	- 22.901* **	-23.005***	I(1)
CV	5.7 13	2.193	-1.968	- 19.158 ***	- 26.790* **	-28.150***	I(1)
GH	4.7 58	0.316	-1.634	- 20.463 ***	- 25.560* **	-25.627***	I(1)
GM	4.2 32	1.511	-0.992	- 28.125 ***	- 29.539* **	-29.758***	I(1)
GN	7.1 98	1.222	-1.159	- 18.280 ***	- 27.406* **	-27.669***	I(1)
GW	4.4 96	0.563	-0.648	- 22.776 ***	- 24.570* **	-24.641***	I(1)
LR	3.0 76	1.234	-0.558	- 26.332 ***	- 26.955* **	-27.255***	I(1)
ML	3.5 23	0.17	-1.648	- 16.510 ***	- 20.553* **	-20.618***	I(1)
NE	2.7 82	-0.366	-1.105	- 26.527 ***	- 26.009* **	-25.995***	I(1)

Variab les	No ne	invaria ble	invariable and inclination/t rend	None	Invaria ble	invariable and inclination/t rend	conclus ion
NG	3.8 27	0.19	-1.665	- 5.496* **	- 10.172* **	-10.229***	I(1)
SL	3.9 09	0.216	-1.476	- 11.332 ***	- 14.524* **	-14.571***	I(1)
SN	6.1 32	3.207	0.381	- 5.288* **	- 5.958** *	-5.885***	I(1)
TG	7.4 79	4.682	0.579	- 20.141 ***	- 24.571* **	-28.066***	I(1)

*** Represents null hypothesis rejection, the unit root of the series is at 1% significance.

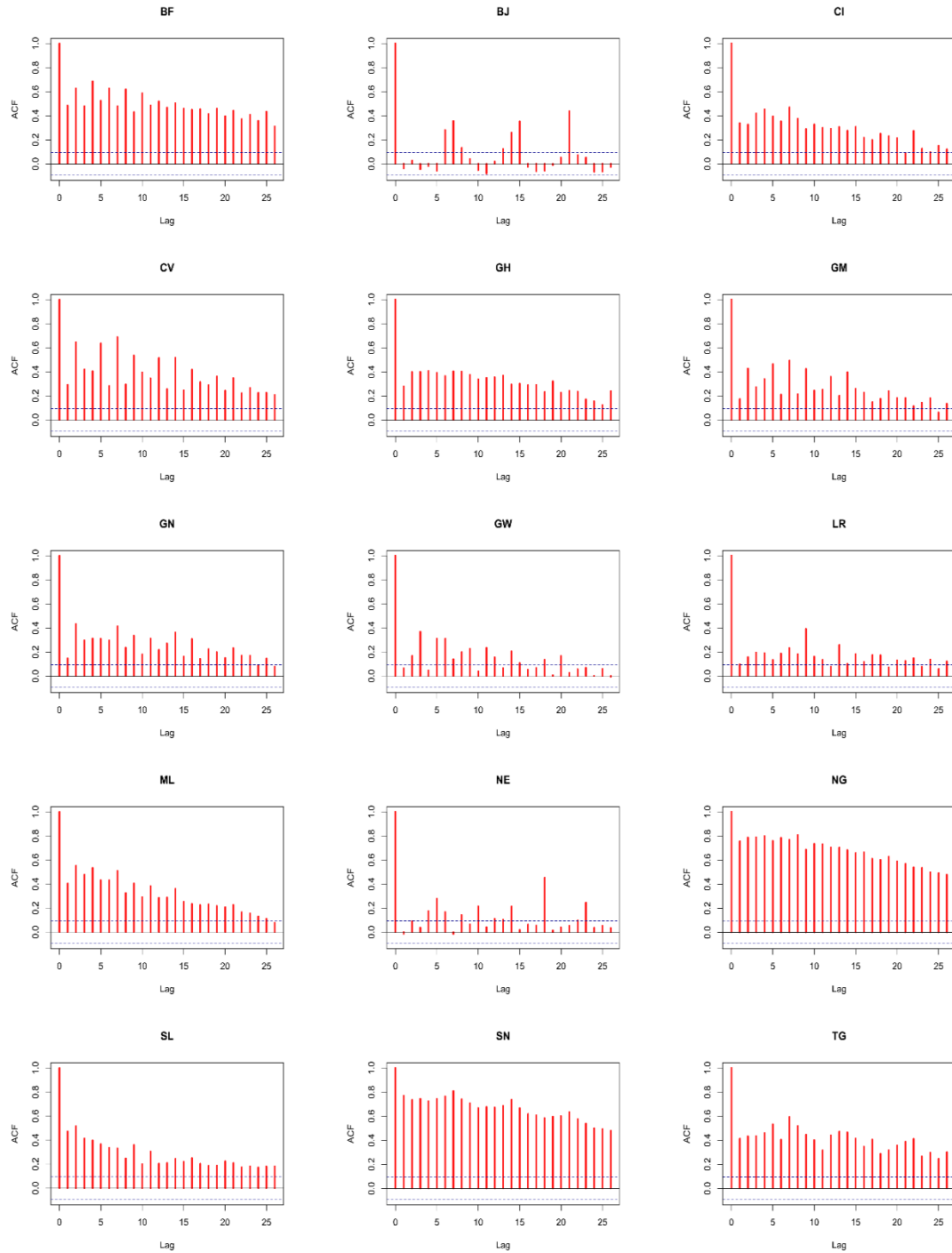


Figure 11: Autocorrelation Function of the primary variation of Cumulative COVID-19 Case for ECOWAS Countries

Possible moving order is evaluated by the autocorrelation while possible auto-regressive order was applied to determine by PACF. Figures 14 and 15 depict the graphs of the ACF and PACF respectively.

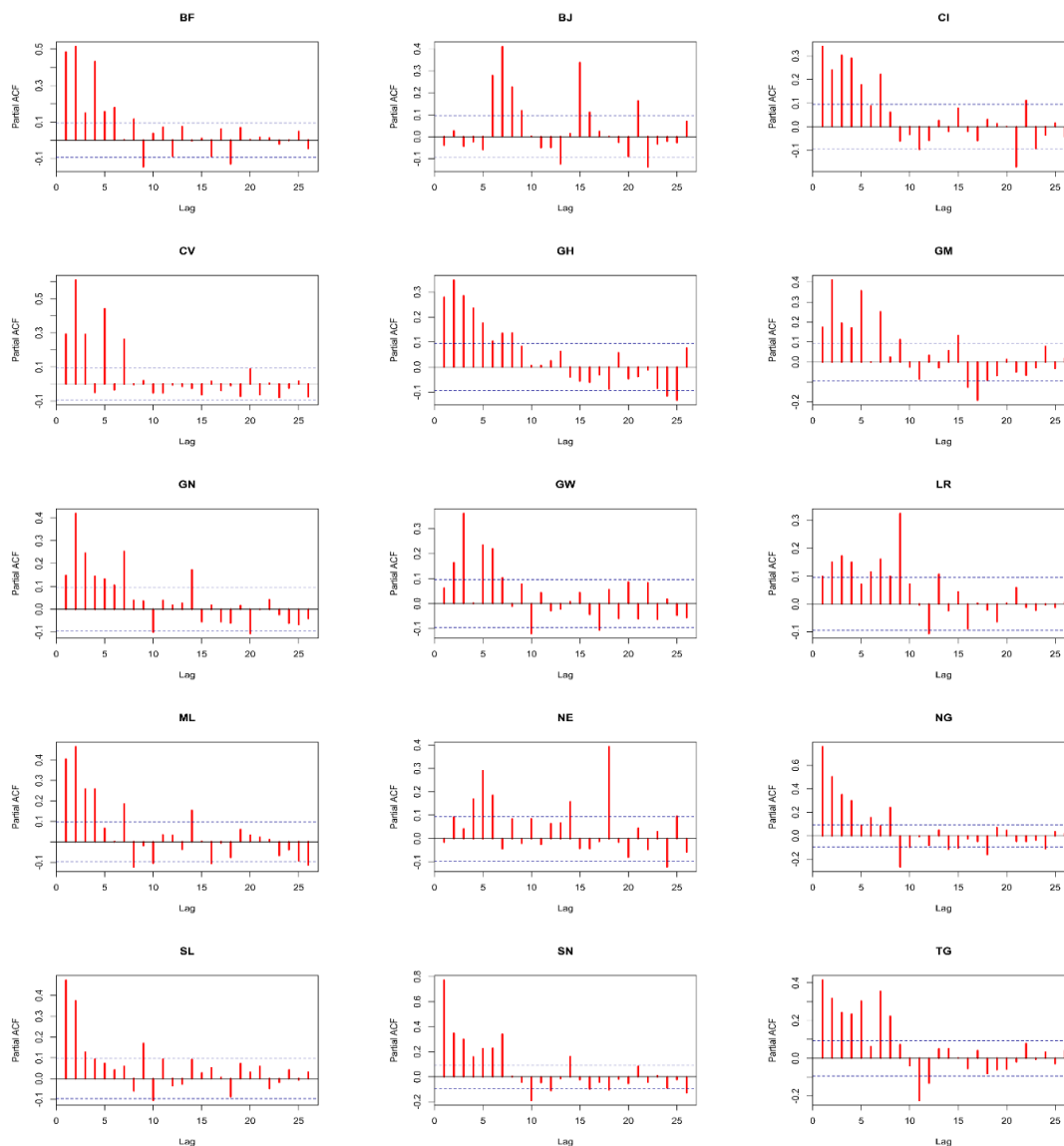


Figure 12: PACF of the initial variation of Cumulative incidences for West African nations

Forecast evaluation for the training sample is presented in Table 3. The table contains the forecast statistics of the best performing models for the CCC for the training sample. Based on statistical indices RMSE, ARIMAGLS is the best performing model for four countries, while ARIMAML-ARIMA GLS is the best for seven countries; ARIMAML is the best for three countries. Only in the case of Cote D'Ivoire AUTOARIMA happens to be the best model. It is noticeable that the RMSE of ARIMAML is much lower than that of AUTOARIMA for all the countries except Cabo Verde, Cote D'Ivoire, Sierra Leone and Guinea Bissau. The AUTOARIMA outperforms other models for Cabo Verde, while

AUTOARIMA and ARIMAML have the same estimates for the last three countries. In addition to this, it is observable that ARIMAGLS improves the forecast accuracy a bit for all the countries except Cabo Verde.

Table 5: Training sample Assessment outcome

Models used	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Panel A: Burkina Faso						
ARIMAML	35.753940	19.445740	1.297034	1.345297	0.002606	0.714019
AUTOARIMA	56.554490	33.256670	1.647193	1.606828	0.004123	0.927168
ARIMAGLS	35.753780	19.418900	1.305899	1.359271	0.002606	0.722071
ARIMAML- ARIMAGLS	35.753290	19.432080	1.301446	1.352223	0.002606	0.717968
Panel B: Benin						
ARIMAML	47.085700	26.943410	12.815130	6.165776	0.006119	6.539295
AUTOARIMA	73.836520	44.124330	3.718659	3.577145	0.009596	0.872731
ARIMAGLS	51.959010	29.788180	24.196570	9.410529	0.006757	11.298730
ARIMAML- ARIMAGLS	48.063600	27.178100	18.312540	7.919193	0.006248	8.811751
Panel C: Cote D'Ivoire						
ARIMAML	127.426200	69.579000	8.438018	3.557386	0.002622	4.333505
AUTOARIMA	127.426200	69.579000	8.438018	3.557386	0.002622	4.333505
ARIMAGLS	127.423500	69.466430	8.108744	3.491126	0.002622	4.142305
ARIMAML- ARIMAGLS	127.423800	69.521210	8.273296	3.524525	0.002622	4.237808
Panel D: Cabo Verde						
ARIMAML	57.338700	30.308660	5.977955	4.205276	0.002550	1.361919
AUTOARIMA	56.207580	30.311820	4.370995	3.653276	0.002493	1.092574
ARIMAGLS	57.362250	30.185890	1.870129	1.974683	0.002551	0.982670
ARIMAML- ARIMAGLS	57.341320	30.224200	3.814714	3.202685	0.002550	1.087047
Panel E: Ghana						
ARIMAML	253.914300	162.485600	12.539630	3.981014	0.002326	8.439755
AUTOARIMA	362.874200	267.797700	1.997441	1.972932	0.003319	0.744331
ARIMAGLS	253.905600	162.242600	11.546440	3.849326	0.002326	7.605292

Models used	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
ARIMAML-	253.906100	162.357800	12.037140	3.910905	0.002326	8.022306
ARIMAGLS						

Panel F: Gambia

ARIMAML	27.525360	13.552750	7.384877	6.046966	0.003959	1.853391
AUTOARIMA	42.967290	24.418330	2.879910	2.851205	0.006185	0.966830
ARIMAGLS	27.524390	13.523630	7.132454	5.882828	0.003958	1.784544
ARIMAML-	27.524520	13.537920	7.258657	5.965219	0.003958	1.818819
ARIMAGLS						

Panel G: Guinea

ARIMAML	42.816840	30.291380	6.894573	3.931124	0.001719	2.921342
AUTOARIMA	99.196220	78.974840	2.136562	2.172763	0.003977	0.732991
ARIMAGLS	42.820720	30.262250	6.297667	3.773340	0.001720	2.627702
ARIMAML-	42.818140	30.276610	6.596118	3.854295	0.001720	2.773974
ARIMAGLS						

Panel H: Guinea Bissau

ARIMAML	18.687360	9.840952	2.172158	1.984021	0.003809	0.610291
AUTOARIMA	18.687360	9.840952	2.172158	1.984021	0.003809	0.610291
ARIMAGLS	18.687270	9.815966	2.146481	1.973860	0.003809	0.617432
ARIMAML-	18.687230	9.828305	2.159295	1.978925	0.003809	0.613807
ARIMAGLS						

Panel I: Liberia

ARIMAML	9.162097	4.850161	2.793990	2.362586	0.003120	1.254766
AUTOARIMA	12.232800	7.409246	2.153113	2.091775	0.004164	0.836006
ARIMAGLS	9.161940	4.843813	2.718351	2.316417	0.003120	1.226379
ARIMAML-	9.161953	4.846868	2.756164	2.339612	0.003120	1.240415
ARIMAGLS						

Panel J: Mali

ARIMAML	37.884530	20.138320	1.690162	1.456238	0.002951	0.841306
AUTOARIMA	38.282030	20.422980	1.385906	1.517089	0.002982	0.753752
ARIMAGLS	37.884110	20.135570	1.661189	1.441783	0.002951	0.818413
ARIMAML-	37.884090	20.136890	1.675502	1.448875	0.002951	0.829739
ARIMAGLS						

Models used	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Panel K: Niger						
ARIMAML	38.299560	14.483870	5.598795	2.944476	0.006597	1.365124
AUTOARIMA	81.163740	39.585650	2.730086	2.759480	0.013925	0.792262
ARIMAGLS	38.297930	14.514670	5.461052	2.931652	0.006597	1.327775
ARIMAML- ARIMAGLS	38.297920	14.497180	5.529858	2.938091	0.006597	1.346264
Panel L: Nigeria						
ARIMAML	207.574600	117.715600	12.543480	6.138207	0.001149	4.789871
AUTOARIMA	234.237200	136.145000	1.646047	1.705520	0.001297	0.834606
ARIMAGLS	208.299900	117.294100	1.435163	1.573133	0.001153	0.900101
ARIMAML- ARIMAGLS	207.727300	117.333000	6.764882	4.492818	0.001150	2.434468
Panel M: Sierra Leone						
ARIMAML	12.175520	6.407529	2.545493	1.873909	0.002303	1.446721
AUTOARIMA	12.175520	6.407529	2.545493	1.873909	0.002303	1.446721
ARIMAGLS	12.175450	6.403291	2.504861	1.851391	0.002303	1.417956
ARIMAML- ARIMAGLS	12.175440	6.405410	2.525177	1.862676	0.002303	1.432304
Panel N: Senegal						
ARIMAML	44.318350	28.697340	1.503405	1.381445	0.001053	0.971118
AUTOARIMA	65.697620	46.642540	1.231426	1.245330	0.001559	0.675517
ARIMAGLS	44.318850	28.672020	1.330413	1.283929	0.001053	0.886259
ARIMAML- ARIMAGLS	44.315300	28.682680	1.416677	1.334063	0.001052	0.922632
Panel O: Togo						
ARIMAML	37.364770	21.400760	3.978480	3.160694	0.003624	1.682209
AUTOARIMA	42.089980	23.128780	1.538623	1.601028	0.004080	0.962236
ARIMAGLS	37.363830	21.398280	3.771889	3.056500	0.003624	1.582242
ARIMAML- ARIMAGLS	37.363760	21.399280	3.875137	3.109029	0.003624	1.631803

The forecast evaluation for the testing set of data is illustrated in Table four. It contains the forecast statistics of the best performing models for the CCC for the testing sample. The table shows that ARIMA with generalized least square performs better than other models for eight nations; ARIMA with maximum likelihood beats other models for six countries. It also worth noting that RMSE and Theil U_1 agree in chooses the same model as the best in most of the cases. It is also important to note that whenever other exactness measures (Maximum Absolute Error, Maximum Absolute Percentage Error, SMAPE and Theil) contradict RMSE, the RMSE measure supersedes.

