Seba Susan

Design and Development of Time Series Forecasting Models for COVID-19 Prediction

Delhi Technological University

Document Details

Submission ID
trn:oid:::27535:98713186

Submission Date
May 31, 2025, 8:56 PM GMT+5:30

Download Date
May 31, 2025, 9:01 PM GMT+5:30

File Name
thesis-WithoutTitlehead.pdf

File Size
4.5 MB

157 Pages

52,071 Words

284,910 Characters

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Small Matches (less than 10 words)

Exclusions

- 8 Excluded Sources

Match Groups

268 Not Cited or Quoted 7%
Matches with neither in-text citation nor quotation marks

31 Missing Quotations 1%
Matches that are still very similar to source material

3 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
5% Publications
4% Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

Replaced Characters
35 suspect characters on 13 pages
Letters are swapped with similar characters from another alphabet.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **268** Not Cited or Quoted 7%
  Matches with neither in-text citation nor quotation marks
- **31** Missing Quotations 1%
  Matches that are still very similar to source material
- **3** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 5% Publications
- 4% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	Liverpool John Moores University on 2022-03-01 <1%
2	Internet	pubmed.ncbi.nlm.nih.gov <1%
3	Internet	www.esp.org <1%
4	Publication	Nghiem Van Tinh, Nguyen Cong Dieu. "A NEW HYBRID FUZZY TIME SERIES FOREC... <1%
5	Internet	vdoc.pub <1%
6	Publication	Samit Bhanja, Banani Ghose, Abhishek Das. "Multi-Step-Ahead Time Series Foreca... <1%
7	Publication	Sibarama Panigrahi, H.S. Behera. "A study on leading machine learning technique... <1%
8	Internet	conservancy.umn.edu <1%
9	Publication	Abdelkader Dairi, Fouzi Harrou, Ying Sun, Sofiane Khadraoui. "Short-Term Foreca... <1%
10	Publication	Niteesh Kumar, Harendra Kumar, Kamal Kumar. "A Study for Plausible Third Wav... <1%

11 Submitted works
University of Auckland on 2021-09-19 <1%

12 Internet
www.mdpi.com <1%

13 Submitted works
Middle Tennessee State University on 2023-05-22 <1%

14 Submitted works
University of Southampton on 2024-09-08 <1%

15 Internet
ipfs.io <1%

16 Internet
jurnal.unprimdn.ac.id <1%

17 Publication
Nora El-Rashidy, Samir Abdelrazik, Tamer Abuhmed, Eslam Amer, Farman Ali, Jon... <1%

18 Submitted works
University of Newcastle on 2020-11-18 <1%

19 Submitted works
University of Witwatersrand on 2022-12-09 <1%

20 Internet
www.medrxiv.org <1%

21 Internet
zaguan.unizar.es <1%

22 Submitted works
University of Liverpool on 2022-04-15 <1%

23 Internet
www.degruyter.com <1%

24 Publication
"International Conference on Innovative Computing and Communications", Sprin... <1%

25 Publication
Derickson, Ian Gabriel. "Analysis of COVID-19 in Rural America", University of Mo... <1%

26 Publication
Mahua Bose, Kalyani Mali. "Designing fuzzy time series forecasting models: A sur... <1%

27 Submitted works
IUBH - Internationale Hochschule Bad Honnef-Bonn on 2022-02-28 <1%

28 Publication
Yang Ye, Abhishek Pandey, Carolyn Bawden, Dewan Md. Sumsuzzman et al. "Inte... <1%

29 Internet
assets.researchsquare.com <1%

30 Internet
dokumen.pub <1%

31 Publication
"Proceedings of International Conference on Communication and Computational... <1%

32 Publication
Hossein Abbasimehr, Reza Paki, Aram Bahrini. "Improving the performance of de... <1%

33 Submitted works
University of Teesside on 2022-05-04 <1%

34 Publication
Xunying Zhao, Xueyao Wu, Jinyu Xiao, Li Zhang, Yu Hao, Chenghan Xiao, Ben Zha... <1%

35 Submitted works
Liverpool John Moores University on 2022-11-29 <1%

36 Submitted works
National Institute of Technology, Silchar on 2023-09-30 <1%

37 Submitted works
The University of Buckingham on 2023-12-04 <1%

38 Publication
Fengqian Ding, Chao Luo. "Interpretable cognitive learning with spatial attention... <1%

39 Publication

Firuz Kamalov, Khairan Rajab, Aswani Kumar Cherukuri, Ashraf Elnagar, Murodbe... <1%

40 Publication

Sourabh Shastri, Kuljeet Singh, Monu Deswal, Sachin Kumar, Vibhakar Mansotra. ... <1%

41 Publication

Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical ... <1%

42 Submitted works

Universite de Sfax on 2024-12-07 <1%

43 Publication

Yingyu Yin, Iman Ahmadianfar, Faten Khalid Karim, Hela Elmannai. "Advanced for... <1%

44 Internet

link.springer.com <1%

45 Internet

www.ijert.org <1%

46 Internet

www.jmess.org <1%

47 Publication

"Assessing COVID-19 and Other Pandemics and Epidemics using Computational ... <1%

48 Submitted works

Liverpool John Moores University on 2023-05-28 <1%

49 Publication

Nghiem Van Tinh. "Chapter 3 A New Refined Forecasting Model Based on the Hig... <1%

50 Publication

Shashi Kant, Devendra Agarwal, Praveen Kumar Shukla. "Chapter 2 Improving th... <1%

51 Publication

Somayeh Bakhtiari Ramezani, Amin Amirlatifi, Shahram Rahimi. "A novel compar... <1%

52 Submitted works

University of Bristol on 2023-08-30 <1%

53 Publication
Wafa F. Alfwzan, Kinda Abuasbeh, Ali Raza, Zunair Zeb, Muath Awadalla, Norah Al... <1%

54 Internet
philpapers.org <1%

55 Internet
uwspace.uwaterloo.ca <1%

56 Submitted works
Institute of Graduate Studies, UiTM on 2017-03-09 <1%

57 Submitted works
Universitat Politècnica de València on 2021-07-19 <1%

58 Publication
Aindrila Saha, Vartika Mishra, Santanu Kumar Rath. "Prediction of growth in COVI... <1%

59 Submitted works
University of Leeds on 2022-09-12 <1%

60 Submitted works
Coventry University on 2016-01-15 <1%

61 Submitted works
Sardar Vallabhbhai National Inst. of Tech.Surat on 2023-06-14 <1%

62 Submitted works
University of Ulster on 2025-05-16 <1%

63 Submitted works
IIT Delhi on 2015-03-27 <1%

64 Publication
Kuldeep Dhama, Firzan Nainu, Andri Frediansyah, Mohd. Iqbal Yatoo et al. "Globa... <1%

65 Submitted works
The Scientific & Technological Research Council of Turkey (TUBITAK) on 2023-06-06 <1%

66 Publication
Wandera Ogana, Victor Ogesa Juma, Wallace D. Bulimo, Vincent Nandwa Chiteri. ... <1%

67 Publication
Abdalla, Ahmad. "A Trustworthy Deep Reinforcement Learning Framework for Sli... <1%

68 Submitted works
Northcentral on 2021-12-23 <1%

69 Publication
Yuri Zelenkov, Ivan Reshettsov. "Analysis of the COVID-19 pandemic using a com... <1%

70 Internet
dmame.rabek.org <1%

71 Internet
export.arxiv.org <1%

72 Publication
Ahmad Daryanto. "An Introduction to Quantitative Research Methods for Market... <1%

73 Publication
J.G. Carvalho, C.T. Costa. "Non-iterative procedure incorporated into the fuzzy ide... <1%

74 Publication
Reza Rajabi, Siamak Haji Yakhchali. "Optimizing cash flow in construction portfoli... <1%

75 Publication
Schlieper, Philipp. "Machine Learning for Industry 4.0: Methods and Applications ... <1%

76 Submitted works
The African Institute for Mathematical Sciences on 2023-05-14 <1%

77 Submitted works
University of Liverpool on 2022-04-27 <1%

78 Submitted works
University of St. Gallen on 2021-12-16 <1%

79 Internet
repository.netecweb.org <1%

80 Publication
Ahmed, Supriyo Shafkat. "Machine Learning-Driven Demand Prediction to Minimi... <1%

81 Submitted works
Maharishi University Of Information Technology, Lucknow on 2025-01-16 <1%

82 Publication
Ping Jiang, Hufang Yang, Ranran Li, Chen Li. "Inbound tourism demand forecasti... <1%

83 Publication
Samit Bhanja, Abhishek Das. "An air quality forecasting method using fuzzy time ... <1%

84 Submitted works
University of Glasgow on 2025-03-26 <1%

85 Internet
covid19evidencereviews.saskhealthauthority.ca <1%

86 Submitted works
universititeknologimara on 2024-12-30 <1%

87 Submitted works
wunu on 2025-03-15 <1%

88 Internet
www2.mdpi.com <1%

89 Publication
Anil Utku. "Deep Learning Based Hybrid Prediction Model for Predicting the Spre... <1%

90 Submitted works
Anna University on 2024-07-08 <1%

91 Submitted works
Aston University on 2022-10-14 <1%

92 Submitted works
Dundalk Institute of Technology on 2022-06-13 <1%

93 Publication
Hossein Abbasimehr, Amirreza Behboodi, Aram Bahrini. "A novel hybrid model to... <1%

94 Publication
Jayanthi Devaraj, Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, G.M. Shafiull... <1%

95 Publication
K.E. ArunKumar, Dinesh V. Kalaga, Ch. Mohan Sai Kumar, Masahiro Kawaji, Timot... <1%

96 Submitted works
Liverpool John Moores University on 2022-07-15 <1%

97 Submitted works
Liverpool John Moores University on 2023-10-25 <1%

98 Publication
Mujtaba Haidari. "Prediction of COVID-19 Cases in Afghanistan Using ARIMA mod... <1%

99 Publication
Radha Mohan Pattanayak, Sibarama Panigrahi, H. S. Behera. "High-Order Fuzzy Ti... <1%

100 Publication
Ruobin Gao, Okan Duru, Kum Fai Yuen. "High-dimensional lag structure optimiza... <1%

101 Publication
Satya Prakash, Anand Singh Jalal, Pooja Pathak. "Infectious Disease Time Series ... <1%

102 Submitted works
Universiti Kebangsaan Malaysia on 2024-06-24 <1%

103 Submitted works
University of Greenwich on 2023-04-28 <1%

104 Submitted works
University of Kent at Canterbury on 2025-04-03 <1%

105 Submitted works
University of Nottingham on 2024-09-02 <1%

106 Publication
Xiao Ning, Linlin Jia, Yongyue Wei, Xi-An Li, Feng Chen. "Epi-DNNs: Epidemiologica... <1%

107 Publication
Zhanybai T. Zhusubaliyev, Alexander N. Churilov, Alexander Medvedev. "Complex... <1%

108 Internet
iris.polito.it <1%

109 Internet

www.wjgnet.com <1%

110 Publication

"Advances in Artificial Intelligence -- IBERAMIA 2014", Springer Science and Busin... <1%

111 Publication

"Artificial Intelligence and Soft Computing", Springer Science and Business Media... <1%

112 Publication

"Computational Intelligence in Data Mining", Springer Science and Business Medi... <1%

113 Publication

"Recent Advances on Hybrid Intelligent Systems", Springer Science and Business ... <1%

114 Publication

Amit Kumar Tyagi, Shrikant Tiwari. "AI and Blockchain in Smart Grids - Fundamen... <1%

115 Submitted works

Associatie K.U.Leuven on 2025-05-21 <1%

116 Submitted works

Birkbeck College on 2025-05-22 <1%

117 Publication

Cheng, C.H.. "Entropy-based and trapezoid fuzzification-based fuzzy time series ... <1%

118 Submitted works

De Montfort University on 2021-02-15 <1%

119 Publication

Douglas Martins, Amit Bhaya, Fernando Pazos. "N-Step-Ahead Optimal Control of ... <1%

120 Submitted works

Erasmus University of Rotterdam on 2024-02-01 <1%

121 Publication

Gerrity, Thomas. "STRAINER: State Transcript Rating for Informed News Entity Re... <1%