Table 6: Testing sample Assessment outcome

Models employed	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Panel A: Burkina Faso						
ARIMAML	5.372085	3.871513	0.028527	0.028526	0.000198	0.865353
AUTOARIMA	47.456230	47.033800	0.347482	0.348097	0.001756	7.684733
ARIMAGLS	5.305577	3.734083	0.027509	0.027509	0.000196	0.854660
ARIMAML- ARIMAGLS	5.337517	3.801341	0.028007	0.028007	0.000197	0.859793
Panel B: Benin						
ARIMAML	245.731000	75.193300	0.683793	0.700848	0.013861	0.899269
AUTOARIMA	319.851000	128.239000	1.270499	1.306599	0.018112	1.196983
ARIMAGLS	268.427100	75.907060	0.708758	0.734715	0.015160	0.991865
ARIMAML- ARIMAGLS	252.377500	73.358110	0.674869	0.696482	0.014245	0.933340
Panel C: Cote D'Ivoire						
ARIMAML	79.145280	54.565690	0.107086	0.107117	0.000800	0.620500
AUTOARIMA	79.145280	54.565690	0.107086	0.107117	0.000800	0.620500
ARIMAGLS	79.118240	54.493310	0.106941	0.106972	0.000799	0.620307
ARIMAML- ARIMAGLS	79.131200	54.529500	0.107013	0.107044	0.000799	0.620399
Panel D: Cabo Verde						
ARIMAML	40.269020	28.827220	0.090976	0.090916	0.000616	0.458678
AUTOARIMA	42.556050	29.069320	0.091902	0.091824	0.000650	0.484413
ARIMAGLS	41.174340	29.033120	0.091709	0.091644	0.000629	0.468327

Models employed	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
ARIMAML- ARIMAGLS	40.710030	28.896360	0.091230	0.091168	0.000622	0.463353
Panel E: Ghana						
ARIMAML	320.886700	196.227500	0.186775	0.187011	0.001589	0.743299
AUTOARIMA	634.771000	530.565700	0.522859	0.524714	0.003151	1.508296
ARIMAGLS	320.800000	196.027100	0.186564	0.186798	0.001589	0.743153
ARIMAML- ARIMAGLS	320.841100	196.119600	0.186661	0.186896	0.001589	0.743220
Panel F: Gambia						
ARIMAML	64.886070	37.354300	0.485910	0.487644	0.004469	0.798639
AUTOARIMA	72.309630	42.231140	0.541940	0.544193	0.004978	0.885317
ARIMAGLS	64.884980	37.377300	0.486194	0.487893	0.004469	0.798423
ARIMAML- ARIMAGLS	64.885130	37.365570	0.486048	0.487765	0.004469	0.798526
Panel G: Guinea						
ARIMAML	60.831690	43.491860	0.169569	0.169667	0.001214	0.635713
AUTOARIMA	112.050000	87.798140	0.344350	0.344934	0.002237	1.161002
ARIMAGLS	60.841120	43.433610	0.169323	0.169420	0.001214	0.635729
ARIMAML- ARIMAGLS	60.835480	43.462730	0.169446	0.169543	0.001214	0.635712
Panel H: Guinea Bissau						
ARIMAML	21.143380	13.968640	0.301148	0.301637	0.002442	0.661632
AUTOARIMA	21.143380	13.968640	0.301148	0.301637	0.002442	0.661632
ARIMAGLS	21.111450	13.944920	0.300631	0.301108	0.002438	0.660552
ARIMAML- ARIMAGLS	21.127150	13.956660	0.300886	0.301370	0.002440	0.661085
Panel I: Liberia						
ARIMAML	59.540520	32.248110	0.857621	0.867199	0.006741	0.790564
AUTOARIMA	68.621470	44.716890	1.146319	1.146679	0.007759	0.866023
ARIMAGLS	59.489550	32.343350	0.859250	0.868653	0.006735	0.788902

Models employed	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
ARIMAML- ARIMAGLS	59.514080	32.295730	0.858435	0.867926	0.006738	0.789725
Panel J: Mali						
ARIMAML	6.699812	5.727888	0.039499	0.039493	0.000231	0.822180
AUTOARIMA	5.516364	4.148252	0.028569	0.028570	0.000190	0.673473
ARIMAGLS	6.596852	5.627404	0.038804	0.038797	0.000228	0.809385
ARIMAML- ARIMAGLS	6.647855	5.677443	0.039150	0.039144	0.000229	0.815724
Panel K: Niger						
ARIMAML	7.576834	5.851198	0.104770	0.104776	0.000681	0.879970
AUTOARIMA	17.415460	14.558800	0.262036	0.261583	0.001564	1.987962
ARIMAGLS	7.583728	5.822714	0.104252	0.104261	0.000682	0.880460
ARIMAML- ARIMAGLS	7.579721	5.836956	0.104511	0.104518	0.000681	0.880147
Panel L: Nigeria						
ARIMAML	132.423400	79.964350	0.045086	0.045089	0.000384	0.369288
AUTOARIMA	158.295800	94.635480	0.053040	0.053048	0.000459	0.439679
ARIMAGLS	133.564900	79.617080	0.044806	0.044808	0.000387	0.372193
ARIMAML- ARIMAGLS	132.867400	79.546500	0.044799	0.044802	0.000385	0.370356
Panel M: Sierra Leone						
ARIMAML	17.995340	11.262070	0.218379	0.218776	0.001611	0.498836
AUTOARIMA	17.995340	11.262070	0.218379	0.218776	0.001611	0.498836
ARIMAGLS	17.964680	11.262070	0.218315	0.218706	0.001608	0.497861
ARIMAML- ARIMAGLS	17.979830	11.262000	0.218346	0.218740	0.001609	0.498344
Panel N: Senegal						
ARIMAML	175.094800	96.345070	0.177540	0.177902	0.001635	0.417941
AUTOARIMA	261.029600	171.479200	0.321148	0.321776	0.002438	0.607449
ARIMAGLS	175.318100	96.639530	0.177854	0.178206	0.001637	0.417738
ARIMAML- ARIMAGLS	175.194800	96.487780	0.177686	0.178043	0.001636	0.417820

Models employed	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Panel O: Togo						
ARIMAML	74.972790	42.004770	0.243453	0.243769	0.002435	0.577294
AUTOARIMA	115.421900	78.775500	0.483778	0.485943	0.003758	0.951941
ARIMAGLS	74.893850	41.928960	0.242980	0.243283	0.002433	0.576363
ARIMAML-	74.932120	41.966860	0.243216	0.243526	0.002434	0.576819
ARIMAGLS						

Figures 13 and 14 present the forecast comparison graphs for every West African country for the training and testing samples respectively. It is observable that the row graphs for the definite as well as forecast indices of the cumulative cases are scarcely distinguishable. This implies the validity of using ARIMA in forecasting the CCC.

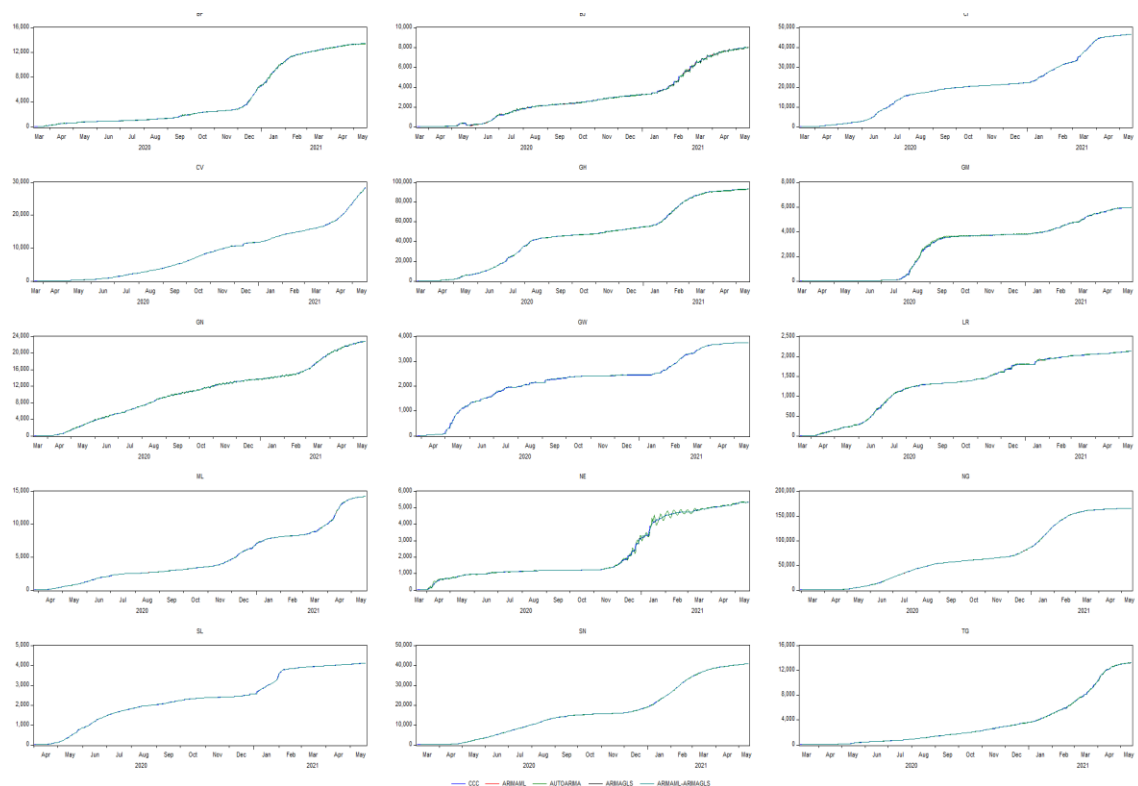


Figure 13: Forecast comparison graph for the training sample

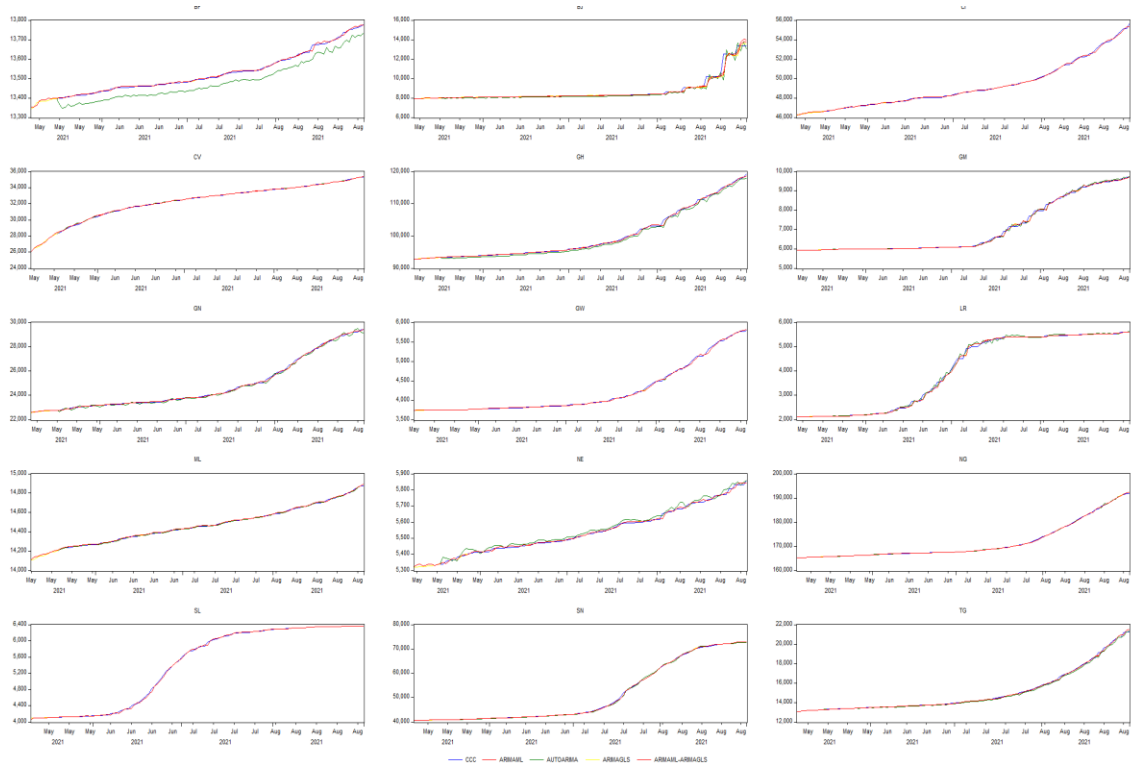


Figure 14: Forecast comparison graph for the testing sample

Figures 15 and 16 present the Taylor diagrams for the training and testing and samples respectively. Largest dot shows the fittest model in regards to RMSE. As shown in Figure 15, ARIMAML-ARIMAGLS happens to have the highest predictive accurateness for 7 nations, ARIMA with Generalized Least Squares for 4 nations and ARIMA with Maximum Likelihood for 3 nations. Figure 16, on the other hand, shows that ARIMAGLS beats other models in forecasting accuracy for eight countries, ARIMAML for six countries. In short, Figures 15 and 16 visually summarize Tables 5 and 6 respectively.

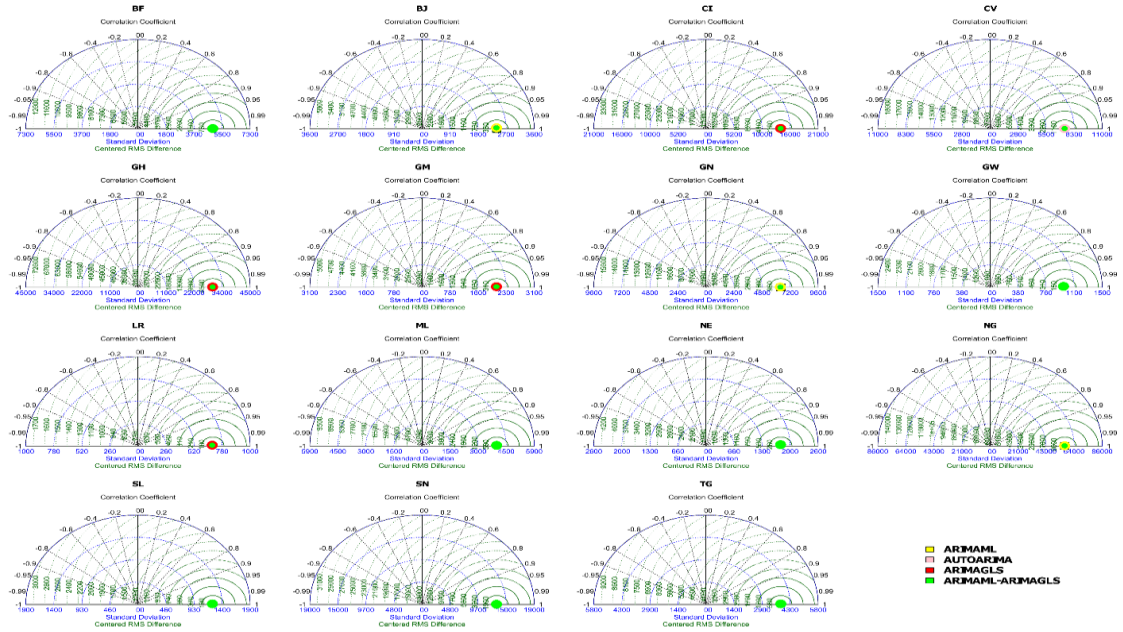


Figure 15: Taylor illustration for the training sample

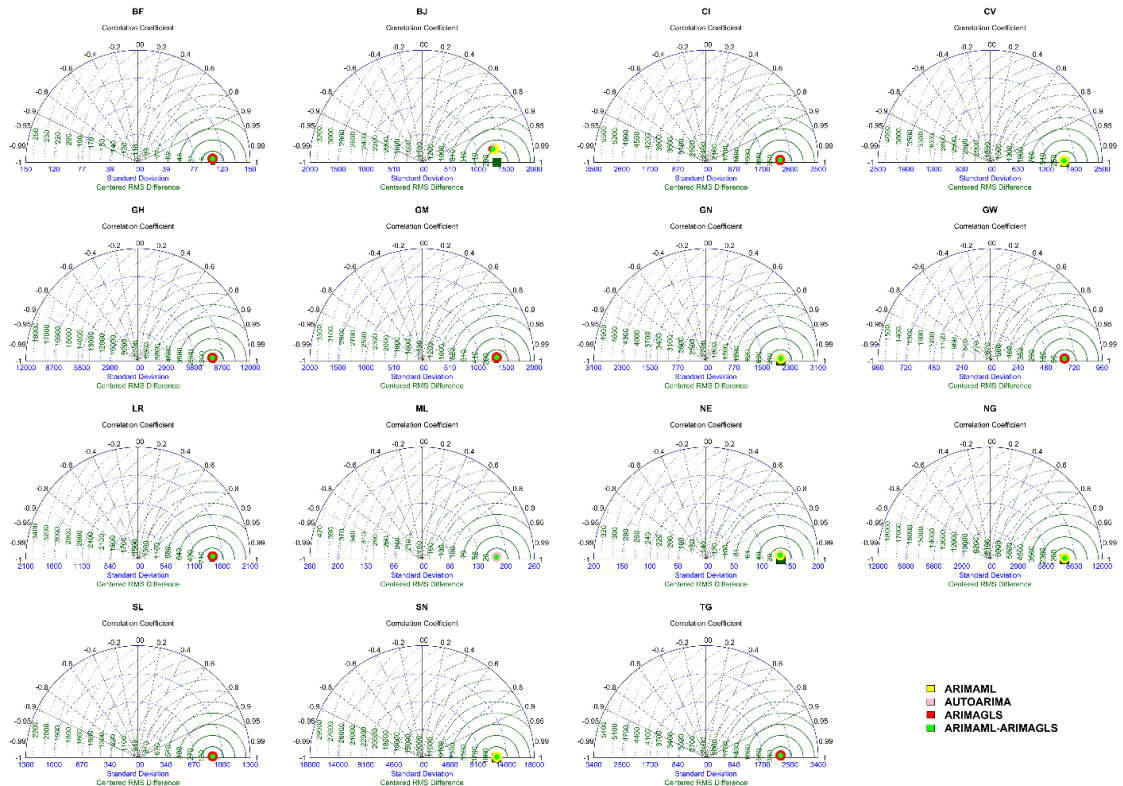


Figure 16: Taylor diagram for the testing sample

Figures 17 and 18 depict the fan diagram for both the training data set and testing one respectively. The figure or shape reports the RMSE of the 4 models. Therefore, the smaller the fan plot's sector angles the better model predictive performance of the area of the fan plot, the better the predictive performance of the model. The two figures conform with the information depicted by Figures 17 and 18.

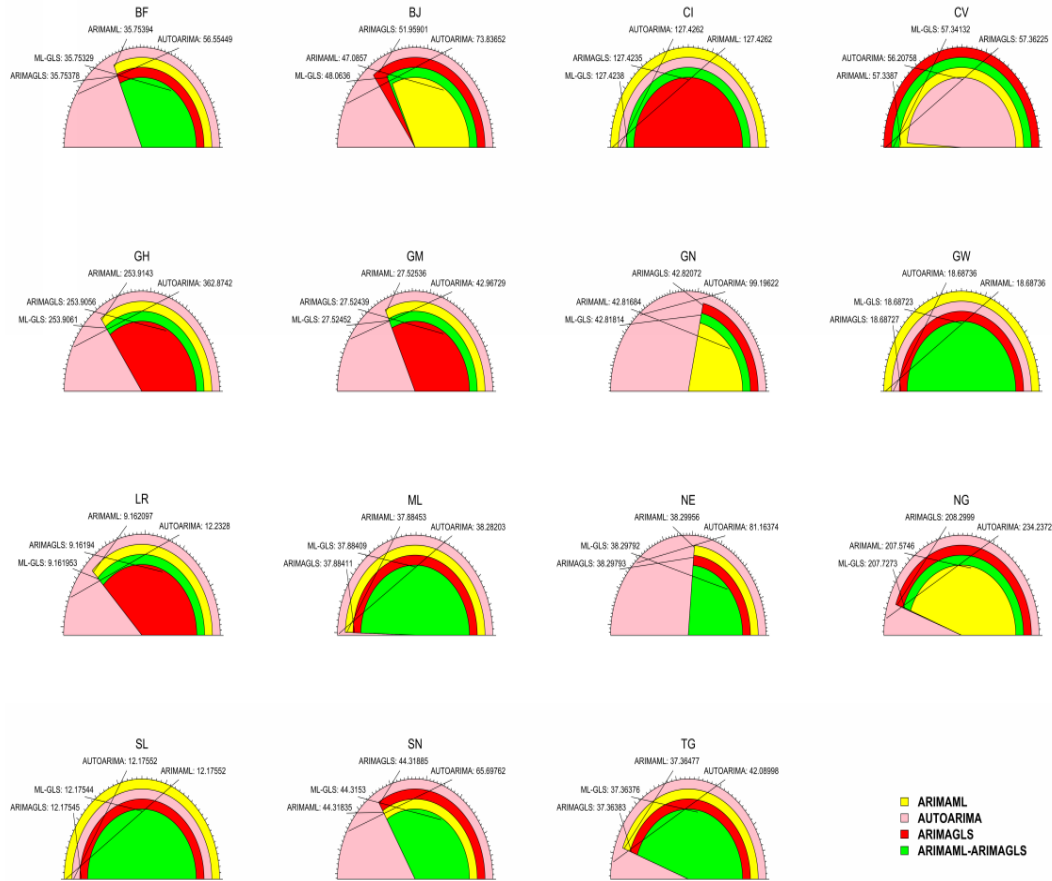


Figure 17: Fan plot for the training sample

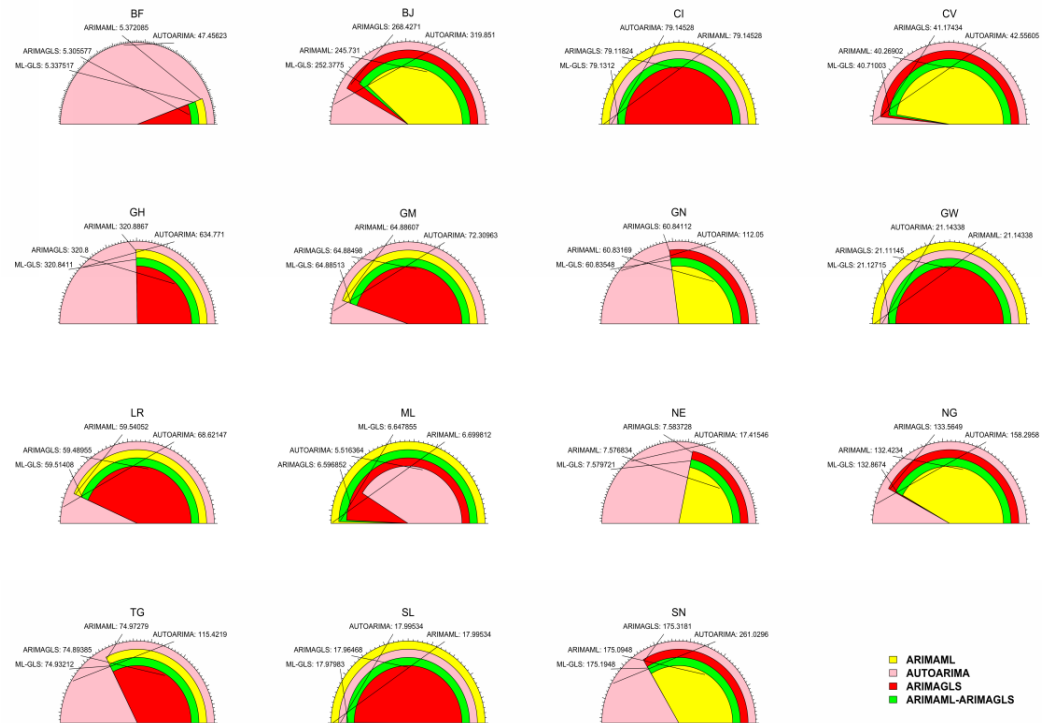


Figure 18: Fan scheme of the RMSE for the testing sample

The foregoing analysis has evaluated the best four forecasting performance models presented in this research. Based on the individual forecast statistics, it is observed that ARIMAGLS and ARIMAML are the best performing models, while ensembling the two models tend to produce the best forecast accuracy. The ARIMAGLS model outperforms other models because GLS produces discrepant empirical suitability and parameter estimations. Finally, assembling the ARIMAML and ARIMAGLS produce best forecasting accuracy in most cases in comparison to individual models due to the suitability of combining forecasts.

4.2 Phase II: Results and Discussions for the Machine Learning Based Modeling.

4.2.1 Results of the separate models:

Numerous time lags were employed so as to achieve the Markov strong point of the earlier incidences with regards to the present issue. It is revealed that up to seven duration lags ($d - 7$), solid association occurs amid present and preceding cases. Meanwhile, preceding incidences up to seven-day duration are very important to the present incidence of coronavirus disease. Therefore, in regards to the prediction of COVID-19 incidences in African states, the subsequent were employed as the inputs;

$$DC^i = f(DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}, DC^i_{(d-6)}, DC^i_{(d-7)}) \dots \dots \dots$$

(15)

i stands the African nation under investigation, DC represents daily incidences, $DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}, DC^i_{(d-6)}, DC^i_{(d-7)}$ are the i th state epidemic dataset at preceding duration stages $d-1, d-2, d-3, d-4, d-5, d-6$ and $d-7$ (or 1, 2, 3, 4, 5, 6 and 7 days ago).

Selecting the best dominating inputs is one of the utmost important components of all ML-based estimate; failing to do so may result in errors and inaccurate outcomes (Abdullahi et al., 2019b; Elkiran et al., 2021). As a result, through trials and errors, the greatest input variables that represented the COVID-19 output's most sensitive inputs were chosen for each nation, as displayed in Table seven.

Table 7. The Input variables employed

	Nation	Inputs
North Africa	Morocco	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}$
	Sudan	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}$
South Africa	Namibia	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}$
	South Africa	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}$
East Africa	Uganda	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}$
	Rwanda	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}, DC^i_{(d-6)}$
West Africa	Nigeria	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}, DC^i_{(d-6)}$
	Senegal	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}$
Central Africa	Gabon	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}, DC^i_{(d-6)}$
	Cameroon	$DC^i_{(d-1)}, DC^i_{(d-2)}, DC^i_{(d-3)}, DC^i_{(d-4)}, DC^i_{(d-5)}, DC^i_{(d-6)}$

For artificial neural network models, 3 FFNN techniques were employed in the finding that comprised input layer and hidden layer as well as output layer respectively. The artificial neural network algorithms were simulated by applying Levenberg Marquardt (algorithm) whereas the adaptation learning functions utilized was LEARNGDM and MSE was applied as the assessment function. This research evaluated numerous transfer functions so as to attain the suitable outcomes comprising of $(f(x) = \tanh(x)), (f(x) = 1/(1 + \exp(-x)), (f(x) = \text{sech}(x))$ as well as Gaussian $(f(x) = e^{-x.x})$. Learning

proportion employed was 0.01 and the epoch figure varied from one hundred to three hundred.

For support vector machine approach, Gaussian was the kernel function employed. The benefit of applying it is that; it makes the simulation simple in complex non-linear.

MLR's model finds the associations of linear amid output variable and input variables also employed to match their predictive effectiveness with the machine learning techniques. Tables 8-12 provides the outcomes of the entire established models.

Four statistical performance indicators used in this research which determine the predictive efficacy of the used models in Africa. The error measures containing MSE, RMSE, and MAD possessed no units because the data have been normalized, while the goodness of fitting measure of R^2 has no dimension.

Table 8. Outcomes of the single models for Northern African states

Nation	Model	Training				Validation			
		MAD	MSE	RMSE	R^2	MAD	MSE	RMSE	R^2
Moro	ANN	0.0699	0.0122	0.1106	0.8486	0.0336	0.0019	0.0436	0.8302
	ANFIS	0.0321	0.0035	0.0591	0.9515	0.0204	0.0011	0.0326	0.9154
	SVM	0.0423	0.0059	0.0767	0.9347	0.0185	0.0008	0.0287	0.9185
	MLR	0.0604	0.0115	0.0107	0.9078	0.0208	0.0001	0.0341	0.8405
Sudan	ANN	0.0369	0.0032	0.0564	0.6000	0.0353	0.0028	0.0564	0.4854
	ANFIS	0.0306	0.0029	0.0536	0.8315	0.0213	0.0012	0.0345	0.5343
	SVM	0.0299	0.0041	0.0642	0.5929	0.0242	0.0029	0.0537	0.3330
	MLR	0.0374	0.0048	0.0691	0.3248	0.0443	0.0055	0.0740	0.1135

Table 9. Findings of the employed models for Eastern African countries

Nation	Model	Training				Validation			
		MAD	MSE	RMSE	R^2	MAD	MSE	RMSE	R^2
Uganda	ANN	0.0078	0.0002	0.0126	0.3552	0.0141	0.0060	0.0777	-0.0025
	ANFIS	0.0064	0.0001	0.0113	0.4807	0.0181	0.0056	0.0750	0.0650
	SVM	0.0036	0.0001	0.0090	0.6682	0.0080	0.0060	0.0774	0.0048
	MLR	0.0081	0.0002	0.0134	0.2712	0.0144	0.0060	0.0775	0.0015
Rwanda	ANN	0.0319	0.0072	0.0851	0.9232	0.0112	0.0003	0.0182	0.7012
	ANFIS	0.0233	0.0023	0.0478	0.9205	0.0106	0.0003	0.0185	0.9059
	SVM	0.0432	0.0110	0.1006	0.5824	0.0150	0.0020	0.0446	0.5366
	MLR	0.0485	0.0094	0.0970	0.6417	0.0166	0.0015	0.0392	0.6119

Table 10. Outcomes of the employed models for Western African countries

Nation	Model	Training				Validation			
		MAD	MSE	RMSE	R ²	MAD	MSE	RMSE	R ²
Nigeria	ANN	0.0374	0.0035	0.0592	0.7964	0.0295	0.0031	0.0560	0.7206
	ANFIS	0.0330	0.0029	0.0537	0.8958	0.0247	0.0016	0.0400	0.7699
	SVM	0.0387	0.0071	0.0844	0.7947	0.0288	0.0032	0.0562	0.4316
	MLR	0.0514	0.0088	0.0936	0.7312	0.0351	0.0041	0.0643	0.3015
Senegal	ANN	0.0284	0.0055	0.0737	0.8334	0.0195	0.0009	0.0308	0.6285
	ANFIS	0.0189	0.0017	0.0408	0.9492	0.0187	0.0008	0.0287	0.6765
	SVM	0.0335	0.0063	0.0791	0.8089	0.0185	0.0009	0.0292	0.6563
	MLR	0.0363	0.0061	0.0783	0.8128	0.0187	0.0009	0.0292	0.6647

Table 11. Outcomes of the employed models for the Southern African countries

Nation	Model	Training				Validation			
		MAD	MSE	RMSE	R ²	MAD	MSE	RMSE	R ²
Namibia	ANN	0.0505	0.0076	0.0870	0.5786	0.0394	0.0044	0.0667	0.3827
	ANFIS	0.0255	0.0024	0.0488	0.8884	0.0183	0.0012	0.0343	0.8059
	SVM	0.0406	0.0093	0.0965	0.5297	0.0255	0.0050	0.0705	0.2400
	MLR	0.0477	0.0091	0.0955	0.5467	0.0259	0.0048	0.0692	0.2556
South Africa	ANN	0.0351	0.0050	0.0706	0.8924	0.0224	0.0018	0.0427	0.8553
	ANFIS	0.0364	0.0040	0.0630	0.9355	0.0195	0.0011	0.0331	0.8846
	SVM	0.0375	0.0063	0.0796	0.9110	0.0209	0.0015	0.0388	0.8160
	MLR	0.0523	0.0096	0.0977	0.8689	0.0240	0.0022	0.0471	0.7225

Table 12. Outcomes of the employed models for C. African countries

Nation	Model	Training				Validation			
		MAD	MSE	RMSE	R ²	MAD	MSE	RMSE	R ²
Gabon	ANN	0.0567	0.0106	0.1006	0.8575	0.0447	0.0079	0.0888	0.5866
	ANFIS	0.0490	0.0077	0.0878	0.9007	0.0411	0.0055	0.0741	0.6983
	SVM	0.0490	0.0120	0.1097	0.8142	0.0441	0.0103	0.1014	0.5289
	MLR	0.0835	0.0209	0.1193	0.6228	0.0700	0.0142	0.1445	0.4429
Cameroon	ANN	0.0037	0.0006	0.0249	0.9042	0.0085	0.0022	0.0465	0.6531
	ANFIS	0.0097	0.0005	0.0220	0.9225	0.0080	0.0012	0.0341	0.8200
	SVM	0.0089	0.0011	0.0339	0.8230	0.0104	0.0013	0.0354	0.7987
	MLR	0.0091	0.0011	0.0338	0.8231	0.0114	0.0012	0.0349	0.8041

As observed from Table 8 (for Sudan and Morocco), diverse models lead to various results for Sudan and Morocco within the validation and training stages, correspondingly. Regarding the validation stage, it is illustrated that in Morocco, the entire employed models possessed R^2 value above 0.7, and it is a sign of the simulation's precision. In spite of the encouraging results of the employed models, support vector machine (SVM) displays great accuracy having lowest errors and robust fit with Mean Absolute Deviation = 0.0185, Mean Square Error = 0.0008, Root Mean Square Error = 0.0287 and $R^2 = 0.9185$. This is accompanied tightly by Adaptive Neuro Fuzzy Inference System Having Mean Absolute Deviation = 0.0204, Mean Square Error = 0.0011, Root Mean Square Error = 0.0326 and $R^2 = 0.9154$. For the Republic of Sudan, it is observed that the models with the improved performance is ANFIS with MAD = 0.0213, MSE = 0.0012, RMSE = 0.0345 and $R^2 = 0.5343$.

According to Table 9 outcomes for Eastern African states, demonstrate the low predictive efficacy by the applied models in Uganda, the model with maximum performance in the validation stage is adaptive neuro-fuzzy inference system having Mean Absolute Deviation = 0.0181, Mean Square Error = 0.0056, Mean Square Error = 0.0750 and $R^2 = 0.0650$. In spite of the drawdown in the modelling efficacy, it can be seen that ANFIS and ANN have considerable predictive precision beyond 0.7 R^2 value, Adaptive Neuro Fuzzy Inference System formed the utmost proficient outcomes with MAD = 0.0106, MSE = 0.0003, RMSE = 0.0185 and $R^2 = 0.9059$.

Outcomes for the countries in Western Africa are displayed in Table 10. The applied model's performance shows that artificial intelligence (AI) models have the capability of predicting daily incidences in Nigeria, while MLR's models could also be applied.

Table 11 demonstrated the results of daily-affirmed incidences prediction by the 4 employed models for South African countries. In regard to Namibia's outcomes in the stage of validation, less performance are established by the entire models with exclusion of adaptive neuro fuzzy inference system, which possessed Mean Absolute Deviation = 0.0183, Mean Square Error= 0.0012, Root Mean Square Error = 0.0343 and $R^2 = 0.8059$. Regarding the outcomes of Southern Africa illustrated by Table 11 in the stage of validation, it can be observed that the entire models possessed greater prediction efficacy. ANFIS has the paramount modelling proficiency having Mean Absolute Deviation =

0.0195, Mean Square Error = 0.0011, Root Mean Square Error = 0.0331 and $R^2 = 0.8846$. Moreover, the subsequent utmost effective model is Artificial Neural Network, followed by Support Vector Machine and traditional Multi Linear Regression model is the slightest in respect to predictive.

In respect to the outcomes of the Central African nations comprising of Cameroon and Gabon are demonstrated by Table 12. According to the validation step results for Gabon, ANFIS produced the maximum precision having Mean Absolute Deviation of = 0.0411, Mean Square Error = 0.0055, Root Mean Square Error = 0.0741 and $R^2 = 0.6983$, followed by ANN with MAD = 0.0447, MSE = 0.0079, RMSE = 0.0888 and $R^2 = 0.5866$.

Regarding Cameroon in Table 12, Adaptive Neuro Fuzzy Inference System also demonstrated good performance with Mean Absolute Deviation = 0.0080, Mean Square Error = 0.0012, Root Mean Square Error = 0.0341 and $R^2 = 0.8200$. In spite of linearity of MLR models, it still created consistent performance when matched with SVM and ANN. The MLR's projective ability is essentially not inexplicable.

Individual model performance can be matched and measured graphical by the Figure 2 by means of radar graph. Moreover, the radar diagram possessed the capability to gather quite a lot of models into one chart for stress-free as well as easy matching. Regarding the R^2 , the broader the inner outlines are, the greater the accuracy of the individual models and the smaller the internal line the lower the predictive precision.

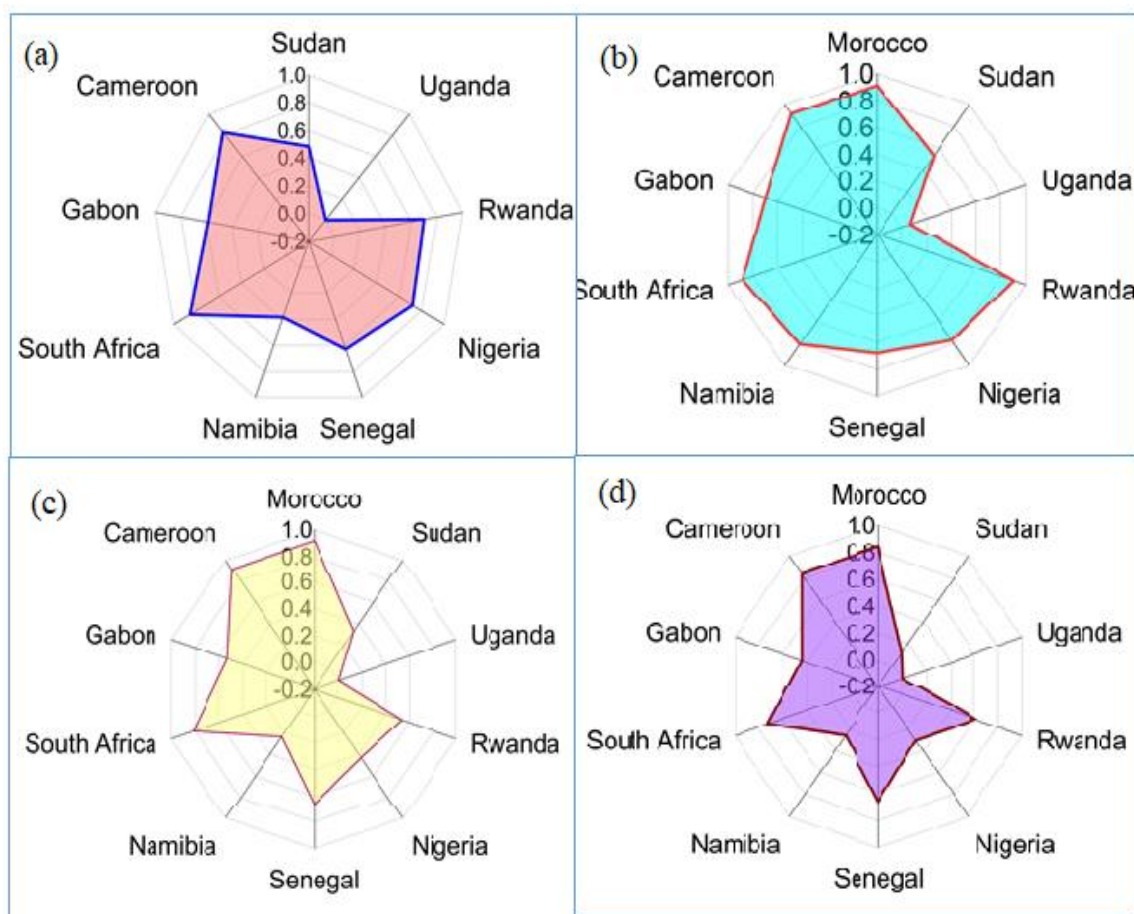
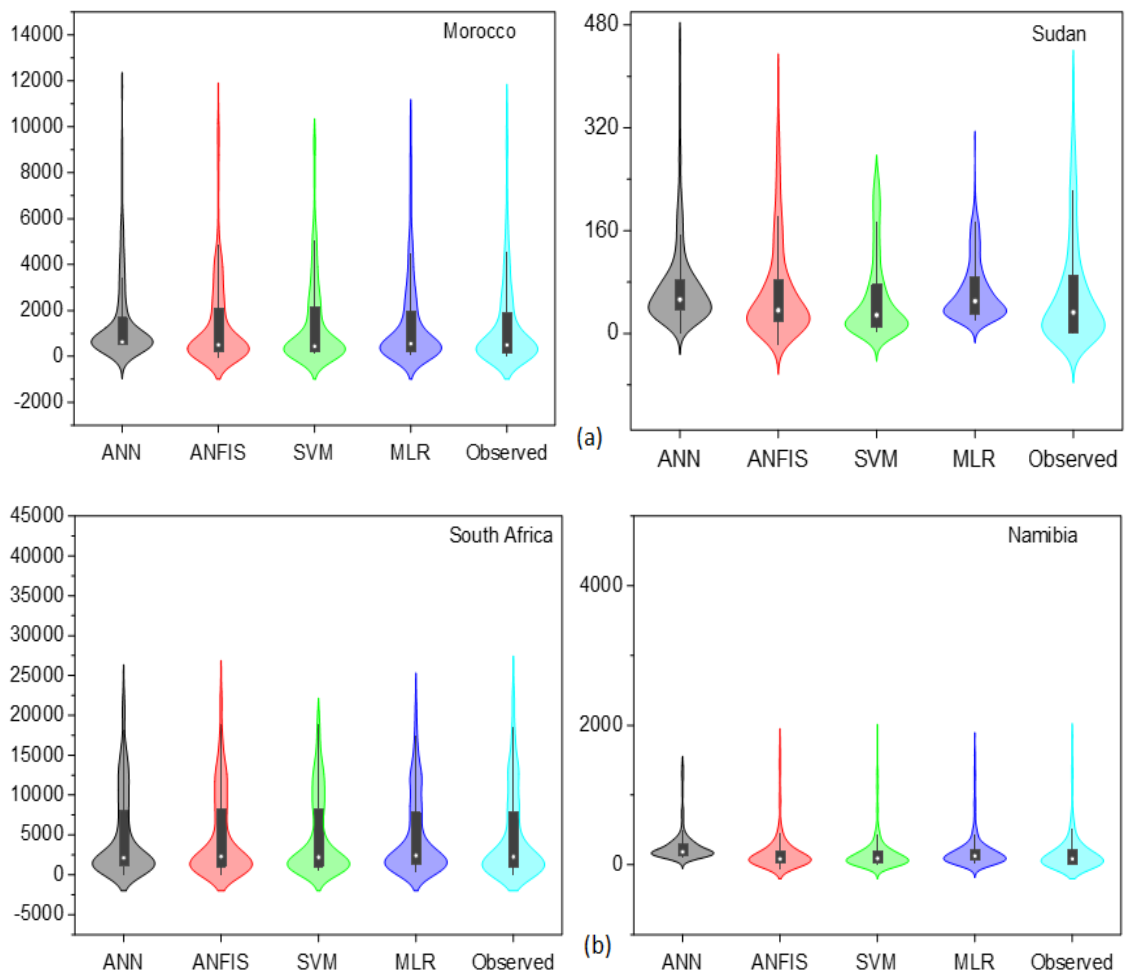


Figure 19. Comparison of individual model performance on R2 (1) ANN model, (2) ANFIS model, (3) SVM model (4) MLR model

For artificial neural network model (Figure 19 (1-4)), with encouraging outcomes for everyday coronavirus disease 2019 test, ANN is the best for Morocco and South Africa. Nevertheless, it is crucial to recognize that the quantity of daily verified COVID-19 instances may not always be relevant to the precision and effectiveness of the predictive algorithms. Authorities' strict preventive measures, like lockdowns, people isolation, applying sanitizers, etc., are crucial to the detection of cases and effective modelling. For instance, Cameroon has fewer cases than Nigeria and numerous other nations combined. However, the steps made by the Cameroonian government to curb the impact and feast of the highly contagious virus make it simpler to disentangle the unknowns connected with the COVID-19.