122 Submitted works

Higher Education Commission Pakistan on 2018-06-04 <1%

123 Submitted works
Higher Education Commission Pakistan on 2023-07-19 <1%

124 Submitted works
Imperial College of Science, Technology and Medicine on 2021-06-04 <1%

125 Publication
Jair Paulino de Sales, Paulo S.G. de Mattos Neto, Paulo R.A. Firmino. "A dynamic e... <1%

126 Publication
Liu, Qing. "Quantifying COVID-19: Modeling and Evaluation", University of Techno... <1%

127 Submitted works
Liverpool John Moores University on 2023-03-13 <1%

128 Publication
Long, Jie. "Deep Learning Algorithms for Time-Dependent Partial Differential Equ... <1%

129 Publication
Mahda Delshad, Mohammad-Javad Sanaei, Atieh Pourbagheri-Sigaroodi, Davood ... <1%

130 Submitted works
Manchester Metropolitan University on 2024-10-01 <1%

131 Publication
Manotosh Mandal, Soovoojeet Jana, Swapan Kumar Nandi, Anupam Khatua, Saya... <1%

132 Submitted works
Queen's University of Belfast on 2016-12-02 <1%

133 Publication
Shaojing Zhou, Jinguo Li, Kai Zhang, Mi Wen, Qijie Guan. "An Accurate Ensemble F... <1%

134 Publication
Sujata Dash, Subhendu Kumar Pani, Joel J. P. C. Rodrigues, Babita Majhi. "Deep Le... <1%

135 Submitted works
Universiti Malaysia Pahang on 2016-04-25 <1%

136 Submitted works
University of Stirling on 2024-09-06 <1%

137 Submitted works
University of Western Macedonia on 2023-06-27 <1%

138 Publication
Vijay Kumar, Dilbag Singh, Manjit Kaur, Robertas Damaševičius. "Overview of cur... <1%

139 Publication
Vincent S. Tseng, Josh Jia-Ching Ying, Stephen T.C. Wong, Diane J. Cook, Jiming Liu... <1%

140 Submitted works
Xi'an Jiaotong-Liverpool University on 2024-04-30 <1%

141 Publication
Zongxi Qu, Yutong Li, Xia Jiang, Chunhua Niu. "An innovative ensemble model ba... <1%

142 Internet
iceb.johogo.com <1%

143 Internet
repositorio.ufmg.br <1%

144 Internet
studylib.net <1%

145 Internet
www.journal-aquaticscience.com <1%

146 Internet
www.nature.com <1%

147 Internet
www.uyik.org <1%

148 Publication
"Advances in Swarm Intelligence", Springer Nature, 2016 <1%

149 Publication
"Artificial Intelligence and Machine Learning Methods in COVID-19 and Related H... <1%

150 Publication
"Computer Vision – ECCV 2024 Workshops", Springer Science and Business Media ... <1%

151 Publication
Abbas Rajabifard, Greg Foliente, Daniel Paez. "COVID-19 Pandemic, Geospatial Inf... <1%

152 Publication
Ahmed Gad, Wafaa I. M. Ibrahim, Abdelnaser S. Abdrabou. "A Mixed Effects Chan... <1%

153 Publication
Aminatus Sa'adah, Roberd Saragih, Dewi Handayani. "Optimal control design of t... <1%

154 Publication
Amit Kumar Tyagi, Ajith Abraham. "Recurrent Neural Networks", CRC Press, 2022 <1%

155 Publication
Andrew B. Lawson, Sudipto Banerjee, Robert P. Haining, Maria Dolores Ugarte. "H... <1%

156 Publication
Arthur C. Vargas Pinto, Larissa C. C. da Silva, Petrônio C. L. Silva, Frederico G. Gui... <1%

157 Submitted works
Bahcesehir University on 2023-12-28 <1%

158 Publication
Barenya Bikash Hazarika, Deepak Gupta. "Modelling and forecasting of COVID-19... <1%

159 Publication
Braimah Joseph Odunayo, Edike Nnamdi, Fabio Mathias Correa. "Initial Growth R... <1%

160 Publication
Budati Anil Kumar, Akella Ramakrishna, Goutham Makkena, Gheorghita Ghinea. ... <1%

161 Publication
Chen, S.M.. "Multivariate fuzzy forecasting based on fuzzy time series and autom... <1%

162 Publication
Cheng Cheng, Elayaraja Aruchunan, Muhamad Hifzhudin Noor Aziz. "Leveraging ... <1%

163 Submitted works
Chiang Mai University on 2024-05-17 <1%

164 Publication
Chitwadgi, Basavaraj Sanjiv. "Manufacturing Production Demand Forecasting Usi... <1%

165 Publication
Damayanthi Dahanayake, Miruna Rabindrakumar. "Transformative Applied Rese... <1%

166 Publication
Ellen Kuhl. "Computational Epidemiology", Springer Science and Business Media ... <1%

167 Submitted works
Federal University of Technology on 2023-12-04 <1%

168 Publication
Forster, Ryan. "Environment First: Inferring Adaptation to Environmental Pressur... <1%

169 Publication
Gihan Jayatilaka, Jameel Hassan, Umar Marikkar, Rumali Perera et al. "Use of Arti... <1%

170 Publication
Heu, Cherryle. "Characterizing Patterns and Identifying Prediction Models to Sim... <1%

171 Publication
Hossein Abbasimehr, Reza Paki. "Prediction of COVID-19 confirmed cases combini... <1%

172 Submitted works
International Maritime College, Oman on 2019-05-22 <1%

173 Publication
Junling Luo, Zhongliang Zhang, Yao Fu, Feng Rao. "Time series prediction of COVI... <1%

174 Publication
Kasasbeh, Anemone Ahmad. "Applying Artificial Intelligence and Machine Learni... <1%

175 Publication
Lecture Notes in Computer Science, 2005. <1%

176 Publication
Lee, L.W.. "Temperature prediction and TAIFEX forecasting based on fuzzy logical ... <1%

177 Publication
Li, S.T.. "Deterministic fuzzy time series model for forecasting enrollments", Com... <1%

178 Submitted works
Liverpool John Moores University on 2023-11-30 <1%

179 Submitted works
Madan Mohan Malaviya University of Technology on 2019-07-07 <1%

180 Publication
Mario Muñoz-Organero, Patricia Callejo, Miguel Ángel Hombrados-Herrera. "A ne... <1%

181 Publication
Mathonsi, Thabang. "Anomaly Detection Using Time Series Forecasting with Dee... <1%

182 Publication
Mohd Sakib, Tamanna Siddiqui. "Multi-Network-Based Ensemble Deep Learning ... <1%

183 Submitted works
Northern Marianas College on 2023-02-18 <1%

184 Publication
Ogundare, Timilehin. "Surrogate Assisted by Genetic Algorithm for Simulation of ... <1%

185 Publication
Rasul Enayatifar, Hossein Javedani Sadaei, Abdul Hanan Abdullah, Abdullah Gani. ... <1%

186 Publication
Roberto De Vogli. "Managing and Preventing Pandemics - Lessons From COVID-1... <1%

187 Submitted works
Rutgers University, New Brunswick on 2021-03-01 <1%

188 Publication
Sharif Naser Makhadmeh, Mohammed Azmi Al-Betar, Iyad Abu Doush, Mohamm... <1%

189 Publication
Studies in Computational Intelligence, 2014. <1%

190 Publication
Sujin Ahn, Minhae Kwon. "Reproduction Factor Based Latent Epidemic Model Infe... <1%

191 Submitted works
Telkom University on 2025-03-13 <1%

192 Submitted works
Telkom University on 2025-04-28 <1%

193 Submitted works
University College London on 2020-10-13 <1%

194 Submitted works
University of Birmingham on 2023-09-10 <1%

195 Submitted works
University of Essex on 2023-11-24 <1%

196 Submitted works
University of Essex on 2025-03-21 <1%

197 Submitted works
University of Greenwich on 2023-01-16 <1%

198 Submitted works
University of London External System on 2021-09-09 <1%

199 Submitted works
University of Melbourne on 2020-10-02 <1%

200 Submitted works
University of Nottingham on 2017-04-02 <1%

201 Submitted works
University of Oklahoma on 2024-01-10 <1%

202 Submitted works
University of Oxford on 2021-04-28 <1%

203 Submitted works
University of Wales Swansea on 2024-04-15 <1%

204 Submitted works
Van Lang University on 2023-07-18 <1%

205 Publication
Venkatesh Ambalarajan, Ankamma Rao Mallela, Prasantha Bharathi Dhandapani,... <1%

206 Publication
Yasminah Alali, Fouzi Harrou, Ying Sun. "A proficient approach to forecast COVID-... <1%

207 Internet
aisberg.unibg.it <1%

208 Internet
cs.uni-muenster.de <1%

209 Internet
dspace.univ-setif.dz:8888 <1%

210 Internet
ecmtb2024.org <1%

211 Internet
ijetst.in <1%

212 Internet
ir.uitm.edu.my <1%

213 Internet
jbpe.ssau.ru <1%

214 Internet
journals.plos.org <1%

215 Internet
mural.maynoothuniversity.ie <1%

216 Internet
proceedings.neurips.cc <1%

217 Internet
repository.uin-suska.ac.id <1%

218 Internet
stax.strath.ac.uk <1%

219 Internet
web.archive.org <1%

220 Internet
www.journal-fea.com <1%

Design and Development of Time Series Forecasting Models for COVID-19 Prediction

Naresh Kumar

ABSTRACT

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) triggered the COVID-19 pandemic, which became a global health crisis with severe impacts on humanity. The first wave of COVID-19 was reported in most countries at the beginning of 2020, and the World Health Organization (WHO) declared it a pandemic on March 11, 2020. During the early phase of the pandemic, most countries relied on non-pharmaceutical interventions, such as bans on international travel, mandatory face masks, quarantine protocols, contact tracing, and complete lockdowns, to curb the spread of the virus. Governments faced numerous challenges, including educating the public about the pandemic, managing resources, ensuring medical facilities, and addressing economic impacts. Therefore, identifying future cases and predicting the spread of the virus became critical for healthcare systems to take proactive measures and minimize casualties. Consequently, predictive analysis of pandemics emerged as a vital research area, aiding healthcare services and governments in planning and controlling the spread of COVID-19. The highly contagious SARS-CoV-2 virus spread rapidly and evolved into numerous mutants and variants, leading to second and third waves of infections across many countries. Over time, vaccines were developed to mitigate the impact of COVID-19, adding further complexity to the dynamics of COVID-19 time series data. This has underscored the importance of developing models capable of handling highly dynamic and non-stationary data for accurate time series forecasting associated with multiple waves of the pandemic driven by successive mutations of the virus.

In this thesis various aspects of the COVID-19 pandemic related to time series forecasting and modeling are analyzed using state-of-the-art methods and novel forecasting models. Initially, the well-known forecasting models namely, Autoregressive Integrated Moving Average (ARIMA), Facebook Prophet (FB-Prophet), Exponential smoothing models (ETS), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM) are evaluated and compared using diverse datasets spanning different timelines of the pandemic. LSTM outperformed all the other compared models for the time series forecasting of the COVID-19 cases. Fuzzy time series (FTS) models are particularly effective at handling uncertain and imprecise time series data, particularly when the underlying patterns are nonlinear. Evolutionary optimization has proven to be a powerful approach for hyperparameter tuning, achieving strong results in complex problem-solving with minimal computational expense. There are three main hyperparameters of a FTS model:- i) number of intervals, ii) length of intervals, iii) fuzzy order. Therefore, two new algorithms based on FTS with hyperparameter optimization using PSO are proposed, namely, nested-FTS-PSO, and exhaustive-search-FTS-PSO. The forecasting results of the algorithms are compared with ARIMA,

FB-Prophet, FTS, and FTS-PSO models using COVID-19 cases from 10 highly affected countries. Exhaustive-search-FTS-PSO algorithm outperformed all the other compared models. The integration of fuzzy techniques with deep learning has emerged as a promising research area, as it enhances both interpretability and explainability of deep learning based systems. Consequently, the development of a hybrid fuzzy time series forecasting model that combines FTS, deep learning, and swarm intelligence is envisioned to achieve more accurate time series forecasting of dynamic, non-stationary data pertaining to multiple waves of the pandemic caused by successive mutations of the virus. Deep learning models, specifically, stacked-LSTM, bidirectional-LSTM, convolution-LSTM, attention-LSTM, attention-bi-LSTM are integrated with FTS and PSO. The hybrid models are compared with the ETS, ARIMA, ANN, LSTM, and FTSF-PSO models. Hybrid of FTS, PSO, and attention-bi-LSTM outperformed all the other compared models on the USA and India COVID-19 datasets.

Compartmental epidemiological models are among the most traditional and widely used approaches to represent the progression of an epidemic. Advancements in Artificial Intelligence (AI) hold significant potential to aid in combating pandemics. Fully leveraging the capabilities of AI, epidemiological modeling, and optimization techniques in an integrated forecasting solution is crucial for predicting the impact of a pandemic. Therefore, Susceptible-Infected-Recovered-Deceased (SIRD) epidemiological model integrating with PSO and deep learning (stacked-LSTM) is proposed to model the evolution of the COVID-19 pandemic in India, UK and the USA. Time-varying model parameters are used to deal with multiple waves of the COVID-19. The proposed hybrid model outperformed the stacked-LSTM and hybrid of SIRD and PSO. Further, a novel epidemiological compartmental model, which provides realistic projections of epidemic spread based on viral characteristics, is proposed, that incorporates time-varying hyper-parameters and deep learning models. The COVID-19 time series data is highly dynamic in nature due to rapidly changing transmission rates and government policy measures such as lockdowns and vaccination campaigns. Recognizing the numerous factors influencing epidemic spread, a 10-compartmental epidemiological model is presented, incorporating restriction policies, multi-dose vaccinations, and vaccine efficacy. COVID-19 case studies of the USA and India are carried out for demonstrating the efficacy of the proposed approach. The proposed approach outperformed ETS, ARIMA, ANN, LSTM, and SEIRD models in the performance evaluation.

Future work may enhance the forecasting accuracy of the proposed models by integrating advanced optimization algorithms and deep learning techniques. Time series forecasting models have the potential to drive changes in the society by enhancing decision-making processes and optimizing resource allocation across various industries. From this perspective, the models introduced in this thesis can be utilized across various fields, including economics, healthcare, and public policy, which can lead to significant positive societal outcomes.

Table of Contents

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 COVID-19 Pandemic

Humanity has encountered numerous epidemics and pandemics in the history [78]. The COVID-19 is a recent pandemic that emerged in December 2019, and quickly escalated into a global health crisis [239]. Coronaviruses (CoV), known for causing high-mortality diseases in humans, were first identified in 2002, with bats recognized as their ecological origin [163]. The Severe Acute Respiratory Syndrome coronavirus 2 (??) belongs to the coronavirus family which is responsible for the COVID-19 pandemic. The common symptoms of the virus are respiratory issues, fever, dry cough, fatigue, sore throat, body aches, headaches, and, in some cases, diarrhea, nausea, and loss of taste or smell. In severe instances, it can lead to death [177]. The World Health Organization (??) officially declared COVID-19 a pandemic on March 11, 2020 [96]. The disease, caused by SARS-CoV-2, is characterized by a highly replicative genome in human cells, enabling rapid spread [248, 225]. Since its emergence in December 2019, the novel coronavirus has infected billions of people globally, and claiming the lives of millions of people [207]. The transmission of COVID-19 occurs in three primary stages which is described below.

- **Local outbreak**: During this stage, the chain of virus transmission is traceable, and the source of infection can be identified. Most cases are confined to close contacts, such as family or friends, or result from local exposure.
- **Community transmission**: At this stage, the source of infection becomes untraceable, and cases multiply through clusters within communities.
- **Large-scale transmission**: This stage is marked by the rapid spread of the virus to different regions, driven by widespread and uncontrolled human mobility.

COVID-19 exhibits a high mutation rate and spreads rapidly. Infected people from this virus experience severe respiratory problems. Specifically, people who is suffering from chronic diseases like diabetes or cardiovascular disease or having weakened immune system [207]. Controlling the disease poses significant challenges, as infected individuals may remain asymptomatic or show symptoms only after a long time period. Measures such as social distancing, widespread testing to identify positive cases, and isolating infected individuals are crucial to prevent the spreading of the virus [80]. The first wave of COVID-19 was reported in most countries at the beginning of the year 2020 [3]. Initially, limited knowledge about the virus posed challenges in combating its spread. Therefore, governments implemented preventive measures such as ban on international flights, mandatory face masking,

isolation, quarantine, contact tracing, thermal scanning, and complete lockdowns. During the early phase of the pandemic, non-pharmaceutical interventions were the primary strategies to prevent the spread, including social distancing, quarantine protocols, market closures, and restrictions on public gatherings [170].

Governments faced numerous challenges, including educating the public, managing resources, sustaining the economy, and providing adequate medical facilities. Medical centers were overwhelmed during the peak of the pandemic, necessitating the expansion of emergency and intensive care units to accommodate the growing number of patients. Various restriction policies were enforced to control the spread. Extensive research was accelerated to develop vaccines against the SARS-CoV-2. As vaccines were developed, vaccination programs were rolled out in phased manner based on priority and availability. Over the course of the pandemic, SARS-CoV-2 mutated into several variants, leading to multiple waves [79, 74]. These variants raised concerns due to their higher transmission rates [164]. Increased infectiousness or resistance to immunity by an variant further complicated efforts to control the virus. Further, the spread of COVID-19 was influenced by government-imposed restrictions, vaccination programs, healthcare infrastructure, and other related factors. WHO declared on May 4, 2023, that COVID-19 was no longer a global health emergency [126].