One can see that the performance of the models in terms of precision is alike across the selected nations by looking at the model performance in Figure 19(1-4). South Africa and Morocco are the two countries where the models are most accurate, accompanied by Republic of Cameroon, Republic of Nigeria, Rwanda, Senegal, Gabon, Republic of Sudan, Namibia, as well as Uganda. Furthermore, by looking at Figure 22(a-d), it is clear that ANFIS performed better in virtually all states. This is because it effectively chains the benefits of fuzzy's logic with NN.

When comparing performance of the models between Tables 8-12 and Figure 23, it can be concluded that Morocco has the highest level of model accuracy. Figure 22 provides time series plots for Morocco from 05/07/2021 to 12/12/2021.



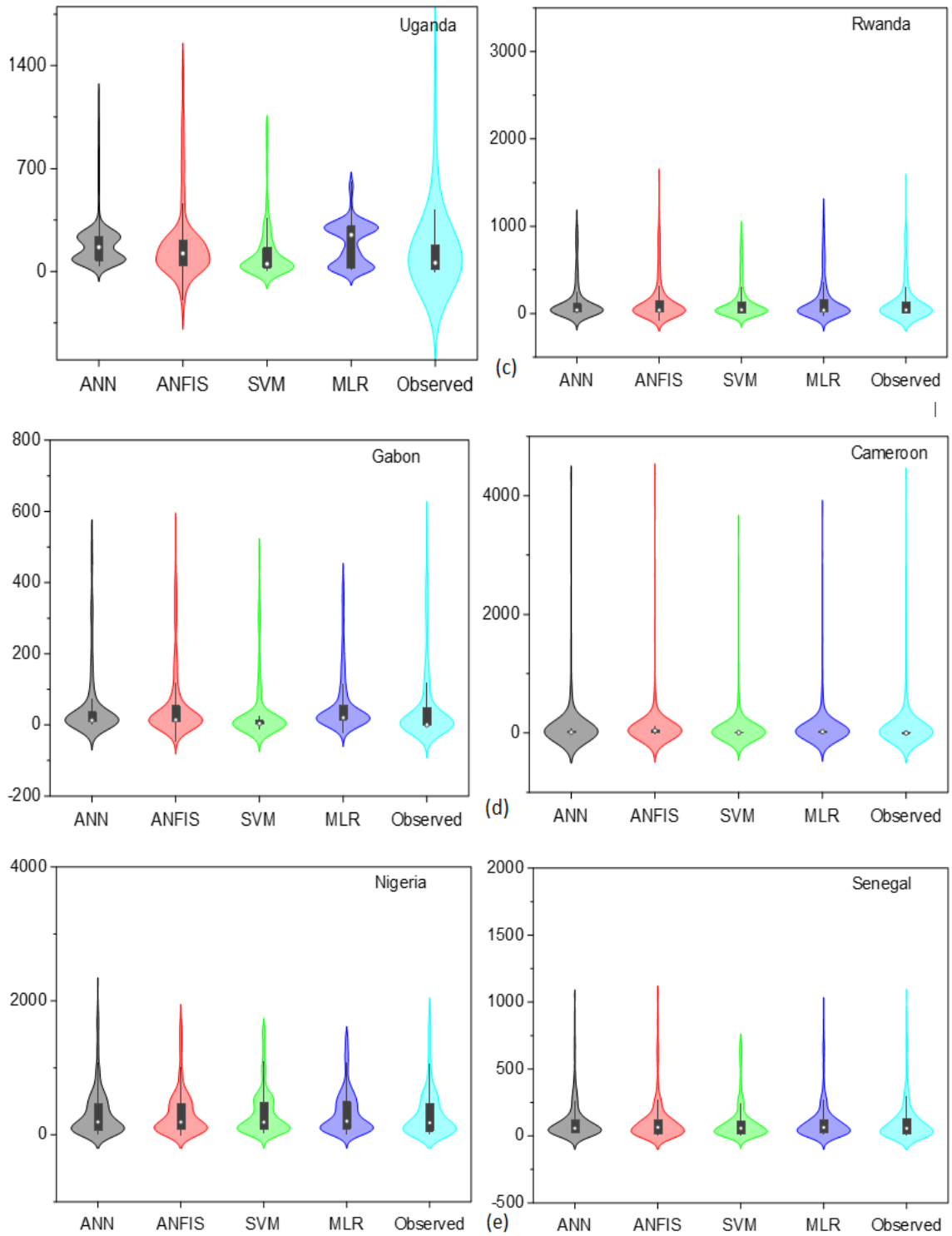


Figure 20: Statistical performance of developed models versus observed values for (a) Northern African nations, (b) Southern African nations, (c) Eastern African nations, (d) Central African nations (e) Western African states

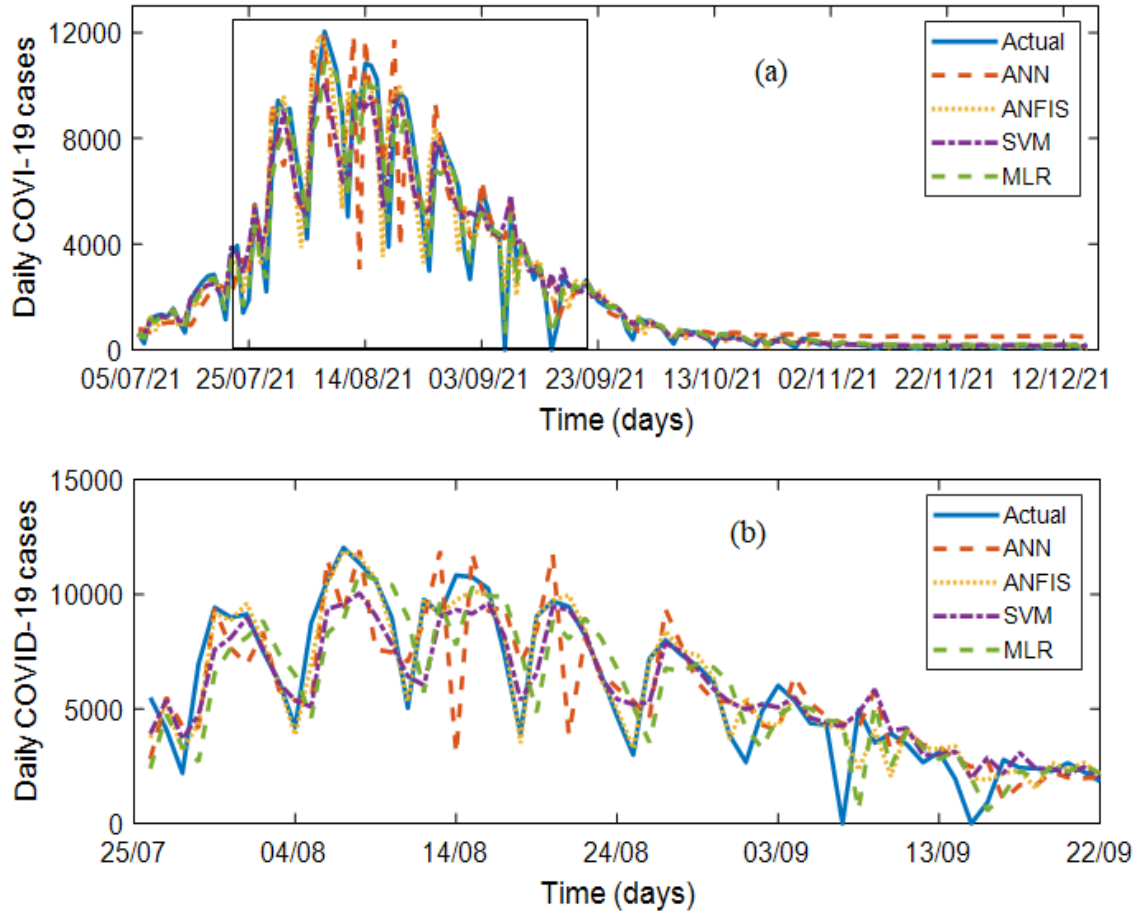


Figure 21. Matching between projected daily-affirmed cases of COVID-19 and observed incidences within the stage of validation for (a) broad set of the data (b) Zoom view

4.2.2 Ensemble Approaches Results

Figure 23 (a) shows how each model performed for Morocco. The Figure shows that every model, in general, follows the inclination of the observed dataset. Due to the oscillations of huge values, near observation of expected values could not be noticed visibly. As a result, Figure 23 (b) is shown with values zoomed for accuracy in comparing anticipated values to actual data values. Despite the fact that Morocco outperformed all other nations in terms of model performance, there is still space for improvement as a closer examination reveals large discrepancies between anticipated as well as observed values. Consequently, ensemble approaches according to SVM-E and ANN-E are applied to increase the modeling precision. The outcomes of the ensemble models are illustrated in Tables 13 and 14.

Table 13. Outcomes of the employed novel ensemble approach (ANN-E)

Section	Country	Training				Validation			
		MAD	MSE	RMSE	R ²	MAD	MSE	RMSE	R ²
Northern									
Africa	Morocco	0.0289	0.0025	0.0500	0.9653	0.0206	0.0009	0.0300	0.9282
	Sudan	0.0210	0.0012	0.0343	0.8338	0.0265	0.0023	0.0478	0.6299
Southern									
Africa	Namibia	0.0277	0.0025	0.0497	0.8968	0.0204	0.0011	0.0330	0.7980
	South Africa	0.0276	0.0027	0.0519	0.9366	0.0203	0.0011	0.0328	0.9219
Eastern									
Africa	Uganda	0.0032	0.0000	0.0050	0.9958	0.0031	0.0000	0.0064	0.8314
	Rwanda	0.0265	0.0047	0.0686	0.9207	0.0111	0.0003	0.0185	0.8059
Western									
Africa	Nigeria	0.0248	0.0014	0.0368	0.9118	0.0315	0.0026	0.0510	0.7926
	Senegal	0.0168	0.0016	0.0394	0.9526	0.0185	0.0008	0.0288	0.6756
Central African states									
	Gabon	0.0485	0.0088	0.0939	0.8954	0.0404	0.0058	0.0761	0.6550
	Cameroon	0.0042	0.0002	0.0145	0.9674	0.0073	0.0002	0.0155	0.9616

Table 14. SVM-E outcomes

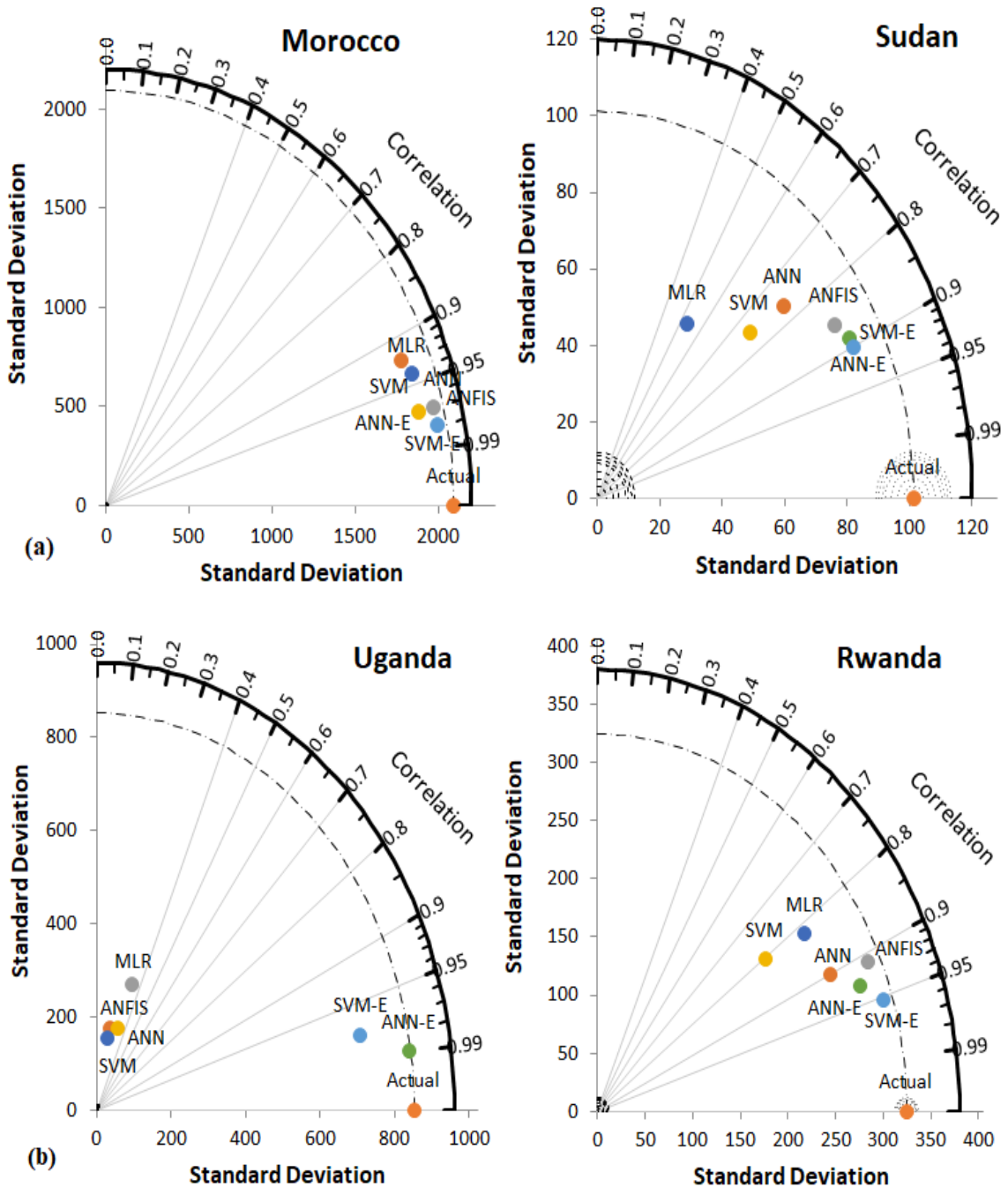
Section	Nation	Training				Validation			
		MAD	MSE	RMSE	R ²	MAD	MSE	RMSE	R ²
North									
Africa	Morocco	0.0299	0.0028	0.0527	0.9615	0.0202	0.0009	0.0302	0.9276
	Sudan	0.0251	0.0024	0.0488	0.8408	0.0208	0.0011	0.0336	0.6142
Southern									
Africa	Namibia	0.0277	0.0025	0.0498	0.8978	0.0181	0.0011	0.0328	0.7980
	South Africa	0.0289	0.0028	0.0531	0.9383	0.0206	0.0010	0.0323	0.9180
Eastern									
Africa	Uganda	0.0058	0.0002	0.0146	0.9645	0.0040	0.0001	0.0087	0.6943
	Rwanda	0.0295	0.0020	0.0452	0.9157	0.0204	0.0008	0.0283	0.8139
Western									
Africa	Nigeria	0.0331	0.0027	0.0520	0.8930	0.0286	0.0016	0.0406	0.7845
	Senegal	0.0205	0.0022	0.0466	0.9337	0.0193	0.0009	0.0293	0.6629
Central									
Africa	Gabon	0.0506	0.0072	0.7175	0.8999	0.0429	0.0055	0.0849	0.7175
	Cameroon	0.0083	0.0002	0.0133	0.9728	0.0097	0.0004	0.0189	0.9429

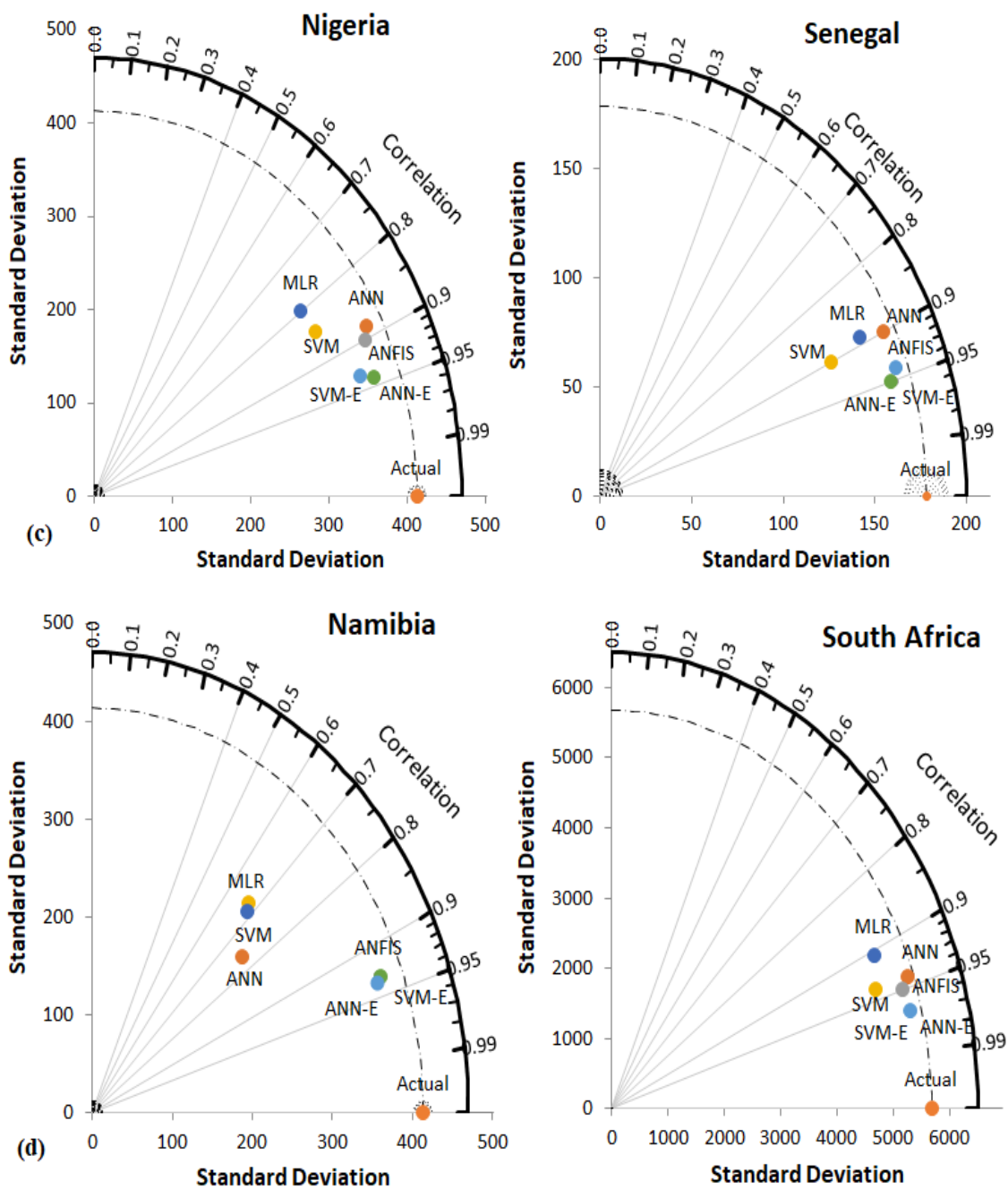
Ensemble approaches outcomes demonstrate significant enhancement having minimal errors and elevated R2 values that are typically greater than 0.9. When ANN ensemble (Table thirteen) outcomes are compared to those from single models (Tables 8–12), it is clear that performance has improved significantly. In the validation step, the ANN ensemble increased the predictive precision of the individual models of Morocco, Sudan republic, Namibian state, South Africa, Ugandan state, republic of Rwanda, republic of Nigeria, Senegal, Gabon, and Cameroon state, correspondingly, by 10%, 14%, 42%, 6%, 83%, 11%, 7%, 5%, 7%, and 31%.

With reduced performance from the individual models and the ensemble approaches fared better, according to the results. For instance, Uganda, the nation with the lowest performance of the individual models, saw the biggest increase in model performance, amounting to 83%. On the other side, it was discovered that countries with the highest single modeling accuracy experienced the least increase in their efficiency as a result of ensemble approaches. For instance, South Africa and Morocco were shown to have efficiency improvements of just 10% and 6% despite having the highest successful modeling of the daily-incidences of the deadly virus COVID-19 by single models. This means that single models with weak or bad performance would leave a large space to improve prediction, whereas single models with high performance would leave a small gap to boost modeling performance. Nevertheless, the findings of this work demonstrate that, for daily-affirmed cases in African nations, the ensemble model not only significantly outperforms weak performance single models, but also significantly outperforms models (single) with elevated. For example, Cameroon, one of the nations with the greatest individual model performance, has seen a 31% improvement in performance.

Tables 13 and 14 can be compared to show that ensemble models perform similarly, which may be because they both combined single models using a similar manner. Tables 13 and 14 make it clear that there is little performance difference between ANN ensemble as well as SVM ensemble. Support Vector Machine Ensemble has shown to be more accurate than ANN-E for some nations, such as Rwanda and Gabon, while ANN-E edged slightly higher for others, such as Morocco and Sudan.

It is indicated that, no improved algorithm regarding the ensemble modeling, because any ensemble (kernel) might result in great improvements. The outcomes of the entire ensemble and single and models are matched using by Taylor diagram (Figure 23).





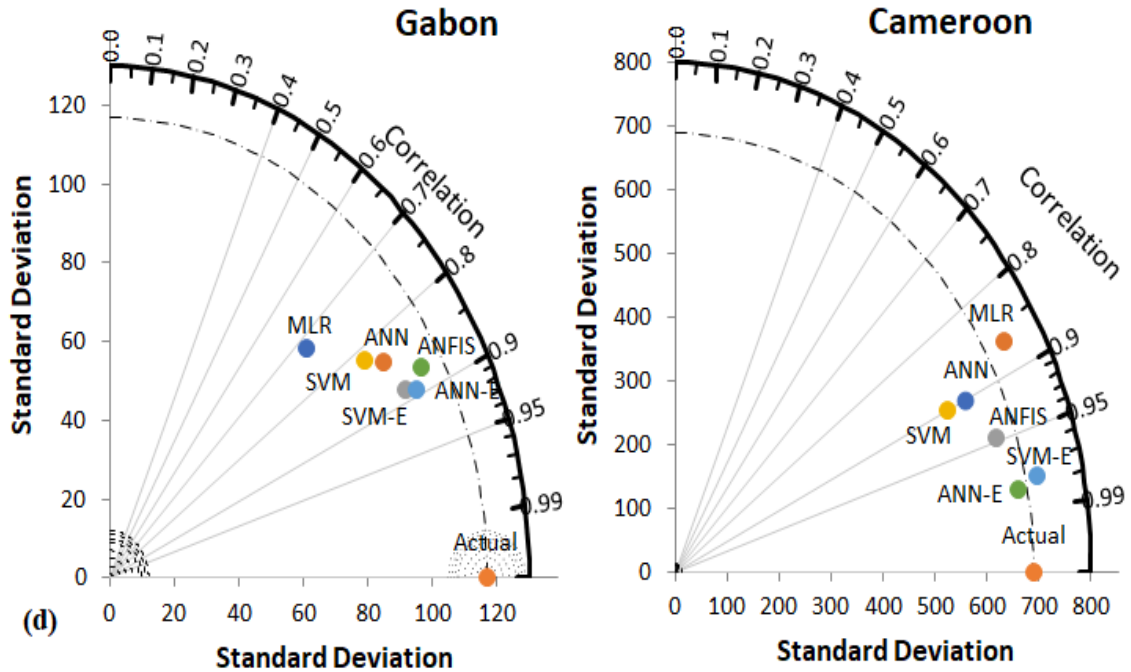


Figure 22. Matching of the models performance for (a) Northern African states, (b) Eastern African countries, (c) Western African countries, (d) South African countries and (e) Central Africa states

In the diagram, correlation coefficients, Root Mean Square Error as well as Standard Deviation are employed to examine the resemblance amid observed records and projecting models. The observed set of data along the abscissa as a circle whereas based on the distance from the detected dataset, the other models displayed their performance.

As observed in Figure 23, according to the values of CC, the SVM-E and ANN-E possessed lesser values (near to one) that imply the maximum efficient and reliable daily-verified cases of COVID-19 modelling through all Nations. Is a sign that apart from the presented supremacy of ensemble procedures beyond individual models employed, which were according to the statistical indicators used; diagrammatically, ensemble approaches beat other models. Regarding Root Mean Square Error, it is detected that ensemble based approaches outcomes possessed lowest values of errors and hereafter, also result in utmost precise modelling. In regards to standard deviation, values very close to the real line indicate grater consistency. It might be detected that the novel ensemble procedures established and displayed improved modelling proficiency.

4.2.3 Results of the Cumulative Confirmed Incidences:

Table 15: Outcomes of the engaged models in Northern Africa

		Training				Validation			
Nation	Model	MAD	MSE	RMSE	R2	MAD	MSE	RMSE	R2
Morocco	ANN	0.0023	0.0003	0.0175	0.0042	0.0076	0.0009	0.0297	0.9621
	ANFIS	0.0026	0.0189	0.0435	0.9643	0.0104	0.0002	0.0141	0.9914
	SVM	0.0221	0.0026	0.0505	0.9518	0.0288	0.0010	0.0314	0.9575
	MLR	0.0039	0.0020	0.0448	0.9620	0.0050	0.0000	0.0056	0.9987
Sudan	ANN	0.0016	0.0000	0.0026	0.9999	0.0022	0.0000	0.0032	0.9960
	ANFIS	0.0051	0.0018	0.0428	0.9756	0.0090	0.0003	0.0161	0.9001
	SVM	0.0207	0.0024	0.0486	0.9687	0.0243	0.0006	0.0248	0.7651
	MLR	0.0038	0.0000	0.0045	0.9923	0.0046	0.0020	0.0450	0.9731

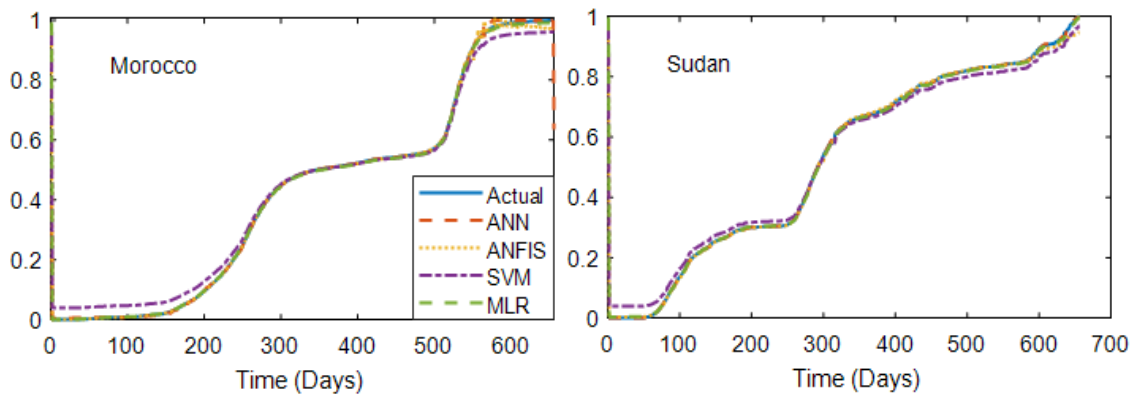


Figure 23: Time series plot for Northern African states

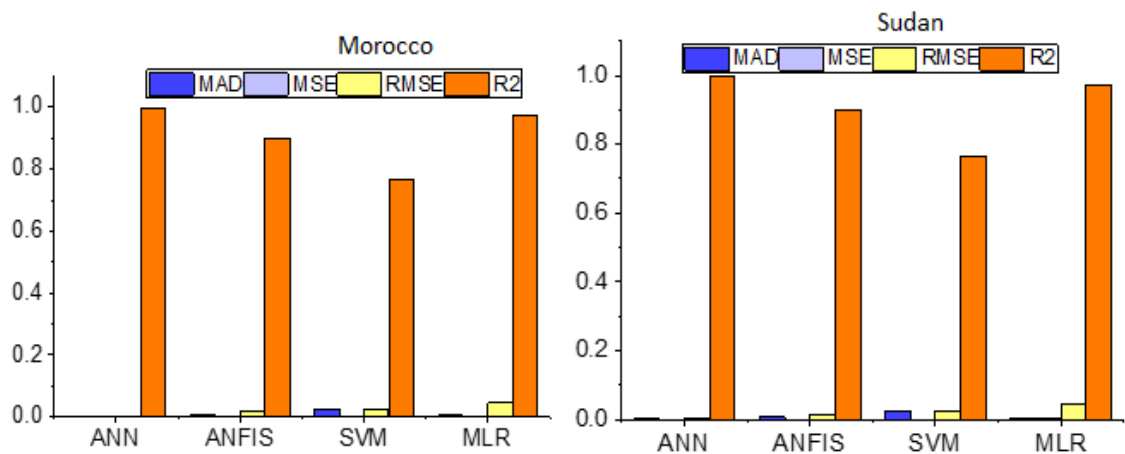


Figure 24: Bar chart illustrating the best performing models for North Africa

Table 15 displayed the outcomes of the used models' for north African states. Regarding the results of Morocco at the validation step, MLR outperform other models with $MAD = 0.0050$, $MSE = 0.0000$, $RMSE = 0.0056$, $R2 = 0.9987$, followed by ANFIS having $MAD = 0.0104$, $MSE = 0.0002$, $RMSE = 0.0141$ and $R2 = 0.9914$. Figure 26 demonstrates the time series in North African countries, whereas figure 27 shows the bar chart with the best performing model within North African nations based on the global statistical indices.

Table 16: Outcomes of the engaged models in Eastern Africa

		Training				Validation			
Nation	Model	MAD	MSE	RMSE	R2	MAD	MSE	RMSE	R2
Uganda	ANN	0.0007	0.0001	0.0009	0.9999	0.0169	0.0124	0.1113	0.4374
	ANFIS	0.0031	0.0018	0.0425	0.9359	0.0280	0.0125	0.1116	0.4340
	SVM	0.0116	0.0020	0.0442	0.9308	0.0570	0.0133	0.1155	0.3940
	MLR	0.0031	0.0020	0.0448	0.9290	0.0221	0.0123	0.1110	0.4402
Rwanda	ANN	0.0003	0.0001	0.0004	0.9999	0.0005	0.0001	0.0006	0.9998
	ANFIS	0.0022	0.0018	0.0426	0.8307	0.0136	0.0003	0.0178	0.9879
	SVM	0.0214	0.0023	0.0481	0.7836	0.0213	0.0005	0.0233	0.9792
	MLR	0.0028	0.0020	0.0447	0.8134	0.0066	0.0001	0.0073	0.9980

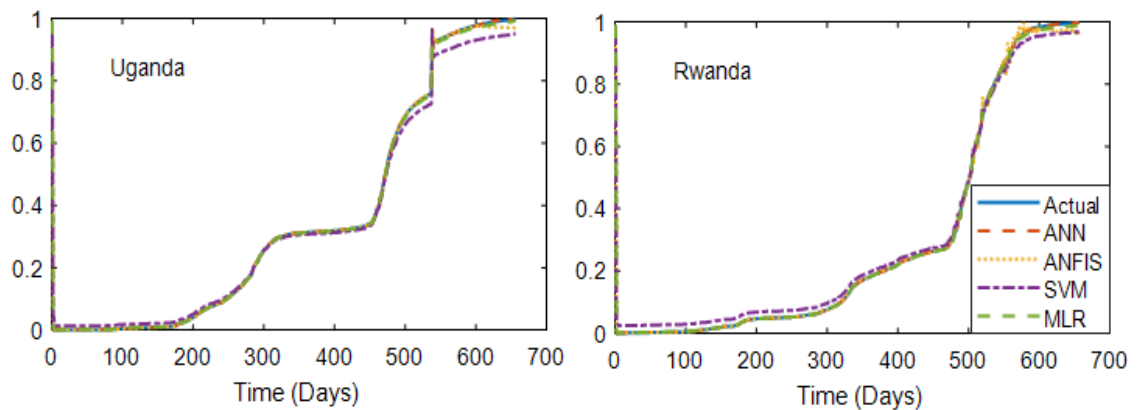


Figure 25: Time series for the cumulative incidences in East Africa

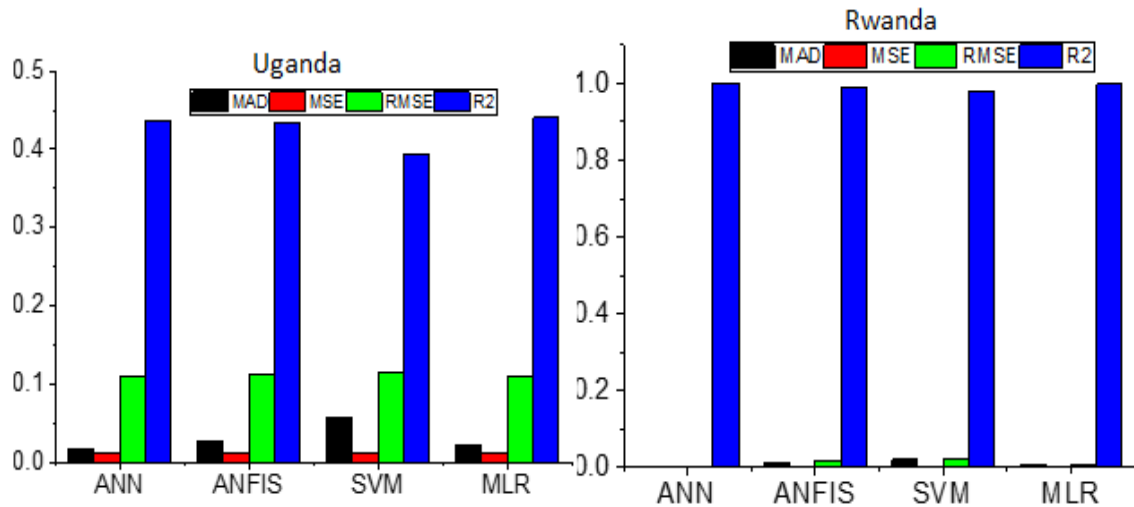


Figure 26: Bar chart illustrating the best performing models in East Africa Table 16, illustrates the finding of the applied models in East African countries. The models' performance in Uganda is poor at the stage of validation, all the models have R2 values less than 0.7. In respect to results for Rwanda, the models displayed a better performance at the validation step where ANN outperform other models with MAD = 0.0005, MSE = 0.0001, RMSE = 0.0006 and R2 = 0.9998 accompanied by MLR with MAD = 0.0066, MSE = 0,0001, RMSE = 0.0073, and R2 = 0.9980. Figure 27 shows the time series plot in East Africa. The models with better prediction accuracy is presented in figure 28 according to global statistical indicators.

Table 17: Outcomes of the engaged models in West Africa
West Africa

Nation	Model	Training				Validation			
		MAD	MSE	RMSE	R2	MAD	MSE	RMSE	R2
Nigeria	ANN	0.0003	0.0000	0.0038	0.9998	0.0002	0.0000	0.0004	1.0000
	ANFIS	0.0072	0.0007	0.0255	0.9917	0.0200	0.0011	0.0336	0.8019
	SVM	0.0251	0.0025	0.0500	0.9681	0.0337	0.0012	0.0343	0.7942
	MLR	0.0039	0.0020	0.0449	0.9743	0.0043	0.0000	0.0045	0.9965
Senegal	ANN	0.0029	0.0020	0.0444	0.9486	0.0091	0.0001	0.0107	0.9912
	ANFIS	0.0026	0.0019	0.0441	0.9493	0.0094	0.0001	0.0113	0.9902
	SVM	0.0133	0.0021	0.0454	0.9463	0.0256	0.0008	0.0275	0.9414
	MLR	0.0033	0.0020	0.0448	0.9476	0.0054	0.0000	0.0058	0.9974

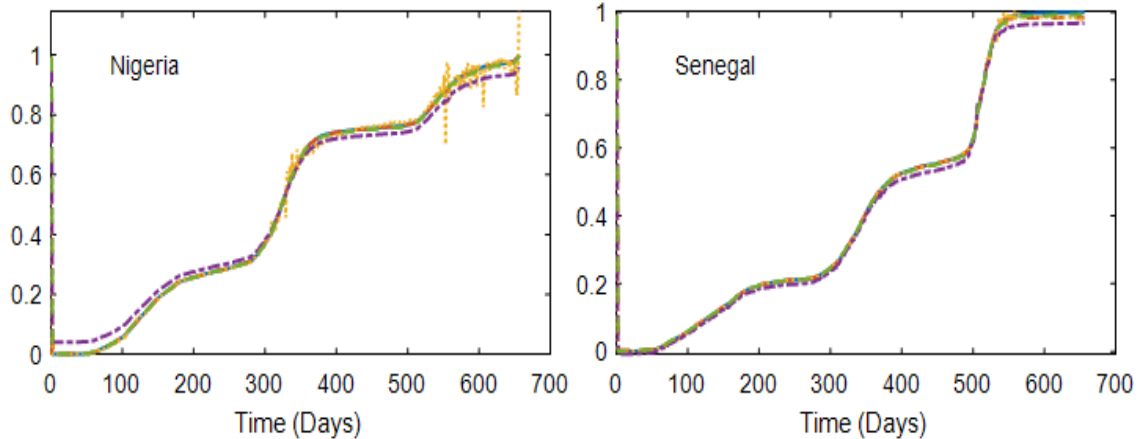


Figure 27: Time series plots for the cumulative incidences in West Africa

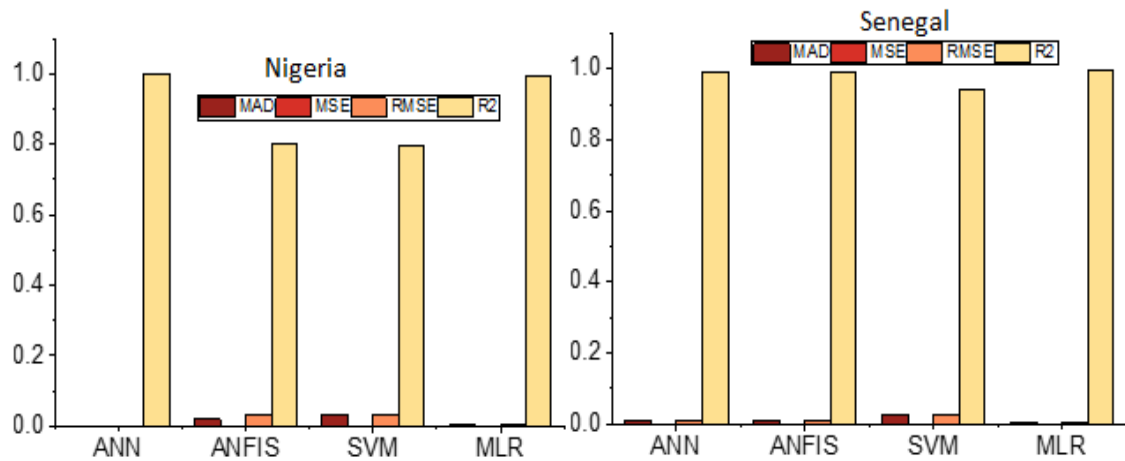


Figure 28: Bar chart demonstrating the best performing models in West Africa

Table 17 demonstrates the best performing model for West African countries. According to the outcomes, the model with better prediction accuracy in Nigeria is Artificial Neural Network having $MAD = 0.0002$, $MSE = 0.0000$, $RMSE = 0.0004$ and $R^2 = 1.000$ followed by MLR having $MAD = 0.0043$, $MSE = 0.0000$, $RMSE = 0.0045$, $R^2 = 0.9965$ at the stage of validation. Regarding the Senegal's result, MLR is the best model with $MAD = 0.0054$, $MSE = 0.0000$, $RMSE = 0.0058$ and $R^2 = 0.0074$. Figure 30 illustrates the best predictive models according to the values of R^2 , whereas figure 29 shows the time series plot for the cumulative incidences in West African states.