1.2 Time series forecasting

Time series forecasting (??) is used to predict future outcome in a sequence of data points based on past observed values. It involves analysis of seasonality, past trends, and patterns in the data to make predictions. TSF is widely used in finance, weather forecasting, energy demand prediction, sales and demand forecasting, healthcare, operational planning, and many more [34, 137]. Key challenges in TSF includes dealing with temporal dependency, stationarity, data noise, outliers, data sparsity, missing values, multivariate data, trend and seasonality [128, 144, 101].

TSF problems have been studied widely in the literature in which COVID-19 forecasting had emerged as one of the complex problem [14]. Time series data of the COVID-19 cases is highly dynamic in nature due to involvement of the multiple factors [81]. The data can be utilized to analyze the spread pattern, and perform the predictive analysis. Such predictions are very useful for shaping policies to control the outbreak of a pandemic. In this direction, numerous studies have been conducted to forecast the spread of COVID-19 using various predictive models [81]. A number of researchers have investigated the baseline prediction models using different datasets from different data sources. There are many aspects of the COVID-19 which have been explored by the researchers such as different training, testing, and forecasting samples in each study. Forecasting models can predict likely impacts of the disease on communities, aiding in epidemic control efforts.

In recent years, fuzzy time series (??) modeling has been adopted in many forecasting studies [244]. FTS offers greater descriptiveness than traditional time series by providing semantic insights into uncertain and fluctuating data. The accuracy of any forecasting model is critical for effective decision making and planning. Existing studies have primarily focused on optimizing one or two

hyperparameters of FTS to enhance forecasting accuracy [195, 42, 167]. FTS techniques have demonstrated improved prediction results in various applications due to use of the optimization techniques. Leveraging these techniques, it is possible to develop hybrid approaches which can improve forecasting results. After facing the global spread of infectious COVID-19, studying FTS forecasting techniques for a pandemic predictions is crucial. In some of the studies, FTS has been explored for forecasting COVID-19 cases [222, 62]. Further, experimentation with additional FTS hyperparameters could yield improved forecasting outcomes. Developing a prediction model with optimized hyperparameters can significantly aid in policy formulation and controlling the spread of the virus in the scenarios similar to the COVID-19.

1.3 Epidemiological compartmental modeling

Epidemiological compartmental models are the most common models that are usually used for modeling the spread of an epidemic or a pandemic [75]. Epidemiological modeling involves dividing a population into distinct compartments to study the progression and spread of an epidemic [145]. In these models, the population is assigned to specific labels, such as susceptible, exposed, infected, suspected, asymptomatic, hospitalized, recovered, dead, etc. [51, 240]. Compartmental models are widely used to analyze the flow and spread patterns of outbreaks. Common baseline models include Susceptible-Infected-Removed (??)[98], Susceptible-Infected-Recovered-Deceased (??) [65], Susceptible-Exposed-Infected-Recovered (??) [215], and Susceptible-Exposed-Infected-Recovered-Deceased (??) [77]. These models use stochastic frameworks to forecast specific measures. These models describe how susceptible (S) individuals may be exposed (E) to a virus, become infected (I), and eventually either recover (R) or succumb (D) to the disease. The likelihood of exposure to infection depends on the contact rate among susceptible individuals. Most compartmental models focus on disease spread within a single population confined to a specific region. These models are unidirectional, as they assume that recovered individuals gain permanent immunity and cannot re-enter the susceptible population.

Numerous studies have utilized compartmental models to analyze the evolution of COVID-19 [160, 198]. Extensions of baseline epidemic models have been developed to more accurately represent the dynamics of the pandemic. However, most of these models assume time-invariant parameters such as infection and recovery rates [194]. Various techniques have been introduced to calibrate these models using COVID-19 time series data [71]. The spread of COVID-19 was influenced by numerous factors, resulting in diverse data types and information. The emergence of new variants and changes in governmental policies led to multiple waves of infection, necessitating the exploration of time-fused and time-variant models to address the dynamic nature of the pandemic. Following significant efforts by virologists worldwide, vaccines were developed in an unprecedentedly short time frame, providing a new avenue for research into the impact of vaccination programs in combating a pandemic. Designing and developing effective epidemiological models have become crucial for generating actionable insights and outcomes to

manage pandemics like COVID-19 effectively.

1.4 Artificial intelligence and optimization techniques

Artificial intelligence (??) assisted forecasting techniques play a vital role in alerting governments to potential health crises by leveraging available data. In the case of COVID-19, numerous factors influenced the spread of SARS-CoV-2, including environmental conditions (temperature, humidity, wind), mobility, lockdown measures, government policies, social distancing, isolation, age, gender, incubation period, mask usage, medical facilities, testing, and vaccination. Given these complexities, it has become increasingly important to address the dynamic nature of pandemics and provide precise, accurate, and timely risk predictions to mitigate casualties in crises like COVID-19. To this end, researchers have explored the potential of AI-integrated forecasting models [5]. Key parameters analyzed to assess the pandemic's impact include daily confirmed cases, daily deaths, daily recoveries, hospitalizations, transmission rates, impact of lockdowns, government policies, population density, and mobility [108, 28]. Researchers have employed predictive models to analyze the spread of COVID-19 during the early stages of the pandemic. These studies focused on modeling COVID-19 cases at various scales, including country, state, city, regional, and global levels, utilizing publicly available databases for model evaluation. Due to the dynamic nature of COVID-19, most research concentrated on short-term forecasting. Numerous machine learning techniques have been proposed, incorporating these factors to enhance prediction accuracy and align with real-world scenarios.

The literature presents a diverse range of Deep Learning (??), Machine Learning (??), and optimization techniques designed to analyze and forecast COVID-19 time series data. Researchers have leveraged ML techniques to study the spread patterns of the COVID-19 pandemic [246, 140]. Several models have been proposed for COVID-19 epidemic modeling using computational and machine learning techniques. For instance, [7] conducted forecasting of confirmed cases, deaths, and recoveries using Artificial Neural Networks (??) alongside five univariate time series models. Machine learning models excel at identifying patterns in data, enabling them to learn from known datasets and generalize to predict future, unseen scenarios. These capabilities support decision-making through predictive analysis of time series data and patterns. The ML modeling process typically involves several steps: data collection, preprocessing, model selection, training, evaluation with performance metrics, tuning, validation, and final prediction or output [56, 84]. Traditional optimization methods often require significant computational resources. Meta-heuristic and evolutionary algorithms, however, offer efficient alternatives for obtaining optimized solutions [119, 220, 191, 150].

Machine learning models were extensively explored to study the effects of lockdowns and restriction policies. Many of these studies combined machine learning or deep learning techniques with epidemiological models to assess the impact of governmental interventions. In the fight against COVID-19 and its

variants, several vaccines with varying efficacies were developed. Vaccination programs were implemented based on priorities and applicability, considering factors such as the number of doses, population coverage, vaccine availability, and medical infrastructure. These factors significantly influenced the forecasting performance of AI models. By integrating data from multiple domains and leveraging AI and optimization techniques, researchers can address the challenges posed by pandemics.

1.5 Hybrid time series forecasting models

Forecasting COVID-19 cases is a complex challenge that cannot be effectively addressed with a single model. AI has demonstrated significant potential in combating the pandemic through various impactful applications. AI offers viable solutions for tackling the spread of COVID-19 by enabling predictive analysis, which has become a critical research focus to support health services and governments in planning and mitigating the spread of the virus. However, numerous challenges in COVID-19 forecasting still require attention. Accurately modeling and predicting the daily spread patterns of the virus can equip healthcare systems to prepare for and manage the anticipated influx of patients effectively.

Hybrid models that integrate machine learning (ML) and computational intelligence (??) with epidemiological modeling are more powerful and efficient than standalone techniques. By combining the strengths of individual approaches, hybrid models address their limitations, enhancing robustness, accuracy, and overall performance [235, 127]. Researchers have explored various models for COVID-19 epidemic modeling using ML and computational methods [200, 111, 82]. These models have been tested on diverse datasets, yielding valuable insights to combat the pandemic. Outcomes include predictions of pandemic spread patterns, peak times, medical resource requirements, and the effects of restriction policies and vaccination efforts. Additionally, ensemble models utilizing techniques such as stacking, bagging, and boosting have been developed to further refine results and improve accuracy [70, 212].

Combining epidemiological models, ML, and computational intelligence can result in highly effective prediction models for pandemics akin to COVID-19. Forecasting infection cases is a critical research area, as it enables healthcare systems, governments, and societies to prepare for combating the virus. Modeling disease spread patterns and predicting their impacts are vital for optimizing planning and managing various services and resources, making this a key focus in pandemic research.

1.6 COVID-19 Variants and Data Sources

Throughout the course of the COVID-19 pandemic, the SARS-CoV-2 virus underwent multiple mutations, resulting in the emergence of several variants [79, 74]. Some variants exhibited increased transmissibility or the ability to evade vaccine-induced immunity [164, 32]. The continuous appearance of new variants has heightened the risk of a renewed public health crisis. These variants are primarily classified into three categories: variants of concern (VOC), variants of interest

(VOI), and variants being monitored (VBM). Details about these variants are available at online link [1], and an overview of SARS-CoV-2 variants of interest can be found at online link [2]. A summary of the well-known SARS-CoV-2 variants is provided in Table 1.1.

Table 1.1: Widely recognized variants of SARS-CoV-2.

WHO Label	Lineage	Status	Reported Country	Title Date	Description
Alpha	B.1.1.7 and Q lineages	VBM	United Kingdom	VOC:Dec 29, 2020; VBM:Sep 21, 2021	It was the first variant which spread quickly around the world.
Beta	B.1.351 and descendant lineages	VBM	South Africa	VOC:Dec 29, 2020; VBM:Sep 21, 2021	It was about 50% more contagious than the Alpha variant according to the CDC.
Gamma	P.1 and descendant lineages	VBM	Brazil	VOC:Dec 29, 2020; VBM:Sep 21, 2021	It was indicated that Gamma variant was 1.4 to 2.2 times more transmissible than Alpha variant.
Delta	B.1.617.2 and descendant lineages	VBM	India	VOC:Jun 15, 2021; VBM:Apr 14, 2022	It was 85% more transmissible than the Alpha variant.
Omicron	B.1.1.529 and descendant lineages	VOC	South Africa	VOC:Nov 26, 2021	It was less severe but more transmissible than the Delta variant.
'VOC': variant of concern, 'VBM': variant being monitored					

The SARS-CoV-2 virus has undergone numerous mutations resulting from genetic changes, giving rise to various variants. These variants have had profound impacts on vaccine efficacy, public health, and the spread patterns of COVID-19. They have demonstrated differences in transmissibility, virulence, and resistance to immune responses. Notably, the Delta and Omicron variants exhibited higher transmissibility, leading to significant surges in infection rates and contributing to increased mortality and morbidity.

A dataset is essential for evaluating the performance of a model. COVID-19 related data sources have played a major role in assessing models using real-world data. Most of the forecasting studies have used these data sources to analyze the spread patterns of COVID-19. The WHO maintained a dashboard to show real-time COVID-19 cases worldwide. It issued public advisories on symptoms, precautions, and updates of the COVID-19 regularly. It has played a major role to

[1] https://www.who.int/activities/tracking-SARS-Cov-2-variants
[2] https://covariants.org/

raising awareness among people. It has helped to contain the spread of the virus. Various organizations maintained publicly accessible data sources to support the research community. These data sources were updated regularly using various techniques ensuring high accuracy. Various statistical and graphical representation have been provided in the dashboard. Different datasets were created segregating the available data based on classifications such as COVID-19 cases of a country or state or region, SARS-Cov-2 variants, and vaccinations data. Researchers have used different datasets from these data sources to train and test their developed models. The proposed models in this research are evaluated using datasets from these data sources. The well-known COVID-19 data sources are included in Table 1.2. The table contains label, description, and publicly accessible url link of a data source.

Table 1.2: Publicly available well-known COVID-19 data sources.

Data Label	Description	Online Link
WHO-Dashboard	WHO dashboard offers latest information on reported COVID-19 cases globally.	https://data.who.int/dashboards
JHU-CSSE Repository	It is a publicly accessible repository of COVID-19 cases managed by the CSSE at JHU, USA.	https://github.com/CSSEGISandData/COVID-19
Worldometer	It provides latest information on COVID-19 cases globally.	https://www.worldometers.info/coronavirus/
OurWorld-In-Data	It offers datasets related to COVID-19 in various file formats.	https://ourworldindata.org/coronavirus, https://github.com/owid/covid-19-data
Harvard-database	It is a health service provided by Harvard University that offers COVID-19 related data.	https://huhs.harvard.edu/covid-19-information#gsc.tab=0
DataWorld	It offers COVID-19-related datasets contributed by organizations and users worldwide.	https://data.world/datasets/covid-19
CDC	It is a data-driven service organization that provides public health related data	https://www.cdc.gov/index.htm
ECDC	It is a European Union agency that manages datasets on COVID-19 and other diseases.	https://www.ecdc.europa.eu/en
HealthGoogle	It is a public repository by Google that includes raw data and visualizations related to COVID-19.	https://health.google.com/covid-19/open-data
Statista	This website offers free access to global COVID-19 facts and figures.	https://www.statista.com/page/covid-19-coronavirus
Flevy-Dashboard	It provides COVID-19 trends and news up to the territorial level.	https://flevy.com/coronavirus
Oxford-database	It offers information on COVID-19 policy measures.	https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker

Continued on next page

Table 1.2 – continued from previous page

Data Label	Description	Online Link
COVID-Tracking-Project	It gathers, verifies, and publishes COVID-19 related data.	https://covidtracking.com

1.7 Performance Evaluation Metrics

The effectiveness and accuracy of a model in solving a problem are assessed using performance metrics. These metrics help assess how well a model performs on a given dataset. They are generally used for comparing different models and selecting the best one for deployment. Performance metrics provide insights into the weaknesses and strengths of a model in machine learning, predictive modeling, and statistical analysis. The following metrics are used to assess the prediction accuracy of the proposed and compared models in this thesis.

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|z_i - \hat{z_i}| \tag{1.7.1}$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{z_i})^2} \tag{1.7.2}$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{z_i - \hat{z_i}}{z_i}\right| \tag{1.7.3}$$

Symmetric Mean Absolute Percentage Error (sMAPE):

$$sMAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{|z_i - \hat{z_i}|}{\frac{z_i + \hat{z_i}}{2}} \tag{1.7.4}$$

Root Relative Squared Error (RRSE):

$$RRSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{z_i} - z_i)^2}{\sum_{i=1}^{N}(\bar{z} - z_i)^2}}, \quad where\ \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i \tag{1.7.5}$$

Mean Absolute Scaled Error (MASE):

$$MASE = \frac{MAE}{MAE_{naive}}, \quad where\ MAE_{naive} = \frac{1}{N-1}\sum_{i=2}^{N}|\hat{z_i} - \hat{z_{i-1}}| \tag{1.7.6}$$

where $\hat{z_i}$ denotes forecasted value and z_i denotes actual value for the i^{th} sample, and N is the sample size.

Providing a controlled environment where modeling results can be observed and analyzed is an essential aspect of any scientific study. The experimental setup ensures that the computing resource is same for all the compared studies, results obtained are reliable, reproducible, and relevant to the research objectives. In this view, the experimental results are obtained using a system equipped with an Intel Core i5 processor running at 2.40 GHz, 4 GB NVIDIA GTX-1650 GPU, and 8 GB of RAM. The proposed and compared models are implemented in Python 3.8 or 3.9 in this thesis. Jupyter notebook editor is used to implement the models and visualize the results.

1.8 Research gaps and research objectives

From an analysis of the literature, the following research gaps which need to be addressed for time series forecasting of the COVID-19.

1. Comparative evaluation of state-of-the-art time series forecasting models tailored for COVID-19 prediction is limited. Empirical Analysis of the models on geospatial separated datasets is necessary to identify an effective and robust model.

2. COVID-19 time series data is highly dynamic in nature due to rapidly changing transmission rates and government policy measures such as lockdowns and vaccination campaigns. There is a need of development of a forecasting model which can handle the dynamism in non-stationary time series data effectively.

3. There is a need to design a model which can incorporate COVID-19 mutant affected population dynamics and produce better prediction results.

4. There is a need to address impact analysis of lockdowns, restrictions policies, and vaccinations on spread of the COVID-19.

The identified research gaps are framed into four research objectives. The framed research objectives are listed below.

1. Comparative analysis of state-of-the-art time series forecasting models for COVID-19 prediction.

2. Development of novel time series forecasting model for COVID-19 prediction.

3. Design of time series forecasting model for predicting COVID-19 for mutant affected population.

4. Design of novel epidemiological model incorporating lockdowns, mobility restrictions, and vaccination for COVID-19 prediction.

1.9 Motivation behind the problems addressed in this thesis

Predictive analysis of spread of an epidemic is become a prominent research area to support health services and governments in planning and mitigating the spread of an infectious disease. COVID-19 is one of a pandemics which has shown drastic dynamics impacting humanity badly. Early studies on COVID-19 were conducted with limited data available during the initial stages of the outbreak. As the virus spreaded globally, more comprehensive information became accessible for analysis. Extensive research was required to understand spread behavior of the SARS-CoV-2 virus. Artificial intelligence can support in shaping policies and controlling the outbreak by identifying spread patterns and providing predictive insights. The datasets of COVID-19 cases from states, countries, and continents can be utilized in the forecasting studies. The studies can incorporate factors such as healthcare infrastructure, government policies, vaccination programs, population demographics, topography, and other conditions into the models for better prediction results.

Prediction techniques are able to provide viable solutions for fighting against a pandemic in several ways such as.

- Design and development of effective and efficient prediction techniques to forecast the outbreak of a pandemic.
- Integration of optimizations and machine learning techniques for the predictions that can help governments in fighting against pandemics such as the COVID-19.
- Design and development of novel epidemiological compartmental models incorporating novel aspects of a pandemic.
- Optimal resource planning based on the prediction results.
- Impact analysis of vaccination programs to know the effectiveness of a vaccine.

1.10 Issues addressed in this thesis and their solutions

- **Problem 1**: Empirical Analysis of the state-of-the-art time series forecasting models on geospatial separated datasets is necessary to identify an effective and robust model.

  Chapter 3 presents empirical analysis of well-known statistical and machine learning models using time series of COVID-19 cases. The study has incorporated different datasets to evaluate the models. Variety of metrics are used to check the performance of the models. Firstly, ARIMA and FB-Prophet models are compared based on forecasting accuracy for infected,

active, recovered and fatality cases of COVID-19 from 10-highly affected countries. ARIMA model outperformed the FB-Prophet for all the datasets. In next experiment, the ARIMA is compared with the LSTM model for the prediction of COVID-19 cases of the USA, where the LSTM outperformed the ARIMA models. Lastly, ETS, ARIMA, LSTM, and ANN models are compared using COVID-19 time series datasets from five countries namely, the USA, India, Italy, Russia, and the UK. The forecasting accuracy is evaluated using MAE, RMSE, RRSE, sMAPE, and MASE performance measures. The Results have shown that the ETS and ARIMA models are not able to handle the dynamic scenarios whereas, the ANN and LSTM models effectively captured temporal dependencies in dynamic situations. The LSTM outperformed the other compared models.

- **Problem 2**: Design and development of a novel forecasting model is needed which can handle dynamic time series data effectively.

  Chapter 4 focuses on developing a COVID-19 forecasting model based on Fuzzy Time Series (FTS) with optimized hyperparameters. The FTS offers semantic interpretation unlike traditional time series models, making it better suited for handling uncertain and fluctuating data. Particle Swarm Optimization (PSO) has attracted significant attention among various optimization methods discussed in the literature in recent years. Consequently, PSO is employed in this chapter to determine optimal solutions for COVID-19 forecasting. In this chapter, two integrated algorithms combining FTS and PSO are proposed namely, Nested-FTS-PSO and Exhaustive-Search-FTS-PSO. These algorithms aim to optimize the number of partitions, the length of partition intervals within the Universe of Discourse (UOD), and the fuzzy order. The predictive analysis covers two distinct phases of the COVID-19 pandemic. The first phase (2020), characterized as limited understanding and the absence of effective containment measures. The second phase (2021), characterized by increased knowledge and the introduction of vaccines in many regions. The proposed models are benchmarked against several state-of-the-art forecasting approaches, including the classic FTS model, FTS-PSO, FB-Prophet, and ARIMA, using daily confirmed COVID-19 case data from both phases. The proposed exhaustive-search-FTS-PSO algorithm demonstrated superior performance in forecasting accuracy among the evaluated methods.

- **Problem 3**: Development of a novel model which can incorporate COVID-19 mutant affected population dynamics to produce better prediction results.

  Chapter 5 explores a range of hybrid forecasting models for predicting COVID-19 cases by integrating the strengths of Fuzzy Time Series (FTS), contextual deep learning, and PSO for hyperparameter tuning. Classical time series models often perform poorly on non-stationary data. A prime example of a non-stationary characteristic is the COVID-19 time series. In case of COVID-19, the forecasting task becomes even more complex due

to the influence of diverse factors such as lockdown measures, population density, SARS-CoV-2 variants, vaccination campaigns, and governmental policies, making it difficult for a single model to deliver accurate predictions. To address this challenge, a comprehensive forecasting framework is proposed, comprising four main components: 1) Data preprocessing, 2) FTS modeling with PSO-based optimization, 3) Deep learning-based forecasting, and 4) Defuzzification and performance evaluation. The data preprocessing stage applies three techniques: i) Differencing, ii) Weekly averaging, and iii) Outlier removal in both training and testing datasets. The number of intervals is determined using an average-based method, while PSO is employed to optimize the fuzzy order and define unequal interval lengths. Fuzzy Logical Relationships (FLRs) are constructed based on the optimal fuzzy order. A deep learning model is then used to learn patterns from the fuzzified data and FLRs, enhancing ability of the model to capture complex dynamics. Specifically, the model incorporates an attention-based Bidirectional LSTM (Bi-LSTM) trained on fuzzified inputs. The proposed hybrid model is evaluated using COVID-19 confirmed case data from the USA, India, the UK, Russia, and Italy. Benchmarking against state-of-the-art forecasting methods shows that the FTS-PSO-attention-Bi-LSTM model achieves superior performance across multiple error metrics, including MAE, RMSE, RRSE, sMAPE, and MASE.

Further, an epidemiological compartmental model is employed to analyze the progression of COVID-19, integrating the SIRD model, PSO, and a stacked-LSTM network. The model dynamically updates its parameters on a weekly basis to account for changes brought about by new infection waves or shifts in government policies. PSO is used to optimize the parameters of the SIRD model. The stacked-LSTM is trained on the optimized SIRD parameters derived from the training data. It then forecasts future parameters for the testing phase. These predicted parameters are subsequently input into the SIRD model to generate time series forecasts. The proposed hybrid framework is evaluated using COVID-19 case data from three countries namely, the USA, the UK, and India. Experimental results demonstrate that hybrid of SIRD, PSO, and stacked-LSTM combined consistently outperforms all baseline models across all datasets, delivering superior forecasting accuracy.

- **Problem 4**: Design and development of an epidemiological compartment model which can address impact analysis of lockdowns, restrictions policies, and vaccinations for COVID-19.

  Chapter 6 presents a novel epidemiological compartmental model that accounts for multi-dose vaccinations, protected individuals, and immunized susceptibles. The proposed model, referred to as the Susceptible-Exposed-Infected-Recovered-Deceased-Protected-Vaccinated (SEIRDPV) model, comprising ten compartments: Susceptible, Exposed, Infected, Recovered, Deceased, Not-fully-vaccinated, Fully-vaccinated, Booster, Protected, and Immunized susceptible. The model introduces a restriction parameter to

capture the influence of government interventions, which reflects the effects of public health measures. This parameter is estimated based on real-world factors such as the emergence of new variants and the implementation of restrictions or advisories during specific time periods. The proposed model parameters are optimized using the PSO algorithm to improve forecasting accuracy. The SEIRDPV model is applied to two case studies involving India and the USA, and its performance is benchmarked against five state-of-the-art models: 1) ETS, 2) ARIMA, 3) ANN, 4) LSTM, and 5) SEIRD. Forecast accuracy is evaluated using standard error metrics namely, MAE, RMSE, RRSE, sMAPE, and MASE. The results indicate that machine learning models tend to perform better in capturing the complex dynamics of pandemic data characterized by multiple waves. Traditional epidemiological models like SEIRD, in contrast, rely on fixed hyperparameters and struggle to adapt to evolving scenarios. However, the proposed SEIRDPV model demonstrates that when hyperparameters are periodically re-estimated based on real-time events, epidemiological models can effectively reflect the dynamic nature of disease spread.

Further, an experimental study is conducted to estimate a resource-optimized restriction parameter. In this work, the traditional SEIRD model is extended by introducing an additional compartment representing hospital admission capacity, resulting in the SEIHRD model. This enhanced model consists of six compartments: Susceptible (S), Exposed (E), Infected (I), Hospitalized (H), Recovered (R), and Deceased (D). The study focuses on optimizing the use of hospital resources by estimating a restriction parameter that ensures the number of infections remains within the limits of available hospital beds. Time-varying model parameters are incorporated to reflect changing pandemic conditions and optimized using the PSO algorithm. Additionally, a multi-variable LSTM-based deep learning model is used to learn and forecast the time-dependent parameters of the SEIHRD model. The framework is evaluated using COVID-19 case data from the USA and India. Results from the study reveal patterns of both over-utilization and under-utilization of hospital resources, demonstrating potential of the proposed model for informing data-driven healthcare policy and capacity planning.

1.11 Organization of the thesis

This thesis is organized into the following chapters:

- Chapter 2: Provides a comprehensive literature review.
- Chapter 3: Focuses on the forecasting framework and evaluation of time series forecasting models.
- Chapter 4: Proposed hybrid FTS forecasting model with optimizations for handling the highly dynamic time series data of COVID-19.

- Chapter 5: Introduces the integration of deep learning and optimization with FTS to enhance forecasting accuracy, presenting the proposed methodology. It also describes a hybrid epidemiological compartmental model integrating with deep learning techniques.
- Chapter 6: Presents multi-factor epidemiological compartmental model, incorporating restriction policies, multi-dose vaccinations, vaccine efficacy, and time-varying parameters.
- Chapter 7: Summarizes the research findings, social impacts, and outlines future research directions.

CHAPTER 2

LITERATURE REVIEW

The COVID-19 forecasting problem has emerged as one of the most complex due to its highly dynamic spread pattern. FTS modeling offers another approach to address these dynamics. Additionally, epidemiological models are well-suited for capturing the evolution of such pandemics. A wide range of time series forecasting problems have been successfully addressed in the literature using artificial intelligence (AI). As a result, AI holds significant potential for predicting the dynamics of the COVID-19 pandemic. In this chapter, COVID-19 prediction related studied are reviewed. Discussion on the related studies is given in the following sections.

2.1 Time series forecasting models for COVID-19

Initially, some of the studies have focused on impact analysis of the governmental policies which were imposed to control the spread of the COVID-19. Ashcroft et al. [16] examined the impact of quarantine duration on the COVID-19 outbreak. Their study empirically determined the distributions of the incubation period, and infectivity. They have evaluated impact of quarantine length on the onward transmission of SARS-CoV-2 from traced contacts of confirmed cases and returning travelers from other countries. Researchers have assessed various forecasting models for predicting COVID-19 trends. In Study [182], the authors evaluated several models using daily COVID-19 case data from the 10 most affected states in Brazil. They found that the stacking ensemble and Support Vector Regression (??) models outperformed Autoregressive Integrated Moving Average (??), Combination-of-Uniform-and-Bias-adjusted-Incremental-Splitting-Trees (CUBIST), Regression-with-Interaction-terms-and-Dummy-variables-for-Group-Effects (RIDGE), and Random Forest (??) models. In study [61], the author developed an ARIMA(p,d,q) model to analyze the epidemiological trends of COVID-19 in the three most affected European countries namely, Spain, Italy, and France. The time series data was used from February 21 to April 15, 2020. The study compared the ARIMA model using different orders (p, d, q), and identified the optimal orders based on the lowest Mean-Absolute-Percentage-Error (??) value. The study concluded that ARIMA model is effective for forecasting COVID-19 trends. Chintalapudi et al. [46] used a seasonal ARIMA model to forecast COVID-19 cases in Italy, analyzing data up to March 31, 2020. They studied the impact of a two-month lockdown, observing a decline in confirmed cases and an increase in recovered cases. Alabi et al. [4] employed the Facebook Prophet model to predict COVID-19 spread, including confirmed and death cases. Arun Kumar et al. [14] examined the forecasting dynamics of cumulative confirmed, recovered, and death cases using ARIMA and seasonal ARIMA models, performing a 60-day forecast for the top 16 affected countries and identifying those

with exponential case growth. Alasafi et al. [6] used Recurrent Neural Network (??) and Long Short-Term Memory (??) models to forecast confirmed and death cases in Morocco, Malaysia, and Saudi Arabia (up to December 3, 2020). These models achieved a precision accuracy of 98.58% and 93.45%, respectively. Lyu et al. [134] analyzed infection doubling times and effective lockdown dates in Victoria, Australia, using COVID-19 data, restriction policies, and mobility data. They proposed a simulation model based on Gradient-Boosted-Regression (GBR), RF, Extreme-Gradient-Boosting (XGBoost) Regression, Decision Tree, SVR, Kernel Ridge, Elastic Net, Linear Regression (LR), Ridge Regression (RR), Bayesian Ridge, and LR models, with the GBR model achieving the best performance. Policies of Indonesian government against COVID-19 were analyzed in [183] using a systematic review approach. The systematic review has served as a vital tool for summarizing the evidences.