Table 18: Results of the applied models for South Africa
South Africa

Country	Model	Training				Validation			
		MAD	MSE	RMSE	R2	MAD	MSE	RMSE	R2
Namibia	ANN	0.0084	0.0023	0.0476	0.9197	0.0143	0.0003	0.0170	0.8961
	ANFIS	0.0091	0.0001	0.0110	0.9567	0.0032	0.0019	0.0434	0.9332
	SVM	0.0051	0.0000	0.0059	0.9874	0.0040	0.0020	0.0451	0.9280
	MLR	0.0385	0.0034	0.0584	0.8789	0.0392	0.0016	0.0394	0.4425
South Africa	ANN	0.0001	0.0001	0.0003	0.9999	0.0003	0.0000	0.0020	0.9994
	ANFIS	0.0024	0.0001	0.0081	0.9983	0.0129	0.0006	0.0239	0.9119
	SVM	0.0194	0.0023	0.0447	0.9389	0.0275	0.0008	0.0287	0.8735
	MLR	0.0041	0.0020	0.0447	0.9476	0.0058	0.0010	0.0065	0.9936

Figure 29: Time series plot for the cumulative incidences in South Africa

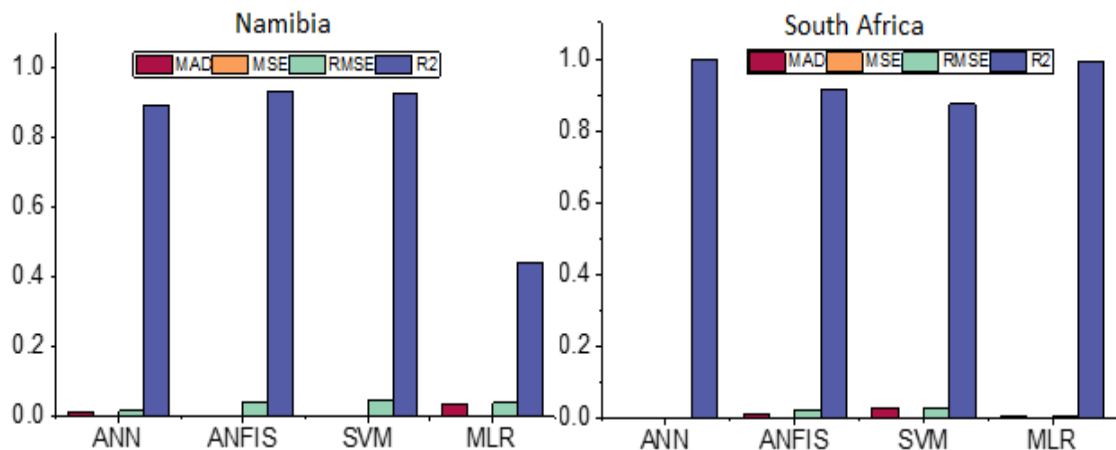
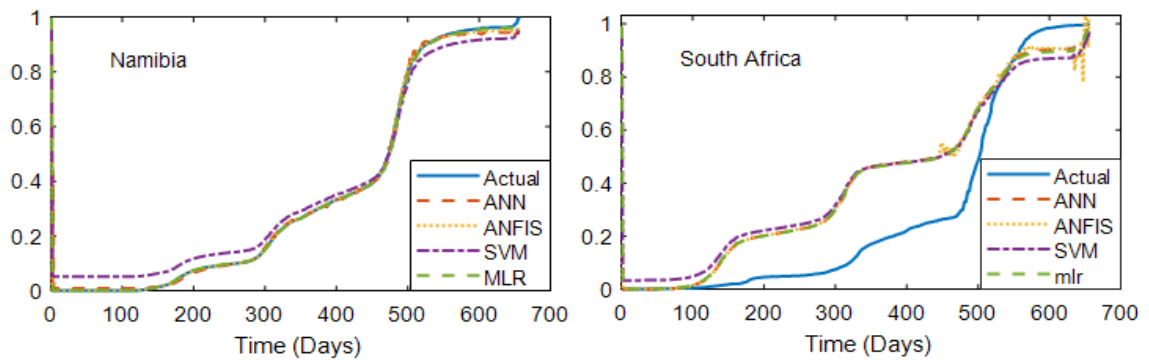


Figure 30: Bar chart illustrating the best performing models in Southern Africa

Table 18 displayed the outcomes of the single models' predictive performance for the cumulative figures in Namibian Republic and South Africa correspondingly. ANFIS is the

best performing model for Namibia and South Africa respectively at the step of validation having MAD = 0.0032, MSE = 0.0019, RMSE = 0.0434 and R2 = 0.9332 accompanied by SVM with MAD = 0.0040, MSE = 0.0020, RMSE = 0.0451 and R2 0.9280. The model with poor cumulative figures prediction performance in Namibia is MLR having values of mean absolute deviation of 0.0392, MSE = 0.0016, RMSE = 0.0394 and R2 = 0.4425. Artificial neural network is the model with better prediction efficiency in South Africa with MAD values of 0.0003, MSE = 0.0000, RMSE = 0.0020 and R2 = 0.9994. Figure 31 illustrates the time series plot of the cumulative figures for both Namibia and South Africa while figure 32 displayed the best predictive models.

Table 19: Outcomes of the engaged models in Central Africa

		Training				Validation			
Nation	Model	MAD	MSE	RMSE	R2	MAD	MSE	RMSE	R2
Gabon	ANN	0.0054	0.0018	0.0419	0.9580	0.0166	0.0009	0.0296	0.9494
	ANFIS	0.0100	0.0002	0.0135	0.9895	0.0041	0.0019	0.0438	0.9543
	SVM	0.0187	0.0004	0.0209	0.9747	0.0198	0.0023	0.0482	0.9445
	MLR	0.0055	0.0000	0.0069	0.9972	0.0045	0.0020	0.0450	0.9517
Cameroon	ANN	0.0042	0.0019	0.0433	0.9631	0.0106	0.0003	0.0179	0.9661
	ANFIS	0.0024	0.0018	0.0425	0.9645	0.0088	0.0003	0.0171	0.9690
	SVM	0.0276	0.0026	0.0513	0.9483	0.0309	0.0010	0.0320	0.8908
	MLR	0.0048	0.0001	0.0081	0.9930	0.0047	0.0021	0.0454	0.9595

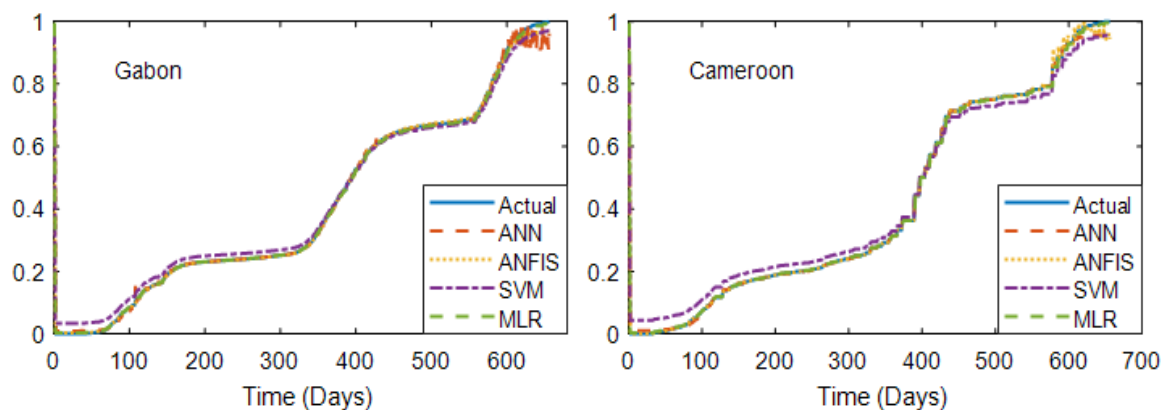


Figure 31: Time series plot of the cumulative incidences for Central African countries

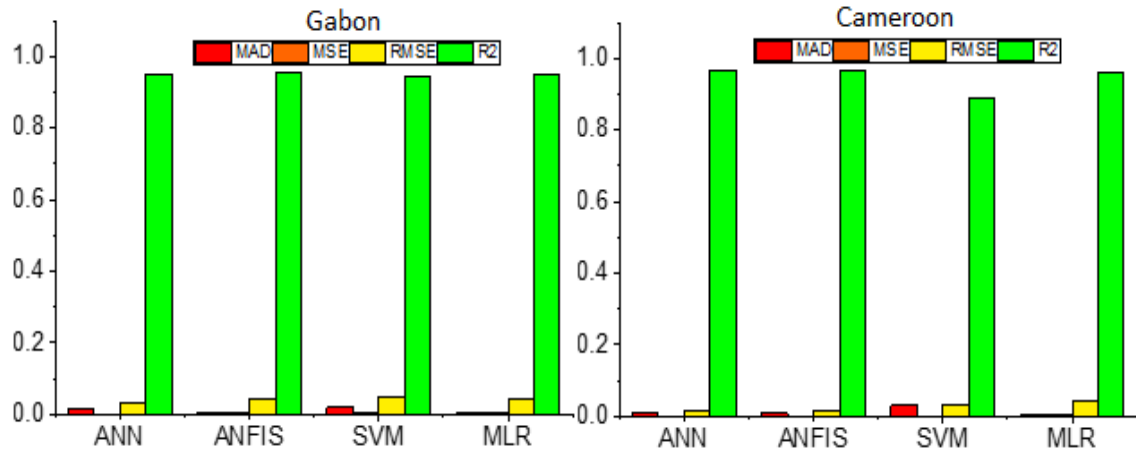


Figure 32: Bar chart illustrating the best performing models in Central African nations.

Table 19 contains the outcomes of the employed single models in Gabon and Cameroon at the step of validation. ANN is the most highly predictive model for Gabon having Mean Absolute Deviation = 0.0166, Mean Square Error = 0.0009, Root Mean Square Error = 0.0296 and R2 = 0.9494, accompanied by ANFIS with MAD = 0.0041, MSE = 0.0019, RMSE = 0.0438 and R2 = 0.9543. The best performing model in Cameroon for Cameroon is Adaptive Neuro Fuzzy Inference System with Mean Absolute Deviation = 0.0088, Mean Square Error = 0.0003, Root Mean Square Error = 0.0171 and R2 0.9690 accompanied by Artificial Neural Network with Mean Absolute Deviation = 0.0106, Mean Square Error = 0.0003, Root Mean Square Error = 0.0179 and R2 = 0.9661.

Figure 31 demonstrates the time series plot of the cumulative figures in both Gabon and Cameroon whereas figure 32 shows bar chart with the highly predictive models based on the global statistical indicators.

CHAPTER V

Discussion

COVID-19 pandemic was brought on by the novel SARS-CoV-2. The virus responsible for the outbreak was first recorded in Wuhan, Republic of China, in the month of December 2019, and it is highly contagious and was initially identified in bats before spreading to dogs and raccoons (Morens, Daszak & Taubenberger et al., 2020). COVID-19 pandemic is responsible for the death of millions number of people worldwide. The negative impact of the novel virus on the people's lives is very high (Hui et al., 2020). This work was carried out to assess the efficacy and accuracy of machine learning models and ARIMA models in predicting and forecasting COVID-19 cases in African regions respectively.

Phase one of this work estimated the performance efficiency of the four models in each West African state. These models are ARIMA projected based on the procedure of Box-Jenkins with maximum likelihood approach (ARIMAML), ARIMA according to automatic routine (AUTOARIMA), ensemble ARIMA together with ML as well as ARIMA with generalized least squares technique. Choice of the desired algorithm among them is evaluated according to the lowest value of the projected data such as SMAPE, RMSE, MAE, MAPE, Theil U_1 as well as Theil U_2 .

Two forecast strategies are used in this study: forecast using individual model, and forecast via ensembling ARIMAML and ARIMAGLS. Simple averaging of the ARIMAML and ARIMAGLS outputs is employed. The output of the ensemble models is then evaluated and weigh against with the individual output of other models. The data set was categorized into: 75% as training sub-sample and testing sub-sample (25%).

The CCC for each country could be modeled as ARIMA (p,1,q) processes, as each is integrated of order one (Akalpler, Ozdeser & Mati, 2017). The order of moving average (q) and order of autoregressive (p) is selected based on the combination that gives minimum SIC along white-noise errors. Since the CCC for each country is one (1), the first discrepancy of apiece cumulative COVID-19 case is taken before estimation.

The outcomes for North African countries show that in the training as well as validation stages, correspondingly, different models for Sudan and Morocco produce varied outcomes. For the validation stage, it is illustrated that the entire used models for Morocco

possessed R² values more than 0.7, which is an indication of the models' precision and correctness.

When comparing the validation stage results for Sudan and Morocco, it can be seen that Morocco's model performances are better. This may be linked to the information that Morocco possessed the highest daily verified cases count, with a supreme value of 12039, compared to Sudan's cumulative figures 1215. The forecasting algorithms were created to offer precise estimate according to the prior practice, but they struggle to function at their best when there are no incidents or cases on a given day but there are incidents or cases on a different day (like in the instance of Sudan).

According to the results for East African states, the models employed in Uganda performed poorly, model with best effectiveness at the validation step is the adaptive neuro-fuzzy inference system. The coronavirus disease's 2019 character, with its rapid or abrupt rise and fall of cases, may be the cause of the lowered or subpar model performance. Rwanda's model performance is comparatively better than Uganda's throughout the validation phase. However, significant differences amongst the models can be seen, which highlight the unpredictability of the confirmed cases in Eastern Africa. ANFIS produced the maximum effective results.

The predictive precision of the used models for West African states demonstrates that MLR models and artificial intelligence algorithms have the ability of making prediction in the republic of Nigeria. The capability of the machine learning based algorithms to take care of the stochastic, nonlinear, and unreliable phenomena related to coronavirus disease in 2019 may be the reason for their improved predictive power.

Regarding the Senegal outcomes, it could be perceived during the process of validation that they are being compared to those of Nigeria. Because, Senegal and Nigeria are in the same region in Africa. Since COVID-19 is frequently acquired through these channels and behaviour, culture, and social interaction are quite similar between the two nations, there is a similarity in the performance of confirmed incidences and predictive models.

For South Africa during the validation stage that the entire models have elevated performance precision records. Due to the fact that South Africa had the elevated figures of coronavirus disease 2019 daily (37875 within the study time), and that the stable movement of the verified incidences aids the models to carryout accurate prediction of

coronavirus disease 2019 in the nation, improving prediction accuracy. Furthermore, the elevated figures of COVID-19 cases in South Africa are associated with identification of SAR-CoV-2 variant (omicron) in the country. South Africa's weather condition is suitable for the novel virus and could be one of the reasons for the highest number of COVID-19 cases in the country. Many environmental factors such as sun radiation as well temperature could affect the novel virus transmissibility. Countries with range of temperature between 16-17 degree Celsius like Spain, Italy, UK and the U.S.A experienced the highest COVID-19 incidences, followed by countries with extreme low temperature like Russia, and countries with a range of temperature between 27–30-degree Celsius experienced low level of COVID-19 cases (Roy, 2020).

Regarding results for the Central African region comprising of Cameroon and Gabon illustrated in the results section, based on the validation stage results for the republic of Gabon, Adaptive Neuro Fuzzy Inference System possessed the supreme predictive precision accompanied by artificial neural network, then support vector machine, and lastly MLR model.

ANFIS also demonstrated better prediction performance for Cameroon. MLR's models are linear, but they nonetheless produce consistent results when compared to ANN as well as SVM. Given that it is a tool for non-linear classification and development has demonstrated greater predictive potential in several discoveries, the multiple linear regression (MLR) model's ability to forecast outcomes is actually not mysterious (Kouadri et al., 2021).

Based on the number of verified cases, the model's performance varies. For artificial neural network algorithm, with encouraging outcomes for average coronavirus disease 2019 assessment, ANN is capable of yielding greatest performance for Morocco and South Africa. Nevertheless, it is crucial to recognize that the quantity of daily verified COVID-19 instances may not always be relevant to the precision and effectiveness of the projecting models. Authorities' strict preventive actions, like the lockdowns, people isolation, applying of hand sanitizers, etc., are crucial to the detection of cases and effective forecasting. For instance, Cameroon has fewer cases than Nigeria and numerous other nations combined. However, the steps made by the Cameroonian administration to

curb the impact and blow-out of the deadly pathogen make it simpler in disentangling the unknowns associated with the COVID-19, this led the models to have dependable forecast. One can see that the nature of the models is alike in terms of performance is similar crosswise countries by looking at the efficacy of the algorithms. The two nations where the models perform best are South Africa and Morocco, which are accompanied by Republic Cameroon, Republic of Nigeria, Republic Rwanda, Senegal, Gabon, Republic of Sudan, Namibia, and Uganda. Furthermore, it is evident from this finding that ANFIS performed better in almost every country. This is due to the fact that it successfully demonstrates the benefit of merging NN and fuzzy logic.

The results of the applied ensemble models demonstrate significant enhancement with minimal errors and elevated R2 values that are typically greater than 0.9. When ANN ensemble outcomes are compared to those from single models, it is clear that performance has improved significantly. In the validation step, the ANN-E increased the modeling precision of ANN algorithm for the republic of Morocco, Sudan, Namibian republic, Southern Africa, republic of Uganda, Rwanda, republic of Nigeria, Senegal republic, Gabon, as well as Cameroon, correspondingly, by 10%, 14%, 42%, 6%, 83%, 11%, 7%, 5%, 7%, and 31%.

With reduced performance from the individual models, ensemble approaches fared well, according to the results (Nourani, Ekiran & Abdullahi, 2019b). Example, Uganda, the nation with the lowest predictive effectiveness of the single algorithms, saw the biggest increase in model performance, amounting to 83%. On the other side, it was discovered that countries with the highest single modeling accuracy experienced the least increase in their efficiency as a result of ensemble approaches. For instance, Republic of Morocco and S. Africa were shown to possessed efficiency improvements of just 10% and 6% despite having the highest successful modeling of the daily-verified COVID-19 instances by individual models. To advance prediction, a single model with poor performance would leave a huge space, whereas a single model with good performance would only leave a tiny space. The performance of the models in Cameroon, one of the countries with the best single model performance, has increased by 31% as a result of the ensemble models. Taylor figure takes in to justification of RMSE amid forecasting by the algorithms and figures detected and correlations manner as well, which reviews or summarize the general

model's performance (Abdullahi et al., 2019). From the diagram, correlation coefficients, root mean square error as well as standard deviation are applied to examine the resemblance amid observed records and the algorithms for the modelling. The performance of the other models was displayed in relation to how far away the observed set data was, and the observed data set was plotted together with the abscissa as a loop (Al-Sultani et al., 2021). Generally, underestimating occurs when the observed values exceed the projected or predicted standard deviation values. On the flip hand, overestimation happens when the projected values are bigger than the standard deviation of the observed figures (Abdullahi et al., 2019c).