A range of forecasting models for Fuzzy Time Series (FTS) have been proposed in the literature, each incorporating different features. A review of these FTS forecasting techniques is presented in [24]. The authors categorized existing studies by classifying the stages of FTS forecasting into five main steps: (1) Defining the universe of discourse (UOD), (2) Partitioning the UOD, (3) Fuzzification, (4) Establishing fuzzy logical relations, and (5) Defuzzification (if necessary). These stages are grouped into two main phases: (a) Data partitioning and (b) Prediction phase. The review discusses data partitioning methods, organizing them in a hierarchical tree structure with clustering algorithms and optimization techniques at the leaf level. Panigrahi and Behera [167] addressed two key challenges in higher-order FTS forecasting: determining the optimal interval length within the UOD and modeling fuzzy logical relationships (??). To determine the optimal interval length, a modified average-based method is employed, while machine learning techniques, specifically LSTM, deep belief network (DBN), and support vector machine (??) are used for modeling FLRs. Tinh [222] proposed a combined FTS and Particle Swarm Optimization (??) model to forecast confirmed COVID-19 cases in Vietnam. The study demonstrated that the FTS-PSO model achieved the best performance with 16 partitions and 5th-order FTS when applied to a one-month COVID-19 dataset from March 4, 2020, to April 7, 2020. Several researchers have applied the FTS optimization concept for COVID-19 forecasting [33]. Melin et al. [146] introduced an ensemble model using neural networks and a type-3 fuzzy system for forecasting COVID-19 confirmed cases, enhancing prediction accuracy with firefly optimization. Xion et al. [231] proposed an improved seagull optimization algorithm (ISOA) combined with XGBoost for FTS forecasting of COVID-19 confirmed cases. They used ISOA to optimize the domain partition and XGBoost for forecasting instead of relying on fuzzy relations. High-order FTS models have been shown to offer better forecasting accuracy compared to first-order FTS models. Consequently, most recent studies have focused on high-order FTS models, as noted in review studies [24, 48]. Furthermore, integrating fuzzy techniques with deep learning has become an emerging area of research to further improve prediction accuracy.

2.2 Epidemiological compartmental modeling for COVID-19

Numerous studies employing epidemiological compartmental models have been conducted to analyze the spread of the COVID-19. These studies have been instrumental in shedding light on how diseases like COVID-19 spread. These studies typically divide populations into compartments and analyze the dynamics using available data. Researchers have used these models to predict the outcomes of infectious diseases and trace their evolutionary trajectories. Both standard epidemic models and their extensions with additional compartments have been extensively evaluated. The SIR model, introduced by Kermack and McKendrick [104], is a foundational epidemic model that has been widely applied to study the COVID-19 pandemic [174, 98, 106]. Similarly, the standard Susceptible-Exposed-Infectious-Removed (SEIR) model has been analyzed using COVID-19 data in numerous studies [99, 105, 136]. To address the limitations of baseline models, researchers have proposed extensions. Sarkar et al. [192] introduced a six-compartment model including susceptible, quarantined susceptible, asymptomatic, infected, isolated infected, and recovered individuals to better capture the dynamics of COVID-19. Another model incorporating governmental interventions and quarantine compartments was proposed by Mandal et al. [141] to study mitigation strategies for COVID-19 transmission. The outcomes and challenges of various epidemic models for COVID-19 predictions have also been explored in [9]. However, most of these models assume time-invariant parameters, limiting their ability to adapt to dynamic changes such as the emergence of new variants, multiple waves, government policies, and vaccination programs. As COVID-19 exhibited multiple waves in its spread, researchers began investigating time-variant parameters to model these fluctuations [237]. Studies such as [66, 13, 240] explored models with time-varying parameters, accounting for factors like restriction policies, SARS-CoV-2 variants, lockdown relaxations, demographics, and vaccination efforts. To enhance prediction accuracy, researchers also experimented with ensemble and hybrid models, combining compartmental frameworks, evolutionary algorithms, fuzzy logic, and machine learning approaches [160, 244, 178, 115]. Following the development of vaccines, governments initiated phased vaccination programs, prompting researchers to analyze the impact of programs and evaluate vaccine efficacy.

Numerous studies examined the effects of vaccination campaigns on controlling the COVID-19 disease [30, 35, 69, 180, 91]. Olivares et al. [162] employed a compartmental epidemic model to study the effects of vaccination, social distancing, and testing on susceptible individuals. Their research quantified the uncertainties related to these mitigation measures and their influence on disease transmission. Similarly, [203] proposed a hybrid model that integrates a nonlinear autoregressive method with radial basis function (NAR-RBF) neural networks, applied within the Susceptible, Infectious, Treated, Recovered (SITR) framework to examine the bi-modal transmission pattern of COVID-19. To improve convergence speed and prediction accuracy, they introduced a novel class of transformations for systems

of ordinary differential equations. These studies provide valuable insights for planning, monitoring, and mitigating the spread of COVID-19.

Furthermore, numerous research efforts have leveraged compartmental and epidemic models to analyze the dynamics of the COVID-19 outbreak [142, 100]. Kong et al. [112] reviewed compartmental structures used in COVID-19 modeling, analyzing three types of models: meta-population models, SEIR-based expanded compartment models, and agent-based models. They observed that SEIR models were adopted to include COVID-19 dynamics, non-pharmaceutical interventions, and age structures but emphasized the need for further research to address mutant strains and vaccination impacts. Saldaña and Hernández [189] provided an overview of mathematical epidemiological models, revisiting deterministic compartmental frameworks inspired by the work of Kermack and McKendrick, along with concepts such as effective reproduction numbers, herd immunity, and vaccination policies. Cangiotti et al. [31] conducted a survey of key Lyapunov functions for epidemic compartmental models, offering readers tools to apply these functions to models of their choice. Xiang et al. [233] reviewed 55 studies on epidemic prediction models and the impact of contact tracing and social isolation strategies on COVID-19. Their findings highlighted the importance of protective measures such as face masks, enhanced quarantine, improved reporting rates, and travel restrictions as the most effective control strategy. They cautioned against over-reliance on mathematical models for formulating control strategies, stressing the need for careful interpretation of prediction results. Lee et al. [123] reviewed 42 articles discussing methods, structures, and the roles of mathematical models in understanding COVID-19 dynamics in Korea. They emphasized the utility of these models in designing control strategies, evaluating epidemic progression, and assessing the effectiveness of containment measures. Chen et al. [44] employed the SEIHRD model to evaluate the risk of a COVID-19 outbreak in Japan. They divided the period from January 6, 2020, to March 31, 2020, into four phases and estimated key parameters based on imported cases. Borri et al. [23] introduced a time-varying Susceptible-Infected-Recovered-Dead (SIRD) model for real-time optimization of epidemic restrictions. Their framework determines a sequence of optimized infection rates to balance health and economic costs. Simulation results using COVID-19 data from Italy suggest that ideal implementation of this approach could reduce deaths by 76.71% compared to actual reported figures.

Zhou et al. [247] analyzed the relationship between mobility patterns, intervention policies, and SARS-CoV-2 transmission in the USA using VAR models and the Toda-Yamamoto Granger causality test. Their study examined spatial and temporal variations across different states and phases, utilizing mobility and COVID-19 policy datasets. Results indicate that face mask mandates and emergency declarations were the most effective NPIs in controlling the spread in the top ten states. Miikkulainen et al. [147] proposed an evolutionary AI-driven model to automatically determine the most effective NPIs. They trained an LSTM-based model on infection data to generate optimal NPI recommendations for mitigating the spread of COVID-19. Martins et al. [143] have proposed N-step-ahead optimal control strategies against an epidemic using a SEIHRD epidemiological

model. The authors have experimented five type of strategies based on isolation of cases, tracking of new cases, interrupt the transmission by restrictions and lockdown, social distancing, vaccination, and omniscient control. The authors have claimed that the control strategies worked well in case of constant model parameters to maintain the level of hospitalizations. Niu et al. [161] have proposed a stochastic SEIHR model to deal with the fluctuations in the number of infections and hospitalizations. The authors have checked disease-free equilibrium using stochastic stability of the model based on the basic reproduction number (R0). The disease-free equilibrium was stochastically stable for less than one value of the R0. Gómez-Corral et al. [73] have proposed a markovian epidemic model to study infectious disease outbreaks under limited medical facilities, specifically accommodation capacity for infected individuals in the hospitals. Ma et al. [135] have considered three types of patients, emergency, COVID-19, and elective for dynamic bed allocation in a health care system during the COVID-19 outbreak. The authors have considered two control mechanisms to maximize the utilization of the public hospital beds. One is the dynamic allocation of beds and second is the moving of elective patients to private hospitals based on a subsidized scheme.

2.3 AI techniques and optimizations for COVID-19 forecasting

Advancements in artificial intelligence have been pivotal in the fight against the COVID-19 pandemic. Nabi et al. [155] conducted forecasting for confirmed and death cases using LSTM, Gated-Recurrent-Unit (??), Convolutional-Neural-Network (??), and Multivariate-CNN models for Brazil, Russia, and the UK, with the CNN model outperforming others. Zeroual et al. [241] conducted a comparative study of five deep learning models: simple RNN, LSTM, Bi-LSTM, GRUs, and Variational AutoEncoder (VAE), to forecast the number of recovered and confirmed COVID-19 cases. The study focused on six countries affected by the virus namely, Australia, China, France, Italy, Spain, and the USA, using historical data spanning 148 days from January 22, 2020, to predict cases 17 days ahead. The results showed that the VAE model outperformed all other models. In another study [197], deep learning models including Stacked LSTM, Bi-LSTM, and Conv-LSTM were employed to conduct a comparative analysis of COVID-19 spread in India and the USA. The authors conducted a one-month ahead forecast and found that the Conv-LSTM model delivered superior performance compared to the other two models.

Several machine learning models have been proposed for COVID-19 time series forecasting, with Long Short-Term Memory (LSTM) [87] emerging as a popular deep learning approach. LSTM and its variants have been explored in various studies to predict COVID-19 trends. Chimmula et al. [45] utilized an LSTM-based deep learning approach to forecast COVID-19 cases in Canada, providing predictions for successive days (e.g., 2nd, 4th, 6th, 8th, 10th, 12th, and 14th days). Their findings indicated that provinces enforcing social distancing guidelines reported fewer infections compared to others. Similarly, Arora et al.

[11] proposed LSTM-based models to forecast COVID-19 cases across Indian states for one-week and one-day intervals. They observed that advanced variants such as convolutional LSTM (Conv-LSTM), stacked LSTM, and bidirectional LSTM (bi-LSTM) demonstrated greater accuracy compared to the standard LSTM model. Xu et al. [234] predicted COVID-19 case numbers in Brazil, India, and Russia using CNN, LSTM, and a hybrid CNN-LSTM model. Numerous studies have explored prediction techniques, with some proposing novel methods to train algorithms on COVID-19 data [246, 148]. Paul et al. [169] reviewed AI-based approaches related to the pandemic, categorizing them into machine learning (ML), deep learning (DL), and hybrid ML-DL techniques. Kalinowska and Orhan [184] reviewed ML and AI-based methods, addressing their adaptability to healthcare systems, strategies for controlling virus spread, vaccine development, and the diagnosis and treatment of infected individuals. In another study, Arun et al. [15] compared forecasting and statistical models using COVID-19 time series data. They employed ARIMA and seasonal-ARIMA (SARIMA) as statistical methods and LSTM and GRU as deep learning techniques, demonstrating that deep learning methods outperformed statistical approaches.

Evolutionary optimization techniques are generally based on nature inspired processes [113, 216]. A number of evolutionary algorithms have been published in the literature. The well-known algorithms are particle swarm optimization (PSO) [103], genetic algorithm (GA) [124], grey wolf optimizer (GWO) [152], dragonfly algorithm [150], firefly algorithm (FA) [236], Bayesian learning [17], simulated annealing (SA) [125], ant-colony optimization (ACO) [29], chaos-enhanced cuckoo search algorithm [89], ant lion optimizer [149], and whale optimization [151]. These algorithms are very effective in handling ambiguous and uncertain scenarios. They can learn and adapt to the dynamic conditions. They can provide optimal or nearly optimal solutions to complex problems with minimal computational effort. Tseng et al. [219] conducted a survey on computational intelligence techniques for combating COVID-19. Their study categorizes existing work in computational intelligence related to the pandemic, focusing on approaches such as evolutionary computation, machine learning, soft computing, and big data analytics in the context of COVID-19.

2.4 Hybrid models for COVID-19 forecasting

Hybrid model using two or more than two techniques is more powerful and efficient than the individual technique. Hybrid models leverage the strengths of more than one techniques in one model. Thus, it improves accuracy, robustness, and performance of the model [235, 127]. From this perspective, researchers have proposed various hybrid models for epidemiological compartmental modeling of the COVID-19 outbreak [116]. Abbasimehr and Paki [1] proposed a hybrid approach combining three deep learning models, namely, LSTM, CNN, and Bayesian Optimization (BO). Among the three models, the LSTM-BO model achieved the lowest SMAPE in six out of ten countries. In [196], a novel hybrid diagnostic strategy integrating fuzzy logic and a deep neural network was proposed to detect COVID-19 patients. The strategy, validated using 10-fold cross-validation,

achieved a detection accuracy of 97.66%. Chang et al. [38] developed a model that integrates feature selection (using LR, DT, RF), instance clustering (using k-means and Expectation-Maximization), and DNN techniques to predict mortality risk in COVID-19 patients. The proposed model outperformed standard neural networks in prediction accuracy. Similarly, Anil Utku [221] employed a hybrid CNN-GRU model to forecast COVID-19 cases in the ten most populous countries and compared its performance against LR, RF, SVM, MLP, CNN, GRU, LSTM, and the base CNN-GRU model. Zhang et al. [243] proposed an interpretable hybrid model combining AR and LSTM, and evaluated it against AR, LSTM, SVM, XGBoost, and RF. The model was applied to forecast COVID-19 cases in Japan, Canada, Brazil, the UK, Italy, Argentina, and Singapore. By integrating AR with predictive power of LSTM, the hybrid model achieved high forecasting accuracy.

Zelenkov and Reshettsov [240] proposed a hybrid model using time-varying SEIRV optimized with genetic algorithm. They analysed different control strategies of the COVID-19 pandemic in Sweden, Germany, UK, and the USA. The model achieved the MAPE values of 2.096%, 2.168%, 1.208% and 1.703% for Sweden, Germany, the UK, and the USA, respectively. Combination of edge RNN and SIRD models has been investigated in [111] for predictive analysis of the COVID-19 infected cases in Italy. The hybrid model outperformed the other compared spatio-temporal models in 3-day ahead forecasting. In this direction, shringi et al. [204] also proposed an hybrid model integrating grey wolf optimization with the SIRD model. Forecasting results using the model are closely aligned with the actual cases of the COVID-19 in India. Janko et al. [95] developed a hybrid epidemiological model combining SEIRD and machine learning techniques to forecast COVID-19 cases. The model consists of three components: socio-economic cost analysis of 12 types of NPIs, case prediction using the SEIRD model, and an NPI optimization prescriptor. SEIRD parameters were dynamically adjusted based on NPIs using various ML models, including linear regression, decision trees, and random forests.

Some of the researchers have investigated ensemble models for the COVID-19 predictions. Qu et al. [175] developed a novel ensemble model combining multiple neural networks with a whale optimization algorithm for modeling the COVID-19 outbreak. The authors highlight the limitations of relying solely on a single forecasting model, arguing that it overlooks the strengths of other approaches. To address this, they integrated four neural networks for predictions, assigning weights to each model and optimizing them using a heuristic optimization algorithm. Saadatmand et al. [186] also investigated ensemble of XGBoost, KNN, RF, bagged-CART, and LogitBoost for the prediction of ICU admission, hospital stay length, and mortality of the COVID-19 patients. The model outperformed with an accuracy over 95% in comparison with the baseline models. Another ensemble model using ANN, CNN and LSTM, with whale optimization is proposed in [212]. The model is used for the prediction of COVID-19 patients. It achieved prediction accuracy more than 90%.

CHAPTER 3

COMPARATIVE ANALYSIS OF STATE-OF-THE-ART TIME SERIES FORECASTING MODELS FOR COVID-19 PREDICTION

[1] Time Series Forecasting (TSF) models are employed to predict future outcomes using historical data. Numerous forecasting models have been proposed in the literature, demonstrating their effectiveness across various datasets and scenarios. The emergence of COVID-19 introduced a complex problem in time series forecasting, highlighting the necessity to evaluate state-of-the-art models on COVID-19 datasets and fine-tune them to adapt to the unique characteristics of the pandemic. In the initial phase of the COVID-19 pandemic, vaccines were not available against the SARS-CoV-2 virus. The primary motivation of this chapter is to model the spread pattern of the SARS-CoV-2 virus and predict its impact. It can help the governments to optimize the planning of public services and resource allocation in the pandemics similar to the COVID-19. By modeling and forecasting the daily spread of the virus, healthcare systems can better prepare to accommodate the anticipated number of patients.

Forecasting models typically follow a series of steps to generate prediction results. The forecasting process consists of three main stages: 1) data preprocessing, 2) modeling, and 3) performance evaluation. During data preprocessing, the dataset is cleaned by eliminating outliers, addressing missing values, and smoothing irregularities. It is then divided into training and testing subsets for model development and evaluation. In the modeling phase, the training data is used to build the model, which is subsequently applied to the testing subset to generate predictions. Finally, various performance metrics can be used to evaluate the forecasting capability and robustness of the model. Accurate forecasting of disease spread is essential as it directly impacts government policies, containment strategies, healthcare systems, and societal interactions. A forecasting model can be selected from various available options, datasets can be sourced from a data repository, and either short-term or long-term forecasting can be conducted based on the requirements. Various measures can be used to evaluate accuracy of the model. A forecasting model generates potential outcomes accordingly. To compare the effectiveness of different models on COVID-19 datasets, this chapter performs a detailed analysis of popular time series forecasting models.

[1]The contents of this chapter are published in a research paper entitled "*COVID-19 pandemic prediction using time series forecasting models*" in international conference on computing, communication and networking technologies (ICCCNT), pp. 1-7. IEEE, 2020.

3.1 COVID-19 spread trends

This section visualizes the spread pattern of the COVID-19 outbreak in two scenarios: 1) the early stage, and 2) the evolved lifespan of the pandemic. For the early stage, time series data of COVID-19 cases was collected from the GitHub repository [94]. This repository is managed by the Center for Systems Science and Engineering (??) at Johns Hopkins University (??), USA. It contains cumulative day-to-day reported COVID-19 cases starting from January 22, 2020.

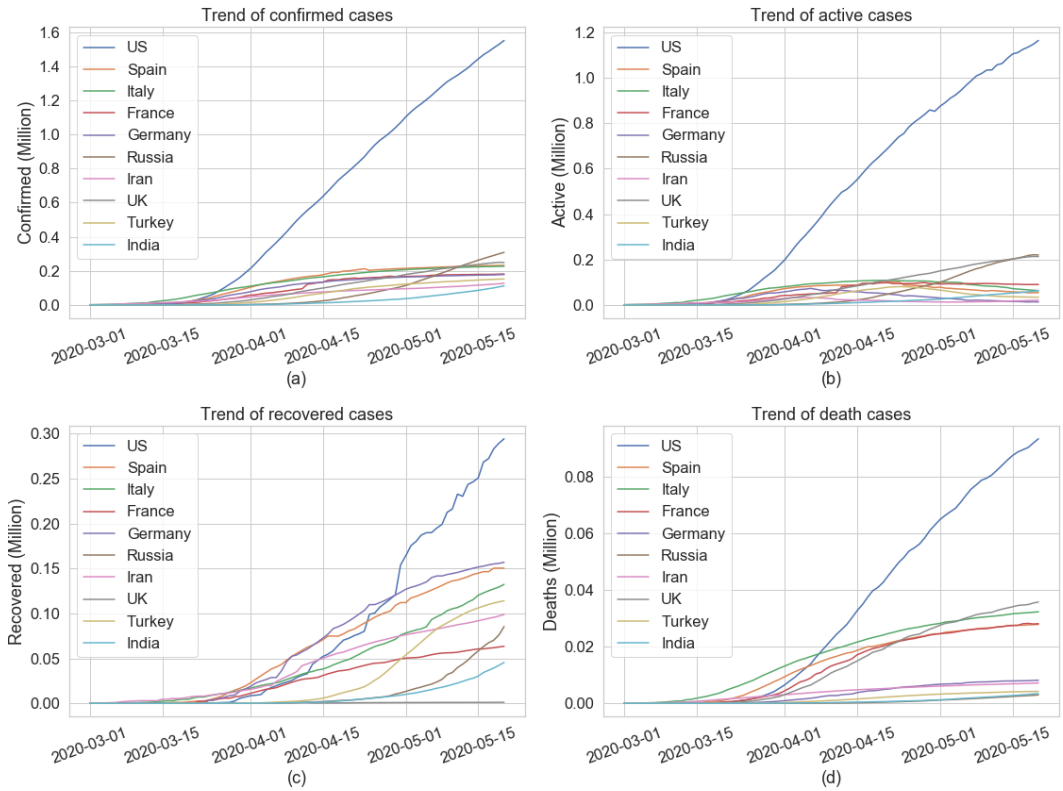


Figure 3.1: Pattern of COVID-19 confirmed, active, recovered, and fatality cases in adopted countries.

In the early stage of the pandemic, the impact of COVID-19 in 10 selected countries from March 1, 2020, to May 20, 2020, is illustrated in Fig. 3.1. These countries were among the most affected by COVID-19 as of May 20, 2020. The pattern of confirmed cases is shown in Fig. 3.1 (a), active cases in Fig. 3.1 (b), recovered cases in Fig. 3.1 (c), deaths cases in Fig. 3.1 (d). Countries ranking from highest to lowest affected, follows the labeling order in the figure. It can be observed that the USA was the most affected, recording the highest number of confirmed and death cases. The USA and Russia continued to show rising trends while the other countries managed to flatten their curves after a certain period.

Figure 3.2 shows the spread pattern of daily confirmed COVID-19 cases over the period, from January 1, 2020, to December 31, 2022, across ten affected countries. Weekly averaging is applied to smooth the time series data. The figure

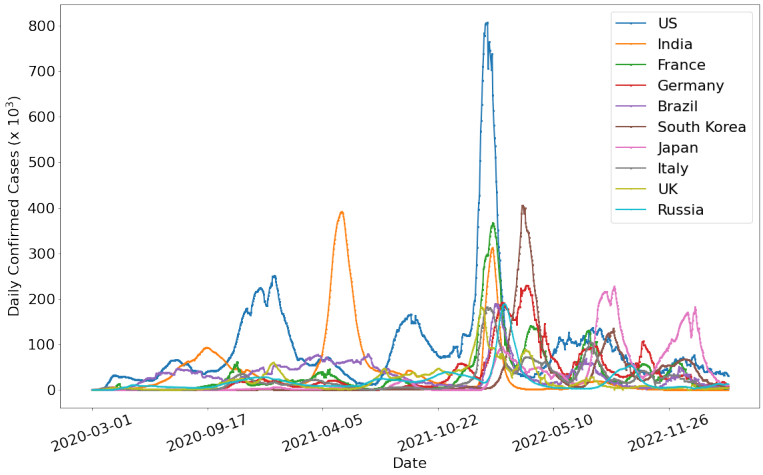
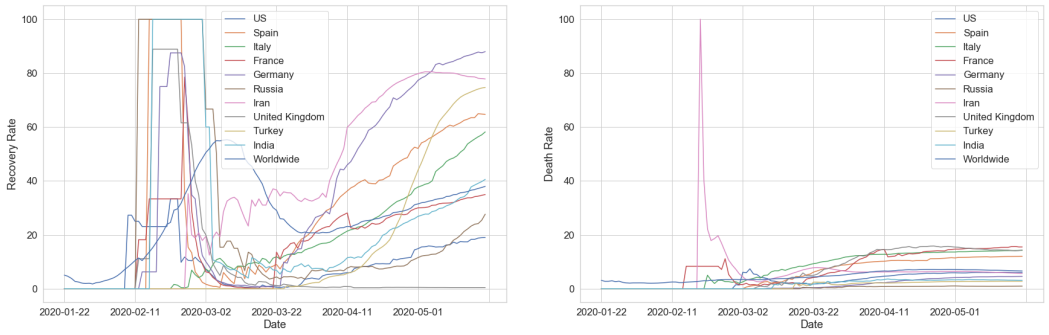


Figure 3.2: Spread pattern of the disease during COVID-19 discourse in highly affected 10 countries.

highlights that the USA was significantly impacted by the pandemic. During the COVID-19 pandemic, the SARS-CoV-2 virus underwent multiple mutations, leading to several waves of infections. The spread pattern was influenced by restrictions imposed by the governments, emergence of virus variants, vaccination campaigns, healthcare infrastructure, and people behavior.



(a) Recovery rate of COVID-19 cases. **(b)** Fatality rate of COVID-19 cases.

Figure 3.3: Forecasting of infected cases using experimented models and comparison with test data of the adopted country

Figure 3.3a illustrates the recovery rate, and Figure 3.3b depicts the fatality rate of COVID-19 patients worldwide and in the selected countries during the early stage of the pandemic. Figure 3.3a shows that Iran had the highest recovery rate, while Turkey shown exponential growth in recovered cases. Other countries exhibited recovery trends that mirrored the growth patterns of confirmed cases. Studies [207, 242, 80] have reported that COVID-19 often resolves naturally over time but can cause severe health complications that may lead to death. Figure 3.3b shows that Iran initially had the highest death rate, which was later overtaken by the USA and France. Spain also experienced a notably high fatality rate

during this time. Other countries managed to control fatalities to some extent by implementing restriction policies and enforcing social distancing measures.

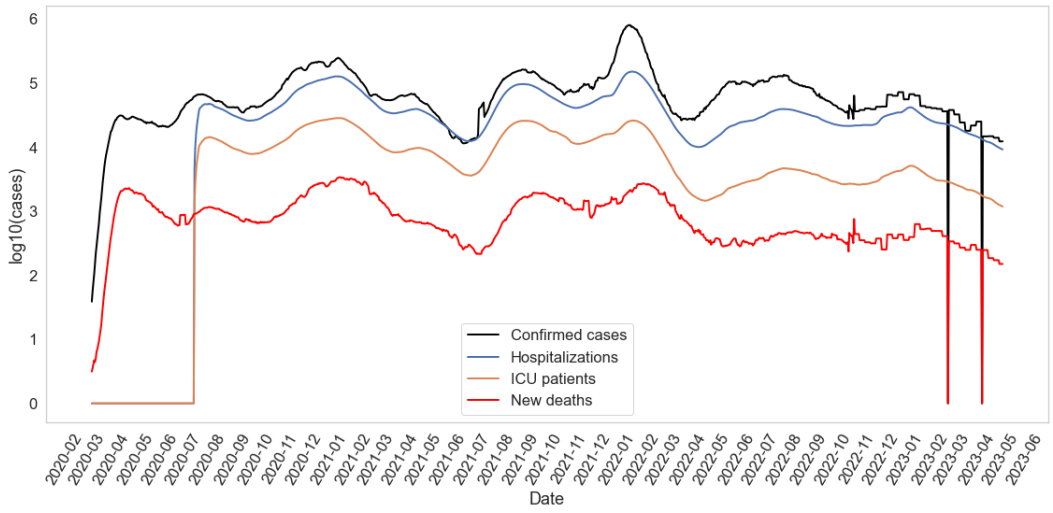


Figure 3.4: COVID-19 deaths, hospitalizations, and ICU patients against the infected cases in the USA.

The number of infections and fatalities over time provides insight into the severity of a disease. In this context, Figure 3.4 presents time series trends of daily death cases, hospitalizations, and ICU admissions relative to COVID-19 infections in the USA. The time series data of the figure 3.4 is sourced from the GitHub repository maintained by Our World in Data (OWID) [93]. The time series spans from March 1, 2020, to May 4, 2023, covering the effective lifespan of the COVID-19 pandemic. A transformation of $log10$ scale is applied on the data for improved visualization, as daily infection cases far exceed hospitalizations, ICU admissions, and death cases. A value of one is added to all the data points to address zero values in the log scale. From the figure, it can be seen that a rise in hospitalizations and ICU admissions are observed due to the lifting of lockdown restrictions. The USA successfully managed the surge in cases due to its robust medical infrastructure.

3.2 Forecasting using ARIMA and FB-Prophet

In this section, ARIMA [230] and Facebook Prophet (??) [217] models are evaluated using COVID-19 forecasting. The ARIMA(p, d, q) is a model formed by integrating the Autoregressive (AR) component and the Moving Average (MA) component. Here, p denotes the order of autoregression, d represents the order of differencing, and q indicates the order of the moving average. The ARIMA can effectively model time series data if the data is stationary, meaning its mean and standard deviation remain constant over time. The differencing parameter d specifies the number of transformations needed to achieve stationarity. The FB-Prophet combines linear and non-linear methods with time as a regressor. The model was developed and made available as open-source software by data science

team at Facebook. FB-Prophet treats training as a curve-fitting exercise. It ignores the temporal dependence of the data. The model supports irregularly spaced observations in the dataset. Key advantages include handling multiple seasonal periods, incorporating custom and predefined holidays, and flexible trend options such as a piecewise linear model or a saturating growth model. Additionally, FB-Prophet takes very less computation time in the data fitting process.