The ensemble-based values exhibit greater modelling performance, as can be observed. In general, the findings gained in this work showed that ensemble-based models can enhance the modelling precision of single models. However, the figure of COVID-19 occurrences and the precautions taken by each nation may have an effect on how well each model simulates the disease. By using ensemble-based models, it is possible to considerably explain and predict the stochastic as well as unreliable nature of the verified cases of COVID-19 in Africa. The employed machine learning models together with conventional multi linear regressions displayed an elevated prediction performance in all the countries within the study area when modelling the cumulative confirmed incidences. The better prediction accuracy of the single models is directly connected with the use of cumulative figures instead of daily cases.

CHAPTER VI

Conclusion and Recommendations

6.1 Conclusion

In regards to ARIMA finding, the forecast performance of the 4 different Autoregressive moving average models, which include the ARIMAML, ARIMAGLS, AUTOARIMA and ARIMAML-ARIMAGLS. Time series data of CCC for fifteen ECOWAS countries is used to estimate the models. The dataset was categorized into 2 sub-samples: 75% (training) and (for testing (25%). ARIMAGLS outperforms other models for seven nations in the training stage and eight countries for the testing sample. In both sub-samples, ARIMA along generalized least squares technique and ensemble ARIMA with maximum likelihood tends to produce better forecast correctness for some countries. This implies the importance of GLS estimation method in improving the forecast accuracy.

In order to anticipate the COVID-19 pandemic in 10 African nations, across all the regions, ensemble procedures dubbed ANN ensemble and SVM ensemble were used. These approaches have the benefit over others in that they consider both the non-linear and linear features of COVID-19 when making the predictions stuffs.

Three machine learning (ML) based models—SVM, ANN and ANFIS were employed firstly as independent models for the coronavirus disease 2019 prediction and forecasting in order to meet the finding's objectives. For comparison, MLR model was also applied. Standalone models' results were then used as the input kernels for ANN and SVM to increase performance.

Because it is one of the most significant problems that the entire human race is presently experiencing, the used ANN ensemble and SVM ensemble were evaluated on COVID-19. The suggested approaches are also generalizable and could be applied to project whichever time series. The comparative evaluation as well as simulation results revealed that the planned ANN-E and SVM ensemble methods could be helpful means for effectiveness regarding enhancing the effectiveness of time series projection, and they outshined all other individual methods tried via the same set of data.

The key and core contributions of this work are: (i) Predictive precision of the machine learning algorithms was enhanced and improved by the applied procedures for daily-affirmed coronavirus disease 2019 prediction in African countries. In spite of the complicated nature of coronavirus disease plague, excellent enhancements in outcomes

were realized using the planned ensemble methods, and could be applied as substitute procedures for all forms of disease epidemic predictions and forecasting, that can aid the policy architects in order to make resolutions on methods to put on as well as the time of their putting into practice. (ii) The applied methods that in situation of disease occurrence, the conventional models of epidemiology along the machine learning based ensemble techniques might be applied for infectious disease prediction of cases. (iii) The performance of ARIMAGLS and Ensemble ARIMA in COVID-19 cases forecast was established.

6.2 Recommendations

The suggested approaches (Machine learning single models and ARIMA models) are also generalizable and can be used to project and forecast whichever time series. To make an effective evaluation of ensemble methods performance, for future research, linear ensemble procedures comprising of simple linear average ensemble (SLAE) as well as weighted linear average ensemble should be employed so as to determine and evaluate the effective or proficient ensemble methods for coronavirus disease 2019 forecasting. Further discoveries should also consider and focus on the ensemble-based models' application for modeling mortality of coronavirus disease 2019 in African nations. Other categories of artificial intelligence models, ensemble kernels like genetic (algorithms), etc. can be applied for more research to determine their precision. Strict actions are essential in order to ensure total elimination of the novel virus. Meanwhile, further studies should focus on modeling the SAR-CoV-2 variants with special emphasis in Africa and examine the role of the new variants in the new cases elevation.

References

- Abdullahi, J., Elkiran, G., & Nourani, V. (2017). Application of artificial neural network to predict reference evapotranspiration in Famagusta, North Cyprus. In *11th International Scientific Conference on Production Engineering Development and Modernization of Production* (pp. 549-554).
- Ayinde, K., Lukman, A., Rauf, R., Alabi, O., Okon, C., & Ayinde, O. (2020). Modeling Nigerian covid-19 cases: A comparative analysis of models and estimators. *Chaos, Solitons & Fractals*, 138; 109911. doi:<https://doi.org/10.1016/j.chaos.2020.109911>.
- Abdullahi, J., Elkiran, G., & Nourani, V. (2019). Artificial intelligence based and linear conventional techniques for reference evapotranspiration modeling. In *International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions* (pp. 197-204). Springer, Cham. https://doi.org/10.1007/978-3-030-35249-3_25
- Abdullahi, J., Iravani, A., Nourani, V., & Elkiran, G. (2019). Application of artificial intelligence based and multiple regression techniques for monthly precipitation modeling in coastal and inland stations. *Desalin. Water Treat.*, 177, 338-349. <https://doi.org/10.5004/dwt.2020.24954>
- Andrew, G.H., Tao, L., & Penghua, W. (2020). Mechanism of SARS-CoV-2 transmission and pathogenesis. *Trends in Immunology, Cell Press*, 41(12). <https://doi.org/10.1016/j.it.2020.10.004>
- Al-Sultani, A. O., Al-Mukhtar, M., Roomi, A. B., Farooque, A. A., Khedher, K. M., & Yaseen, Z. M. (2021). Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access*, 9, 108527-108541. <https://doi.org/10.1109/ACCESS.2021.3100490>
- Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U & Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. *Algorithms*, 13(10), 249. <https://doi.org/10.3390/a13100249>
- Aleem, A., Akbar, Samad., A.B., & Slenker, A.K. (2022). Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). *StatPearls Publishing*. 212(34).

- ArunKumar, K., Kalaga, D., Kumar, M., Chilkoor, G., Kawaji, G., & Brenza, T. (2021). Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-regressive integrated moving average (ARIMA) and seasonal auto-regressive integrated moving average (SARIMA). *Applied Soft Computing*, 103; 107161. <https://doi.org/10.1016/j.asoc.2021.107161>.
- Alabdulrazzaq, H., Alenezi, M., Rawajfih, Y., Alghannam, B., Al-Hassan, A., & Al-Anzi, F. (2021). On the accuracy of ARIMA based prediction of COVID-19 spread. *Results in Physics*, 27; 104509. Doi:<https://doi.org/10.1016/j.rinp.2021.104509>.
- Akalpler, E., Ozdeser, H., & Mati., S. (2017). Trade-volatility relationship in the light of Nigeria and the euro area. *Journal of Applied Economic Sciences*, 12 (2017).
- Bhagat, S. K., Pyrgaki, K., Salih, S. Q., Tiyasha, T., Beyaztas, U., Shahid, S., & Yaseen, Z. M. (2021). Prediction of copper ions adsorption by attapulgitite adsorbent using tuned-artificial intelligence model. *Chemosphere*, 276, 130162 <https://doi.org/10.1016/j.chemosphere.2021.130162>
- Bhattacharya, M., Chatterjee, S., Sharma, A.R., Agoramoorthy, G., & Chakraborty, C. (2021). D614G mutation and SARS-CoV-2: impact on S-protein structure, function, infectivity, and immunity. *Appl Microbiol Biotechnol*, 105(24):9035-9045. doi: 10.1007/s00253-021-11676-2
- Box, G., & Jenkins. G. (1976). Time series analysis: Forecasting and control. *San Francisco, Calif: Holden-Day*.
- Box, G., Jenkins, G., Reinsel, G., & Ljung. M. (2015). Time series analysis: Forecasting and control. *John Wiley & Sons*.
- Burki, T.K. (2020). Coronavirus in China. *The Lancet, Respiratory Medicine*, 8 (2020) 238. Doi:10.1016/S2213-2600(20)30056-4.
- Cagliani, R., Forni, D., Clerici, M. & Sironi, M. (2020). Coding potential and sequence conservation of SARS-CoV-2 and related animal viruses. *Infect. Genet. Evol.* 83, 104353.
- Chenar, S. S., & Deng, Z. (2018). Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico

- coast. *Environment international*, 111, 212-223.
<https://doi.org/10.1016/j.envint.2017.11.032>
- Cui, H.Z., Gao, Z.Y., & Liu M. (2020). Structural genomics and interactomics of 2019 Wuhan novel coronavirus, 2019-nCoV, indicate evolutionary conserved functional regions of viral proteins. *bioRxiv, preprint*.
- Chowdhury, R., Heng, K., Shawon, S., Goh, G., Okonofua, D., & Ochoa-Rosales, C. (2020). Dynamic interventions to control COVID-19 pandemic: A multivariate prediction modelling study comparing 16 worldwide countries. *European Journal of Epidemiology*, 35 (2020) 389–399. doi:10.1007/s10654-020-00649-w.
- Chen, L.J., Liu, W.Y., & Zhang, Q. (2020). RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg Microbes Infect*, 9(1):313-319.
- Center for disease control. (2021). Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. <https://www.cdc.gov/mmwr/volumes/70/wr/mm7003e2.htm>
- Cody, B.J., Micheal, F., Bing, C., Hyeryun, C. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews, Molecular Cell Biology*, 23(3).
<https://doi.org/10.1038/s41580-021-00418-x>
- Dickey, D., & Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74 (1979) 427–431. doi:10.1080/01621459.1979.10482531.
- Das, R. (2020). Forecasting incidences of COVID-19 using Box-Jenkins method for the period July 12-septembert 11, 2020: A study on highly affected countries, Chaos, Solitons & Fractals. doi:<https://doi.org/10.1016/j.chaos.2020.110248>.
- Elkiran, G., Nourani, V., Elvis, O., & Abdullahi, J. (2021). Impact of climate change on hydro-climatological parameters in North Cyprus: application of artificial intelligence-based statistical downscaling models. *Journal of Hydroinformatics*, 23(6), 1395-1415. <https://doi.org/10.2166/hydro.2021.091>
- Ekhmaj, A.I. (2012). Prediction of evapotranspiration using artificial neural networks model. *Malaysia International Annual Symposium on sustainability Science and Management. Terengganu*, pp. 937–943.

- Fernández, A. (2020). Structural impact of mutation D614G in SARS- CoV-2 spike protein: enhanced infectivity and therapeutic opportunity. *ACS Med. Chem. Lett*, 11, 1667–1670.
- Firth, A. E. (2020). A putative new SARS-CoV protein, 3c, encoded in an ORF overlapping ORF3a. *J. Gen. Virol.* <https://doi.org/10.1099/jgv.0.001469>.
- Guleryuz, D. (2021). Forecasting outbreak of COVID-19 in Turkey; comparison of Box–Jenkins, brown’s exponential smoothing and long short-term memory models. *Process Safety and Environmental Protection*, 927–935. doi:<https://doi.org/10.1016/j.psep.2021.03.032>.
- Groves, D.C., Rowland-Jones, S.L., Angyal A. (2021). The D614G mutations in the SARS-CoV- 2 spike protein: Implications for viral infectivity, disease severity and vaccine design. *Biochem Biophys Res Commun.* 2021 Jan 29; 538:104-107. doi: 10.1016/j.bbrc.2020.10.109
- Guo, Q., Li, M., & Wang, C.H. (2020). Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. *bioRxiv, preprint.* <https://doi.org/10.1101/2020.01.21.914044>
- Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences*, 75(6), 1-14. <https://doi.org/10.1007/s12665-015-5096-x>
- Harvey, W.T., Carabelli, A.M., & Jackson, B. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **19**, 409–424. <https://doi.org/10.1038/s41579-021-00573-0>
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 395(10223), 497-506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- Hoffmann, M., Klein, H., & Krüger, N. (2020). The novel coronavirus 2019 (2019-nCoV) uses the SARS- coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. *bioRxiv, preprint.* <https://doi.org/10.1101/2020.01.31.929042>

- Huang, Q., & Herrmann., A. (2020). Fast assessment of human receptor-binding capability of 2019 novel coronavirus (2019-nCoV). *bioRxiv, preprint*. <https://doi.org/10.1101/2020.02.01.930537>
- Hui, D., Azhar, E., Madani, T., Ntoumi, F., Kock, R., & Dar., O. (2020). The continuing 2019- nCoV epidemic threat of novel coronaviruses to global health — the latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases, 91* (202),264–266. doi:<https://doi.org/10.1016/j.ijid.2020.01.009>.
- Hodcroft, E.B., Zuber, M., Nadeau, S., Vaughan, T.G., Crawford, K.H.D., Althaus, C.L., Reichmuth, M.L., Bowen, J.E., Walls, A.C., Corti, D., Bloom, J.D., Veessler, D., Mateo, D., Hernando, A., Comas, I., González, & Candelas, F (2020). SeqCOVID-SPAIN consortium; Stadler T, Neher RA. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *MedRxiv, Preprint*. doi: 10.1101/2020.10.25.20219063
- Hyndman, J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(1), 1–22.
- Islam, M., Mahmud, S., Muhammad, L. J., Nooruddin, S., & Ayon, S. I. (2020). Wearable technology to assist the patients infected with novel coronavirus (COVID-19). *SN computer science, 1*(6), 1-9. <https://doi.org/10.1007/s42979-020-00335-4>
- Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review, 136*, 101922. <https://doi.org/10.1016/j.tre.2020.101922>
- Jabra, B., Marwa, Anis, K., Bilel, B., Adel, A., & Habib, H. (2021). "COVID-19 Diagnosis in Chest X-rays Using Deep Learning and Majority Voting" *Applied Sciences* 11, no. 6: 2884. <https://doi.org/10.3390/app11062884>
- Jang, J.S.R., Sun, C.T., & Mizutani, E. (1997). Neuro-fuzzy and Soft Computing—A Computational Approach to Learning and Machine Intelligence. *Prentice Hall, Newb Jersey*.
- Jain, S., Venkataraman, A., Wechsler, M.E., & Peppas, N.A. (2021). Messenger RNA-based vaccines: Past, present, and future directions in the context of the COVID-

- 19 pandemic. *Adv Drug Deliv Rev*, 179:114000. doi: 10.1016/j.addr.2021.114000
- Jaimes, J., André, N., & Millet, J. (2020). Structural modeling of 2019-novel coronavirus (nCoV) spike protein reveals a proteolytically-sensitive activation loop as a distinguishing feature compared to SARS-CoV and related SARS-like coronaviruses. *bioRxiv, preprint*. <https://doi.org/10.1101/2020.02.10.942185>
- Ji, W., Wang, W., & Zhao, X. (2020). Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *J Med Virol*, 92(4):433-440. <https://doi.org/10.1002/jmv.25682>
- Jebara, T. (2012). *Machine learning: discriminative and generative* (Vol. 755). Springer Science & Business Media.
- Kiran, N. R., & Ravi, V. (2008). Software reliability prediction by soft computing techniques. *Journal of Systems and Software*, 81(4), 576-583. <https://doi.org/10.1016/j.jss.2007.05.005>
- Kouadri, S., Elbeltagi, A., Islam, A. R. M., & Kateb, S. (2021). Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). *Applied Water Science*, 11(12), 1-20. <https://doi.org/10.1007/s13201-021-01528-9>
- Kocadagli, O., Baygul, A., Gokmen, N., Incir, S., & Aktan, C. (2022). Clinical prognosis evaluation of COVID-19 patients: An interpretable hybrid machine learning approach. *Current Research in Translational Medicine*, 70(1), 103319. <https://doi.org/10.1016/j.retram.2021.103319>
- King, A., Adams, M., & Carstens, E. (2012). Virus Taxonomy. *Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier, San Diego, USA, p.770-783.
- Lau, h., Khosrawipour, G., Kocbach, V., Mikolajczyk, A., Schubert, J., & J. Bania, J. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of Travel Medicine*. 27(2020). doi:10.1093/jtm/taaa037.

- Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol*, 3(1):237-261.
- Li, M., Wei, D., Liu, T., Liu, Y., Yan, L., Wei, Q., Du, B., & Xu, W. (2019). EDTA functionalized magnetic biochar for Pb (II) removal: Adsorption performance, mechanism and SVM model Prediction. *Separation and Purification Technology*. <https://doi.org/10.1016/j.seppur.2019.115696>
- Lan, J., Ge, J.W., & Yu, J.F. (2020). Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *bioRxiv*, preprint. <https://doi.org/10.1101/2020.02.19.956235>
- Lucas, B., Vahedi, B., & Karimzadeh, M. (2022). A spatiotemporal machine learning approach to forecasting COVID-19 incidence at the county level in the USA. *International Journal of Data Science and Analytics*, 1-20. <https://doi.org/10.1007/s41060-021-00295-9>
- Mahase, E. (2020). China coronavirus: what do we know so far?. <https://doi.org/10.1136/bmj.m308>.
- Mati, S. (2021). Do as your neighbors do? Assessing the impact of lockdown and reopening on the active COVID-19 cases in Nigeria. *Social Science & Medicine*. 270 (2021) 113645. doi:<https://doi.org/10.1016/j.socscimed.2020.113645>.
- Mallano, A., Ascione, A., & Flego., M. (2022). Antibody Response against SARS-CoV-2 Infection: Implications for Diagnosis, Treatment and Vaccine Development. *Int Rev Immunol*, 41(4):393-413. doi: 10.1080/08830185.2021.1929205.
- Meng, T., Cao, H., & Zhang, H. (2020). The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. *bioRxiv*, preprint.
- Mirza, S., Niwalkar, A., Gupta, A., Gautam, S., Anshul, A., Bherwani, H., & Kumar, R. (2022). Is safe distance enough to prevent COVID-19? Dispersion and tracking of aerosols in various artificial ventilation conditions using Open FOAM. *Gondwana Research*. <https://doi.org/10.1016/j.gr.2022.03.013>
- Morens, D. M., Daszak, P., & Taubenberger, J. K. (2020). Escaping Pandora's box—another novel coronavirus. *New England Journal of Medicine*, 382(14), 1293-1295. <https://doi.org/10.1056/NEJMp.2002106>

- Muhammad, L. J., Algehyne, E. A., Usman, S. S., Ahmad, A., Chakraborty, C., & Mohammed, I. A. (2021). Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. *SN computer science*, 2(1), 1-13. <https://doi.org/10.1007/s42979-020-00394-7>
- Mwenda, M., Saasa, N., Sinyange, N., Busby, G., Chipimo, P.J., Hendry, J., Kapona, O., Yingst, S., Hines, J.Z., Minchella, P., Simulundu, E., Changula, K., Nalubamba, K.S., Sawa, H., Kajihara, M., Yamagishi, J., Kapin'a, M., Kapata, N., Fwoloshi, S., Zulu, P., Mulenga, L.B., Agolory, S., Mukonka, V., & Bridges, D.J. (2020). Detection of B.1.351 SARS-CoV-2 Variant Strain - Zambia, December 2020. *MMWR Morb Mortal Wkly Rep*, 26;70(8):280-282. doi: 10.15585/mmwr.mm7008e2
- Naganna, S., Deka, P., Ghorbani, M., Biazar, S., Al-Ansari, N., Yaseen, Z., (2019). Dew Point Temperature Estimation: Application of Artificial Intelligence Model Integrated with Nature-Inspired Optimization Algorithms. *Water*. <https://doi.org/10.3390/w11040742>
- Nourani, V., Elkiran, G., & Abdullahi, J. (2019b). Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *Journal of Hydrology*, 577, 123958. <https://doi.org/10.1016/j.jhydrol.2019.123958>
- Nourani, V., Elkiran, G., & Abdullahi, J. (2020). Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. *Journal of Hydrology*, 581, 124434. <https://doi.org/10.1016/j.jhydrol.2019.124434>
- Phillips, P., & Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika*. 75 (1988) 335–346. <http://www.jstor.org/stable/2336182>.
- Pascarella, S., Ciccozzi, M., Zella, D., Bianchi, M., Benedetti, F., Benvenuto, D., Broccolo, F., Cauda, R., Caruso, A., Angeletti, S., Giovanetti, M., Cassone, A. (2021). SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate? *J Med Virol*, 93(12):6551-6556. doi: 10.1002/jmv.27210.
- Pustake, M., Tambolkar, I., Giri, P., Gandhi, C. (2022). SARS, MERS and CoVID-19: An overview and comparison of clinical, laboratory and radiological features. *J Family Med Prim Care*. 11(1):10-17. doi: 10.4103/jfmpc.jfmpc_839_21

- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., & Hyman, J. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5 (2020) 256–263. doi:<https://doi.org/10.1016/j.idm.2020.02.002>.
- Rothe, C., Schunk, S., Bretzel, F., & Wallrauch, M. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine*, 382 (2020). doi:10.1056/nejmc2001468.
- Roy, I. (2020). The role of temperature on the global spread of COVID-19 and urgent solutions. *International Journal of Environmental Science and Technology*. <https://doi.org/10.1007/s13762-020-02991-8>
- Roy, B., Dhillon, J.K, Habib, N., & Pugazhandhi, B. (2021). Global variants of COVID-19: Current understanding. *Journal of Biomedical Sciences*, 2021;8(1):8-11
- Rostami-Tabar, Rendon-Sanchez, J. (2021) Forecasting COVID-19 daily cases using phone call data, *Applied Soft Computing*, 100 (20). 106932. doi:<https://doi.org/10.1016/j.asoc.2020.106932>.
- Sajda, P. (2006). Machine learning for detection and diagnosis of disease. *Annual review of biomedical engineering*, 8(1), 537-565.
- Sharghi, E., Nourani, V., & Behfar, N. (2018). Earthfill dam seepage analysis using ensemble artificial intelligence based modeling. *Journal of Hydroinformatics*, 20(5), 1071-1084. <https://doi.org/10.2166/hydro.2018.151>
- Scarpino, S. V., & Petri, G. (2019). On the predictability of infectious disease outbreaks. *Nature communications*, 10(1),1-8. <https://doi.org/10.1038/s41467-019-08616-0>
- Sujatha, A., Govindaraju, L., Shivakumar, N., & Devaraj, V. (2021). Fuzzy knowledge based system for suitability of soils in airfield applications. *Civil Engineering Journal*, 7(1), 140-152.
- Talebkeikhah, M., Sadeghtabaghi, Z., & Shabani, M. (2021). A comparison of machine learning approaches for prediction of permeability using well log data in the hydrocarbon reservoirs. *Journal of Human, Earth, and Future*, 2(2), 82-99.