3.2.1 Modeling Dataset

To evaluate the ARIMA and the FB-Prophet models, aggregated day level COVID-19 time series data of 10 countries is adopted in the experiments. Start date of the time series and other detail of the samples are given in Table 3.1. The end date of the COVID-19 time series samples is May 20, 2020. In this experiment, the dataset is split into 80% for training and 20% for testing. Model performance is assessed using COVID-19 data, including confirmed, recovered, death, and active cases.

Table 3.1: COVID-19 cases aggregated samples from each adopted country and worldwide as on May 20, 2020.

Region	Sample size (Days)	Start Date	Confirmed	Recovered	Deaths	Active
Worldwide	120	2020-01-22	4996472	1897466	328115	2770891
US	115	2020-01-27	1551853	294312	93439	1164102
Spain	105	2020-02-06	232555	150376	27888	54291
Italy	106	2020-02-05	227364	132282	32330	62752
France	113	2020-01-29	181700	63472	28135	90093
Germany	110	2020-02-01	178473	156966	8144	13363
Russia	106	2020-02-05	308705	85392	2972	220341
Iran	87	2020-02-24	126949	98808	7183	20958
UK	106	2020-02-05	249619	1116	35786	212717
Turkey	66	2020-03-16	152587	113987	4222	34378
India	107	2020-02-04	112028	45422	3434	63172

3.2.2 Forecasting Results of ARIMA and FB-Prophet

This section presents forecasting performance of the ARIMA and the FB-Prophet models for active, recovered, deceased, and confirmed COVID-19 cases. The models are implemented in Python 3.8. The models were evaluated using standard metrics including MAE, MAPE, RMSE, and RRSE. The analysis incorporated datasets from ten countries: the USA, Spain, Italy, France, Germany, Russia, Iran, the UK, Turkey, and India.

Forecasting of active cases

Active cases represent the number of infected individuals currently under medical supervision. The ARIMA and FB-Prophet models are evaluated on the active cases of COVID-19. As observed in section 3.1 that time series data of active cases is non-stationary. Since the ARIMA requires stationary data, preprocessing techniques are applied to transform the non-stationary active cases data,. First, one-lag differencing is carried out, followed by square root scaling, to achieve stationarity. The stationarity of the transformed data series was verified using the Dickey-Fuller test. Additionally, Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) plots are analyzed to determine appropriate values for order p and q of the ARIMA. In contrast, FB-Prophet does not require data preprocessing and is applied directly to the raw data.

Table 3.2: Forecasting results of the models for the COVID-19 active cases.

Region	Model	MAE	RMSE	RRSE	MAPE
Worldwide	ARIMA(9,1,2)	19141.89	21377.14	0.086	0.816
	FB-Prophet	168452.05	182230.63	0.706	6.943
US	ARIMA(10,1,3)	5732.16	8050.31	0.079	0.586
	FB-Prophet	95766.22	108424.76	1.07	9.12
Spain	ARIMA(8,1,4)	2191.68	2603.02	0.346	3.293
	FB-Prophet	67132.86	69748.42	9.274	109.40
Italy	ARIMA(9,1,3)	3197.25	4266.60	0.320	3.411
	FB-Prophet	26934.34	30963.76	2.325	35.55
France	ARIMA(5,1,4)	10974.15	11489.85	6.166	11.75
	FB-Prophet	44596.16	48195.48	25.864	48.340
Germany	ARIMA(11,1,4)	2114.09	2597.193	0.407	9.052
	FB-Prophet	50902.42	52259.90	8.197	277.26
Russia	ARIMA(10,1,2)	6456.26	6786.96	0.158	4.238
	FB-Prophet	36430.36	40232.57	0.936	20.748
Iran	ARIMA(4,1,2)	328.28	379.79	0.147	2.202
	FB-Prophet	12856.19	12902.11	5.009	82.503
UK	ARIMA(4,1,2)	8090.84	8637.25	0.375	4.66
	FB-Prophet	2954.65	4649.43	0.202	1.481
Turkey	ARIMA(8,1,2)	3631.37	3655.74	0.884	9.485
	FB-Prophet	59801.55	60725.11	14.678	158.59
India	ARIMA(11,1,5)	7007.09	7330.06	0.61	16.74
	FB-Prophet	10245.17	12085.37	1.005	21.429

Table 3.2 presents the forecasting accuracy results for active cases across 10

selected countries and worldwide. Results are shown with the optimal order (p,d,q) values calibrated for the ARIMA model. The ARIMA has shown superior performance using the specified (p,d,q) values in comparison with the FB-Prophet, as shown in the table.

Forecasting of recovered cases

Forecasting performance evaluation study is conducted of the adopted models using the recovered COVID-19 cases data from 10 selected countries and worldwide. Stationarity techniques are applied to prepare the stationary time series to perform the forecasting study using the ARIMA. Conversely, FB-Prophet was directly applied to the raw data without preprocessing to fit the model and generate forecasts.

Table 3.3: Forecasting results of the models for COVID-19 recovered cases.

Region	Model	MAE	RMSE	RRSE	MAPE
Worldwide	ARIMA(9,1,2)	34932.99	36992.53	0.128	2.523
	FB-Prophet	185584.71	214741.61	0.712	12.49
US	ARIMA(5,1,2)	31899.89	33109.68	0.667	15.635
	FB-Prophet	53970.19	57816.45	1.165	24.174
Spain	ARIMA(8,1,4)	9683.45	9774.06	0.786	7.361
	FB-Prophet	3021.53	3766.69	0.303	2.22
Italy	ARIMA(9,1,3)	12910.06	13078.23	0.693	12.78
	FB-Prophet	8721.87	10057.88	0.533	7.881
France	ARIMA(3,1,1)	5780.87	5853.29	1.21	10.574
	FB-Prophet	7323.90	8362.88	1.729	12.613
Germany	ARIMA(5,1,3)	13702.61	13901.04	1.287	9.808
	FB-Prophet	25017.26	28763.20	2.664	16.969
Russia	ARIMA(4,1,0)	2376.69	3212.50	0.141	5.103
	FB-Prophet	26988.80	33858.56	1.484	60.964
Iran	ARIMA(1,1,1)	4213.14	4496.75	0.736	4.933
	FB-Prophet	5638.72	6037.87	0.988	6.267
UK	ARIMA(4,1,2)	78.19	91.12	1.177	8.311
	FB-Prophet	69.11	79.44	1.026	7.326
Turkey	ARIMA(8,1,2)	4242.09	4333.57	0.44	4.321
	FB-Prophet	45986.27	46211.42	4.688	45.536
India	ARIMA(2,1,0)	721.17	1066.65	0.096	2.911
	FB-Prophet	11395.90	14381.55	1.295	42.882

Table 3.3 presents the forecasting performance results for the recovered cases. The prediction results indicate that ARIMA prediction closely matched the actual

values. The ARIMA achieved acceptable MAPE values, with a maximum of 15.6 and a minimum of 2.5 on all the datasets, whereas, FB-Prophet achieved MAPE values with a maximum of 31.822 and a minimum of 3.759 on all the datasets.

Forecasting of death cases

The COVID-19 pandemic has claimed many lives, making it essential to analyze the fatality rate and forecast future cases to help governments take proactive measures. In this section, the forecasting models have been evaluated using death cases across the selected countries and worldwide. Time series data of death cases is transformed into a stationary form for the ARIMA modeling using preprocessing techniques. The FB-Prophet model is applied directly to the raw data to generate predictions.

Table 3.4: Forecasting results of the models for COVID-19 fatality cases.

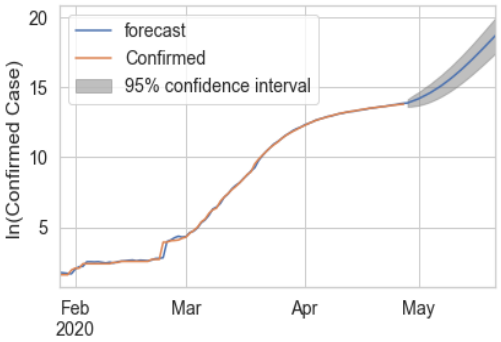
Region	Model	MAE	RMSE	RRSE	MAPE
Worldwide	ARIMA(9,1,2)	661.98	821.20	0.026	0.257
	FB-Prophet	21666.12	24874.45	0.735	7.465
US	ARIMA(2,1,0)	1924.08	1988.71	0.19	2.571
	FB-Prophet	4799.16	5856.54	0.56	5.751
Spain	ARIMA(2,1,0)	940.85	953.08	0.907	3.577
	FB-Prophet	2573.67	2961.76	2.818	9.525
Italy	ARIMA(2,1,0)	1240.10	1254.32	0.892	4.128
	FB-Prophet	3008.94	3433.46	2.443	9.703
France	ARIMA(3,1,1)	1335.79	1355.02	0.983	5.139
	FB-Prophet	6545.93	7270.98	5.274	24.382
Germany	ARIMA(1,1,0)	318.04	341.57	0.668	4.382
	FB-Prophet	1446.64	1668.89	3.262	18.761
Russia	ARIMA(2,1,0)	43.31	48.98	0.082	2.252
	FB-Prophet	628.39	709.50	1.184	30.597
Iran	ARIMA(1,1,1)	836.66	836.86	2.929	12.487
	FB-Prophet	257.44	291.70	1.021	3.759
UK	ARIMA(2,1,0)	959.53	984.02	0.343	3.119
	FB-Prophet	4171.84	4867.84	1.699	12.639
Turkey	ARIMA(8,1,2)	113.54	117.61	0.619	2.909
	FB-Prophet	280.83	312.96	1.647	6.945
India	ARIMA(2,1,0)	48.94	60.75	0.085	2.704
	FB-Prophet	771.35	897.58	1.26	31.822

Table 3.4 presents the prediction accuracy for the death cases due to COVID-19. The results show that ARIMA consistently achieves low prediction errors

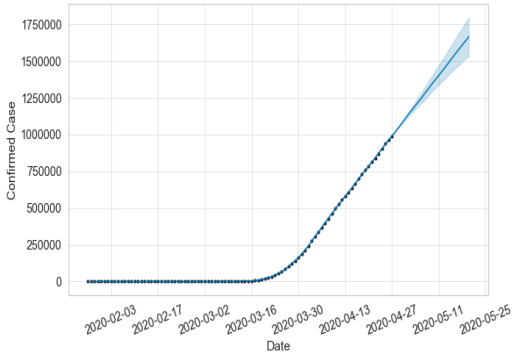
in comparison with the FB-Prophet. These findings indicate that ARIMA is more suitable for accurate forecasting of death cases, enabling better planning of necessary services by the governments.

Forecasting of confirmed cases

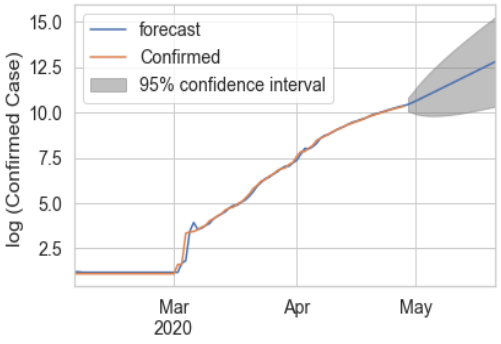
Forecasting accuracy of ARIMA and FB-Prophet models using COVID-19 confirmed cases data is visualized in this section. To achieve this, COVID-19 confirmed cases of two countries, namely, the USA and India are used for the analysis. Prediction results for both ARIMA and FB-Prophet models are illustrated in Figure 3.5.



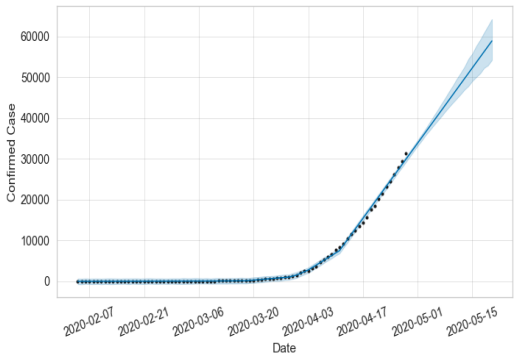
(a) ARIMA Forecasting for USA confirmed cases



(b) FB-Prophet Forecasting for USA confirmed cases



(c) ARIMA Forecasting for India confirmed cases



(d) FB-Prophet Forecasting for India confirmed cases

Figure 3.5: Visualization of prediction results of ARIMA and FB-Prophet for COVID-19 confirmed cases of India and the USA.

In Figure 3.5a, forecasting of confirmed cases of the USA is shown with 95% confidence interval using ARIMA model. The FB-Prophet model also demonstrates a good fit for the USA data, as shown in Figure 3.5b. Similarly, forecasting performance of the ARIMA and FB-prophet on India dataset is depicted in Figure 3.5c and 3.5d, respectively. FB-Prophet is based on progressive modeling, making it suitable for datasets with limited data points. In contrast, ARIMA requires a sufficient amount of data to effectively model and generate accurate predictions.

3.3 Forecasting using ARIMA and LSTM

The ARIMA and LSTM are the highly utilized models for COVID-19 forecasting and modeling. Therefore, in this section, comparative study of ARIMA and LSTM is carried out.

In the experimental setup, ARIMA(p, d, q) parameters are set as; p = 3, d = 0, and q = 1. The stationarity of the time series was examined using the Dickey-Fuller test [54]. The appropriate p and q orders for the ARIMA model were identified through analysis of ACF and PACF plots.

Following [87], the LSTM architecture incorporates a memory cell with three regulatory gates: input, forget, and output. These gates collectively control information flow through the network, while the memory cell maintains temporal state information. LSTM is widely used for long-term forecasting studies due to its robust performance and efficiency. Therefore, LSTM modeling is carried out to check accuracy, flexibility, and robustness on COVID-19 time series data.

The hyperparameter values for the LSTM model are configured as follows: optimizer = Adam, hidden units = 50, dropout rate = 0.2, dense units = 1, activation function = linear, epochs = 500, and MSE as the evaluation metric.

3.3.1 Modeling Dataset

The initial distribution of daily COVID-19 infected cases during the early stages of the pandemic was highly dynamic and non-linear. Therefore, it is an ideal dataset for evaluating the forecasting performance of time series forecasting models. In this view, ARIMA and LSTM models are assessed using one year of daily reported COVID-19 cases data from the USA. The adopted time series period is from April 1, 2020, to March 31, 2021. Smoothing of the data is performed using weekly averaging. The dataset is divided into training and testing samples in the ratio of 80:20.

3.3.2 Forecasting Results of ARIMA and LSTM

In this section, the LSTM is investigated in comparison with the ARIMA model for the COVID-19 forecasting. Forecasting result of ARIMA and LSTM for the USA dataset is shown in Figure 3.6. The figure demonstrates that both ARIMA and LSTM effectively adapt to multi-wave scenario of the pandemic spread. LSTM outperformed ARIMA in case of forecasting the unseen samples during the testing phase. It can be concluded that LSTM is more adaptable to non-stationary time series distributions in comparison with the ARIMA.