- Tada, T., Dcosta, B.M, Samanovic-Golden, M., Herati, R.S., Cornelius, A., Mulligan, M.J., Landau, N.R. (2021). Neutralization of viruses with European, South African, and United States SARS-CoV-2 variant spike proteins by convalescent sera and BNT162b2 mRNA vaccine-elicited antibodies. *bioRxiv* [Preprint], 7:2021.02.05.430003. doi: 10.1101/2021.02.05.430003
- Tang, Y., & S. Wang, S. (2020). Mathematic modeling of COVID-19 in the united states, *Emerging Microbes & Infections*. 827–829. doi:10.1080/22221751.2020.1760146.
- Tang, J.W., Toovey, O.T.R., Harvey, K.N., & Hui, D.D.S. (2021). Introduction of the South African SARS-CoV-2 variant 501Y.V2 into the UK. *J Infect*, 82(4):e8-e10. doi: 10.1016/j.jinf.2021.01.007
- Vapnik, V., (1998). The support vector method of function estimation, in: *Nonlinear Modeling*. Springer, pp. 55–85. https://doi.org/10.1007/978-1-4615-5703-6_3
- Vasireddy, D., Vanaparthy, R., Mohan, G., Malayala, S.V., & Atluri, P (2021). Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? *J Clin Med Res*, 13(6):317-325. doi: 10.14740/jocmr4518
- Wen, X., Si, J., He, Z., Wu, J., Shao, H., & Yu, H. (2015). Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resource Management*. 29 (9), 3195–3209. <https://doi.org/10.1007/s11269-015-0990-2>
- Wrapp, D., Wang, N., & Corbett, K. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483):1260-1263. <https://doi.org/10.1126/science.abb2507>
- Walls, A., Park, Y.J., & Tortorici, M.A. (2020). Structure, function and antigenicity of the SARS CoV-2 spike glycoprotein. *Cell*, 180:281-292.
- World health organization. (2021). The effect of virus variants on COVID-19 vaccines. <https://www.who.int/news-room/feature-stories/detail/the-effects-of-virus-variants-on-covid-19-vaccines>
- World health organization. (2021). Seven things to know about COVID-19 variants in Africa. <https://www.afro.who.int/news/seven-things-know-about-covid-19-variants-africa>

- Wax, R.S., & Christian, M.D. (2020). Practical recommendations for critical care and esthesiology teams caring for novel coronavirus (2019-nCoV) patients. *Can J Anesth/J Can Anesth*, 67:568-576.
- Wit, E., Van, D., Ien, N., & Falzarano, D. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol*, 14(8):523-534.
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T. (2020). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*, 11;27(3):325-328. doi: 10.1016/j.chom.2020.02.001.
- WHO (2020), Key messages and actions for COVID-19 prevention and control in schools
- Xiong, Y., Ma, Y., Ruan, L., Li, D., Lu, C., & Huang, L. (2022). Comparing different machine learning techniques for predicting COVID-19 severity. *Infectious diseases of poverty*, 11(1), 1-9. <https://doi.org/10.1186/s40249-022-00946-4>
- Yu, P., Zhu, J., Zhang, Z., & Y. Han, Y. (2020). A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. *The Journal of Infectious Diseases*. doi:10.1093/infdis/jiaa077.
- Ye, Z.W., Yuan, S., Yuen, K.S., Fung, S.Y., Chan, C.P., & Jin, D.Y. (2020). Zoonotic origins of human coronaviruses. *Int J Biol Sci* 16:1686–1697
- Zhan, Z., Dong, W., Lu, Y., Yang, P., Wang, Q., & Jia, P. (2019). Real-time forecasting of hand-foot-and-mouth disease outbreaks using the integrating compartment model and assimilation filtering. *Scientific reports*, 9(1), 1-9. <https://doi.org/10.1038/s41598-019-38930-y>
- Zhang, W., Davis, B.D., Chen, S.S., Martinez, J.M., Plummer, J.T., & Vail, E. (2021). Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA*, 6;325(13):1324-1326. doi: 10.1001/jama.2021.1612.
- Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D.S., Polykovskiy, D., Shayakhmetov, R., Filimonov, A., Orekhov, P. (2020). Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. *Insilico Med Hong Kong Ltd A* 307: EI. <https://doi.org/10.26434/chemrxiv.12301457>

- Zivkovic, M., Bacanin, N., Venkatachalam, K., Nayyar, A., Djordjevic, A., Strumberger, I., & Al-Turjman, F. (2021). COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustainable Cities and Society*, *66*, 102669. <https://doi.org/10.1016/j.scs.2020.102669>
- Zhou, P., Yang, X., & Wang, X. (2020). Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv*, preprint.
- Zheng, T., Guo, J., He, W., Wang, H., Yu, H., & Ye, H. (2020). Coronavirus disease 2019 (COVID-19) in pregnancy: 2 case reports on maternal and neonatal outcomes in Yichang city, Hubei Province, China. *Medicine (Baltimore)*, *17*;99(29). doi: 10.1097/MD.00000000000021334.
- Zamorano, C., & Grandvaux, N. (2020). ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. *Elife*, *9*;9: e61390. doi: 10.7554/eLife.61390

Appendix X

Turnitin Similarity Report

Zurki Ibrahim Thesis Turnit in

ORJİNALLİK RAPORU

% 11	% 9	% 9	%
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	www.ncbi.nlm.nih.gov İnternet Kaynağı	% 4
2	link.springer.com İnternet Kaynağı	% 1
3	www.esp.org İnternet Kaynağı	% 1
4	www.ttstats.nl İnternet Kaynağı	<% 1
5	www.mdpi.com İnternet Kaynağı	<% 1
6	Zurki Ibrahim, Pinar Tulay, Jazuli Abdullahi. "Multi-region machine learning-based novel ensemble approaches for predicting COVID-19 pandemic in Africa", Environmental Science and Pollution Research, 2022 Yayın	<% 1
7	www.researchgate.net İnternet Kaynağı	<% 1
8	journalcjast.com İnternet Kaynağı	

		<% 1
9	Mohammad Saood Manzar, Mohammed Benaafi, Romulus Costache, Omar Alagha et al. "New generation neurocomputing learning coupled with a hybrid neuro-fuzzy model for quantifying water quality index variable: A case study from Saudi Arabia", Ecological Informatics, 2022 Yayın	<% 1
10	Gemma Massonis, Julio R. Banga, Alejandro F. Villaverde. "Structural identifiability and observability of compartmental models of the COVID-19 pandemic", Annual Reviews in Control, 2020 Yayın	<% 1
11	www.frontiersin.org İnternet Kaynağı	<% 1
12	www.nornesk.no İnternet Kaynağı	<% 1
13	academic.oup.com İnternet Kaynağı	<% 1
14	m.scirp.org İnternet Kaynağı	<% 1
15	Ali Ghaemi, Amin Safari, Hadi Afsharirad, Hossein Shayeghi. "Situational awareness and	<% 1

deficiency warning system in a smart distribution network based on stacking ensemble learning", *Applied Soft Computing*, 2022

Yayın

-
- 16** Abebaw Bizuneh Alemu, U. Jaya Parakash Raju, Abdu Mohammed Seid, Bayile Damtie. "Comparative study of seasonal autoregressive integrated moving average and Holt-Winters modeling for forecasting monthly ground-level ozone", *AIP Advances*, 2023
Yayın <% 1
-
- 17** www.researchsquare.com
İnternet Kaynağı <% 1
-
- 18** Rochelle P. Walensky, Henry T. Walke, Anthony S. Fauci. "SARS-CoV-2 Variants of Concern in the United States—Challenges and Opportunities", *JAMA*, 2021
Yayın <% 1
-
- 19** Abdelgader Alamrouni, Fidan Aslanova, Sagiru Mati, Hamza Sabo Maccido, Afaf. A. Jibril, A. G. Usman, S. I. Abba. "Multi-Regional Modeling of Cumulative COVID-19 Cases Integrated with Environmental Forest Knowledge Estimation: A Deep Learning Ensemble Approach", *International Journal of* <% 1

Environmental Research and Public Health,
2022

Yayın

20 ojin.nursingworld.org <% 1
İnternet Kaynağı

21 Hassan Heidari, Salih Turan Katircioglu,
Narmin Davoudi. "Are Current Account
Deficits Sustainable? New Evidence from Iran
Using Bounds Test Approach to Level
Relationship", Economics: The Open-Access,
Open-Assessment E-Journal, 2012
Yayın

22 Sandro M. Hirabara, Tamires D. A. Serdan,
Renata Gorjao, Laureane N. Masi et al. "SARS-
COV-2 Variants: Differences and Potential of
Immune Evasion", Frontiers in Cellular and
Infection Microbiology, 2022
Yayın

23 Marwa Ben Jabra, Anis Koubaa, Bilel Benjdira,
Adel Ammar, Habib Hamam. "COVID-19
Diagnosis in Chest X-rays Using Deep
Learning and Majority Voting", Applied
Sciences, 2021
Yayın

24 dspace.bracu.ac.bd <% 1
İnternet Kaynağı

25	Laila Elmancy, Hala Alkhatib, Anis Daou. "SARS-CoV-2: An Analysis of the Vaccine Candidates Tested in Combatting and Eliminating the COVID-19 Virus", Vaccines, 2022	<% 1
Yayın		
26	M. Pérez-Abeledo, J.C. Sanz Moreno. "SARS-CoV-2 variants, a still unfinished story", Vacunas (English Edition)	<% 1
İnternet Kaynağı		
27	Nathalie Chazal. "Coronavirus, the King Who Wanted More Than a Crown: From Common to the Highly Pathogenic SARS-CoV-2, Is the Key in the Accessory Genes?", Frontiers in Microbiology, 2021	<% 1
Yayın		
28	sciendo.com	<% 1
İnternet Kaynağı		
29	Suraj Kumar Bhagat, Konstantina Pyrgaki, Sinan Q. Salih, Tiyasha Tiyasha et al. "Prediction of Copper ions adsorption by Attapulgit adsorbent using tuned-artificial intelligence model", Chemosphere, 2021	<% 1
Yayın		
30	authors.library.caltech.edu	<% 1
İnternet Kaynağı		
curve.carleton.ca		

31	İnternet Kaynağı	<% 1
32	dspace.nwu.ac.za İnternet Kaynağı	<% 1
33	event.academiainformationstechnology.org İnternet Kaynağı	<% 1
34	jurnal.itscience.org İnternet Kaynağı	<% 1
35	newsghana.com.gh İnternet Kaynağı	<% 1
36	www.nature.com İnternet Kaynağı	<% 1
37	"Coronavirus Therapeutics – Volume II", Springer Science and Business Media LLC, 2021 Yayın	<% 1
38	Kapil Kumar Malviya. "SARS-CoV-2 infection: Pathogenesis, Immune Responses, Diagnosis", Journal of Pure and Applied Microbiology, 2022 Yayın	<% 1
39	Massimiliano Chetta, Marina Tarsitano, Maria Oro, Maria Riveccio, Nenad Bukvic. "An in silico pipeline approach uncovers a potentially intricate network involving spike SARS-CoV-2 RNA, RNA vaccines, host RNA-binding proteins	<% 1

(RBPs), and host miRNAs at the cellular level",
Journal of Genetic Engineering and
Biotechnology, 2022

Yayın

-
- 40** Pedro F.N. Souza, Felipe P. Mesquita, Jackson L. Amaral, Patrícia G.C. Landim et al. "The human pandemic coronaviruses on the show: The spike glycoprotein as the main actor in the coronaviruses play", International Journal of Biological Macromolecules, 2021
Yayın <% 1
-
- 41** Shahzaib Ahamad, Kanipakam Hema, Shahnawaz Ahmad, Vijay Kumar, Dinesh Gupta. "Insights into the structure and dynamics of SARS-CoV-2 spike glycoprotein double mutant L452R-E484Q", 3 Biotech, 2022
Yayın <% 1
-
- 42** Thais Q. Morcatty, Kim Feddema, K.A.I. Nekaris, Vincent Nijman. "Online trade in wildlife and the lack of response to COVID-19", Environmental Research, 2020
Yayın <% 1
-
- 43** bmcp psychiatry.biomedcentral.com
İnternet Kaynağı <% 1
-
- 44** dergipark.org.tr
İnternet Kaynağı <% 1
-
- 45** docksci.com
İnternet Kaynağı

		<%1
46	dokumen.pub İnternet Kaynağı	<%1
47	ejim.springeropen.com İnternet Kaynağı	<%1
48	en.wikipedia.org İnternet Kaynağı	<%1
49	ileta.org İnternet Kaynağı	<%1
50	ijsk.org İnternet Kaynağı	<%1
51	revistas.uees.edu.ec İnternet Kaynağı	<%1
52	tfb.ba İnternet Kaynağı	<%1
53	uwstartcenter.org İnternet Kaynağı	<%1
54	www.besjournal.com İnternet Kaynağı	<%1
55	www.biorxiv.org İnternet Kaynağı	<%1
56	www.cdc.gov İnternet Kaynağı	<%1

- 57 Firuz Kamalov, Khairan Rajab, Aswani Cherukuri, Ashraf Elnagar, Murodbek Safaraliev. "Deep Learning for Covid-19 Forecasting: state-of-the-art review.", *Neurocomputing*, 2022
Yayın <% 1
-
- 58 Olayinka Sunday Okoh, Nicholas Israel Nii-Trebi, Abdulrokeeb Jakkari, Tosin Titus Olaniran et al. "Epidemiology and genetic diversity of SARS-CoV-2 lineages circulating in Africa", *iScience*, 2022
Yayın <% 1
-
- 59 Sandrine M. Soh, Yeongjun Kim, Chanwoo Kim, Ui Soon Jang, Hye-Ra Lee. "The rapid adaptation of SARS-CoV-2—rise of the variants: transmission and resistance", *Journal of Microbiology*, 2021
Yayın <% 1
-
- 60 Vipul Kumar, Jasdeep Singh, Seyed E. Hasnain, Durai Sundar. "Possible Link between Higher Transmissibility of Alpha, Kappa and Delta Variants of SARS-CoV-2 and Increased Structural Stability of Its Spike Protein and hACE2 Affinity", *International Journal of Molecular Sciences*, 2021
Yayın <% 1
-
- 61 Young-Il Kim, Mark Anthony B. Casel, Young Ki Choi. "Transmissibility and pathogenicity of <% 1

SARS-CoV-2 variants in animal models",
Journal of Microbiology, 2022

Yayın

62 "Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis", Springer Science and Business Media LLC, 2022

<% 1

Yayın

63 "Integrated Risk of Pandemic: Covid-19 Impacts, Resilience and Recommendations", Springer Science and Business Media LLC, 2020

<% 1

Yayın

64 Rossana Segreto, Yuri Deigin. "The genetic structure of SARS.CoV.2 does not rule out a laboratory origin", BioEssays, 2020

<% 1

Yayın

65 Severino Jefferson Ribeiro da Silva, Jessica Catarine Frutuoso do Nascimento, Renata Pessôa Germano Mendes, Klarissa Miranda Guarines et al. "Two Years into the COVID-19 Pandemic: Lessons Learned", ACS Infectious Diseases, 2022

<% 1

Yayın

66 Seyed Mostafa Biazar, Ahmad Fakheri Fard, Vijay P. Singh, Yagob Dinpashoh, Abolfazl Majnooni-Heris. "Estimation of Evaporation

<% 1

from Saline-Water with More Efficient Input Variables", Pure and Applied Geophysics, 2020

Yayın

67 Varun Gupta, Dmitry Ivanov, TsanMing Choi. "Competitive Pricing of Substitute Products under Supply Disruption", Omega, 2020 <% 1

Yayın

68 Wajihul Hasan Khan, Zohra Hashmi, Aditya Goel, Razi Ahmad et al. "COVID-19 Pandemic and Vaccines Update on Challenges and Resolutions", Frontiers in Cellular and Infection Microbiology, 2021 <% 1

Yayın

[Alıntılanı çıkart](#)

[Kapat](#)

[Eşleşmeleri çıkar](#)

< 5 words

[Bibliyografyayı Çıkart](#) üzerinde

CURRICULUM VITAE

1. PERSONAL INFORMATION

NAME, SURNAME:	Zurki Ibrahim
DATE of BIRTH and PLACE:	08/08/1984 Jigawa State, Nigeria
CURRENT OCCUPATION:	Student
ADDRESS of CORRESPONDENCE:	Dutse, Jigawa State, Nigeria
TELEPHONE: / Interphone:	+2348068322809
E-MAIL:	zurkiibrahim@yahoo.com

2. EDUCATION

YEAR	GRADE	UNIVERSITY	FIELD
2010	Undergraduate	Kano University of Science and Technology, Wudil	Applied Biology
2017	Masters in Science	Near East University	Medical Biology and Genetics

3. ACADEMIC EXPERIENCE

PERIOD	TITLE	DEPARTMENT	UNIVERSITY
2012 to date	Academic Staff	Biological Sciences	Sule Lamido University, Kafin Hausa, Nigeria.

4. FIELD OF INTERESTS

FIELDS OF INTERESTS	KEYWORDS
Biological Modeling	MacModeling, Epidemiology, Infectious and Genetic Diseases.
Statistical and Population Genetics	Frequency, Alleles, Genetic Variations