Table 3.5: Forecast accuracy results of ARIMA and LSTM for the COVID-19 time series data pertaining to the USA.

Model	MAE	RMSE	Correlation	MAPE
ARIMA(3, 0, 1)	76577	83352	0.959	1.570
LSTM	3105	4059	0.999	0.040

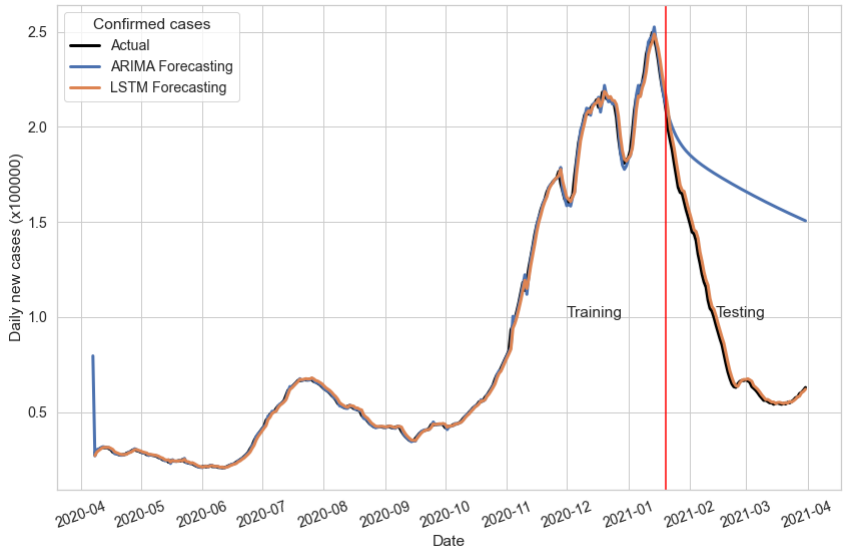


Figure 3.6: Forecasting of daily new cases of the USA using ARIMA and LSTM.

To delve deeper into the analysis, the forecasting results are evaluated using the metrics presented in Table 3.5. Each metric plays a crucial role in assessing the performance dynamics of a model. The table reveals that LSTM outperformed the ARIMA model.

3.4 Time complexity of ARIMA and LSTM

The experimental results in section 3.3 show that the LSTM has outperformed the ARIMA model. Further, we can compare these models on the measure of time complexity which can play a significant role in the model selection for a particular problem. The time complexity of a model depends on several factors, including the length of the time series data, model parameters, and the algorithm used for parameter optimization [205].

3.4.1 ARIMA time complexity

ARIMA(p,d,q) model training time complexity is generalized by $\mathcal{O}(n.k^2)$ [131, 26].
where,

- n is the length of the time series,
- $k = max(p, q)$, the maximum of the autoregressive and moving average orders.

Libraries like *statsmodels* use parameter optimizations which can reduce computational cost. If automatic model selection using grid search over the parameters (p, d, q) is applied, then the training complexity of the ARIMA is given by $\mathcal{O}(n.k^2.g)$.
where,

- g is the number of (p,d,q) combinations evaluated to find optimal values.

Generally the optimizers take 50 to 200 iterations. We will assume worst case iterations in the computational cost of the models which is 200.

3.4.2 LSTM time complexity

Assuming that training of a LSTM with a batch size of 1, sequence length n, and input and output vectors of constant size, then training time complexity of a multi-layer LSTM is given by $\mathcal{O}(e.n.h^2.L)$ [76].

where,

- e = number of training epochs,
- n = sequence length,
- h = number of hidden units,
- L = number of layers

3.4.3 Training time complexity comparison of ARIMA and LSTM

Table 3.6 includes training time complexities of ARIMA and LSTM models carried out in section 3.3.

Table 3.6: Comparative analysis of training time complexity of ARIMA and LSTM.

Model	Training time complexity	Parameters value	Computational cost
ARIMA	$\mathcal{O}(n.k^2.g)$	n=287, k=3, g=200	172200
LSTM	$\mathcal{O}(e.n.h^2.L)$	n=287, e=500, h=50, L=3	21525000

It can be observed from the outcomes in Table 3.6 that the ARIMA is faster then LSTM, and best suited for linear and stationary data. The LSTM has much higher computational cost but it can handle non-linear, non-stationary, and multi-variate sequences better than ARIMA. The LSTM is more suitable for large datasets or complex problems.

3.5 Forecasting using ETS, ANN, ARIMA and LSTM

In this section, four state-of-the-art models, namely, Error-Trend-Seasonality (ETS) [232], ANN [157], ARIMA, and LSTM are compared using COVID-19 forecasting. Five different datasets containing more than one wave of the COVID-19 outbreak are used for the forecasting performance evaluation.

The ?? is a statistical model for time series forecasting. It employs exponential smoothing techniques for forecasting. It works by decomposing a time series into

three main components, namely, 1) Error (E), 2) Trend (T), and 3) Seasonality (S). It is able to effectively handle seasonality and trends in a time series data.

ANNs are computationally modeled after the biological neural architecture of the human brain. This design enables ANNs to perform pattern recognition and predictive modeling tasks. It consists of mainly three layers, namely, 1) Input layer which processes the initial data, 2) Hidden layer(s) which processes inputs through weighted connections and applies activation functions, 3) Output layer which provides the final result.

3.5.1 Modeling Dataset

Five countries significantly impacted by COVID-19, namely, the USA, India, the UK, Russia, and Italy are selected to compare the performance of ETS, ANN, ARIMA, and LSTM models. The time series data of confirmed COVID-19 cases is taken from the GitHub repository [94], maintained by CSSE-JHU, USA. The datasets consist of cumulative COVID-19 confirmed cases covering multiple waves of the pandemic. The Time series data from June 1, 2020, to April 15, 2022 is utilized for the modeling. Data preprocessing is performed using three key techniques: 1) Differencing to eliminate trends, 2) Outlier detection and removal using the Interquartile Range (IQR) method [224], and 3) Weekly averaging to smooth fluctuations in the time series. The datasets are partitioned into training (70%), validation (15%), and testing (15%) subsets for experimental evaluation.

3.5.2 Comparative analysis of ETS, ANN, ARIMA and LSTM

Performance comparison of ETS [232], ARIMA[230], ANN [157], and LSTM [6] models is carried out in this section. ETS model focuses on trend, seasonality, and error. It is best suited for time series with consistent patterns.ARIMA is best on linear relationships and trends. It requires stationarity to perform forecasting. ANN is a non-linear model that uses interconnected neurons to learn patterns from data. LSTM networks, a specialized variant of recurrent neural networks, are specifically designed to model long-range temporal dependencies in sequential data. In this section, these models are trained using training samples and validated using the validation samples. Comparison is performed using testing samples.

The hyperparameter values for the LSTM and ANN are configured as follows: optimizer = Adam, learning rate = 0.005, loss function = Huber, hidden units = 10, dropout rate = 0.2, dense units = 1, activation function = linear, epochs = 500, and MSE as the evaluation metric. Forecasting performance is compared using five type of metrics namely, MAE, RMSE, RRSE, sMAPE, and MASE.

Comparison on India dataset

Table 3.7 presents the forecasting performance of the evaluated models on the test set of the India dataset, with the best-performing metrics highlighted in bold. The results indicate that the ANN model demonstrated superior performance across three of the five evaluation metrics. Meanwhile, the ARIMA model achieved the lowest RMSE, and the LSTM model performed best in terms of RRSE.

Table 3.7: Forecasting accuracy of the models on India dataset (best values are shown in bold)

Model	India				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(MAdM)	60731 ± 0	69662 ± 0	1.249 ± 0	113.66 ± 0	42.41 ± 0
ARIMA(6,1,0)	61255 ± 0	69522 ± 0	1.217 ± 0	122.25 ± 0	115.33 ± 0
ANN	59067 ± 367	65644 ± 423	**1.215 ± 0.41**	**104.05 ± 4**	1.48 ± 0.14
LSTM	**8280 ± 349**	**64319 ± 439**	1.222 ± 0.36	106.92 ± 5	**1.34 ± 0.15**

Comparison on UK dataset

The forecasting performance of the models is assessed using time series data of COVID-19 confirmed cases from the UK. The dataset encompasses multiple waves of COVID-19 within the chosen time period. The results of the forecasting performance on testing samples are presented in Table 3.8.

Table 3.8: Forecasting accuracy of the models on the UK dataset (best values are shown in bold)

Model	UK				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	30953 ± 0	37424 ± 0	2.835 ± 0	90.07 ± 0	26.92 ± 0
ARIMA(4,1,30)	25802 ± 0	35344 ± 0	2.742 ± 0	92.77 ± 0	13.74 ± 0
ANN	24588 ± 297	32149 ± 368	2.623 ± 0.14	82.20 ± 13	1.53 ± 0.12
LSTM	**24374 ± 285**	**31263 ± 345**	**2.42 ± 0.15**	**81.92 ± 14**	**1.49 ± 0.15**

From the results, it is evident that the LSTM model outperformed the others on three out of five metrics, namely, MAE, sMAPE, and MASE. ARIMA achieved the best performance on the RMSE metric, while ETS excelled on the RRSE metric. These findings suggest that LSTM is better equipped to handle multi-wave time series data compared to the other models.

Comparison on Russia dataset

The forecasting performance results of the models on the testing samples of the Russia dataset are presented in Table 3.9. The table reveals that ARIMA outperformed the other models on three out of five metrics, while LSTM demonstrated superior performance on the MAE and MASE metrics.

Table 3.9: Forecasting accuracy of the models on the Russia dataset (best values are shown in bold)

Model	Russia				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	20935 ± 0	26398 ± 0	2.797 ± 0	82.45 ± 0	27.39 ± 0
ARIMA(5,1,10)	23236 ± 0	25352 ± 0	3.171 ± 0	79.86 ± 0	110.61 ± 0
ANN	20356 ± 158	24443 ± 185	1.802 ± 0.29	76.89 ± 3	1.82 ± 0.14
LSTM	**20267 ± 169**	**23831 ± 193**	**1.766 ± 0.24**	**75.45 ± 2**	**1.67 ± 0.11**

Comparison on Italy dataset

The forecasting performance results of the models on the testing samples of the Italy dataset are presented in Table 3.10. The table indicates that LSTM outperformed the other models on four out of five metrics, while ANN achieved the best performance on the MASE metric.

Table 3.10: Forecasting accuracy of the models on the Italy dataset (best values are shown in bold)

Model	Italy				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	36716 ± 0	38917 ± 0	9.436 ± 0	130.83 ± 0	75.06 ± 0
ARIMA(1, 1, 0)	35073 ± 0	37805 ± 0	7.832 ± 0	97.21 ± 0	52.77 ± 0
ANN	34352 ± 348	37518 ± 397	6.67 ± 1.09	**84.62 ± 6**	2.39 ± 0.16
LSTM	**34012 ± 334**	**37114 ± 389**	**6.51 ± 1.12**	89.13 ± 8	**2.27 ± 0.15**

Comparison on USA dataset

The forecasting performance results of the models on the testing samples of the USA dataset are presented in Table 3.11. The table shows that ANN outperformed the other models on three out of five metrics, while LSTM achieved the best performance on the RMSE and RRSE metrics.

Forecasting trend of the compared models

The forecasting results of the models are depicted in Fig. 3.7 to visualize the forecasting performance on the USA dataset of COVID-19 confirmed cases. The performance of ETS, ARIMA, ANN, and LSTM is illustrated in Figures 6.5a, 6.5b, 6.5c, and 6.5d, respectively. From Figures 6.5a and 6.5b, it is evident that

Table 3.11: Forecasting accuracy of the models on the USA dataset (best values are shown in bold)

Model	USA				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	123115 ± 0	145967 ± 0	1.261 ± 0	94.74 ± 0	30.24 ± 0
ARIMA(2,1,0)	157951 ± 0	143898 ± 0	1.735 ± 0	90.55 ± 0	140.63 ± 0
ANN	**113179 ± 364**	141847 ± 396	1.273 ± 0.19	90.37 ± 8	**1.52 ± 0.28**
LSTM	128375 ± 328	**138425 ± 384**	**1.252 ± 0.17**	**88.12 ± 6**	1.62 ± 0.29

the ETS and ARIMA models are not able to handle the multiple waves of the pandemic. In contrast, the ANN and LSTM models effectively capture temporal dependencies, demonstrating robustness in dynamic situations.

3.6 Statistical significance test

Statistical significance tests are the methods to reach a conclusion to support or reject a claim (also called a hypothesis) based on sample data. A claim can be accepted if it is statistically significant at some significance level. A threshold value termed as p-value is to be selected before the test is performed. Traditionally, p-value is kept 1% or 5%. In our analysis, we kept p-value at 5%.

In this section, Friedman test [68, 52] is used followed by Nemenyi post-hoc test [158] for the statistical significance testing of the performance of the experimented models in section 3.5.2. The Friedman test determines whether or not there is a statistically significant difference between the performance mean of the compared models on the evaluation datasets. Null hypothesis (H_0) and alternative hypothesis (H_a) for the test are defined as follows.

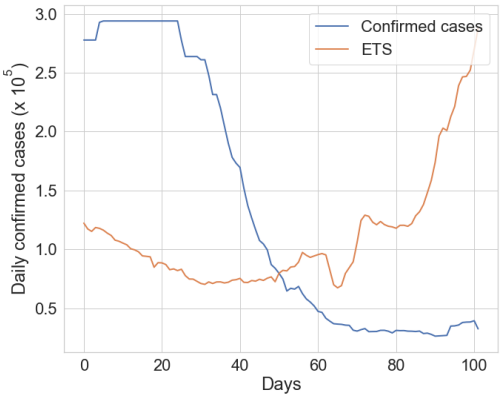
- H_0: The compared models are equivalent, and the mean performance of each model is same.
- H_a: The mean performance of at least one model is different from the rest.

Equations 3.6.1 and 3.6.2 are used for test statistics of the Friedman test.

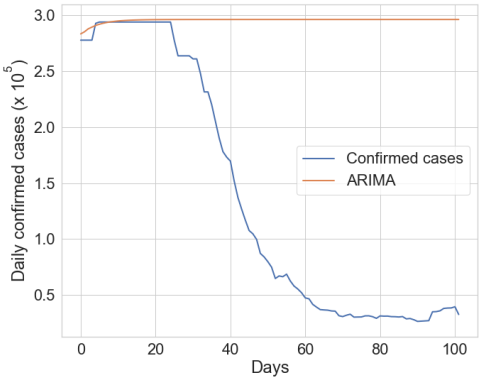
$$\chi_F^2 = \frac{12d}{m(m+1)} \left[\sum_{j=1}^{m} r_j^2 - \frac{m(m+1)^2}{4} \right] \tag{3.6.1}$$

$$F_F = \frac{(d-1)\chi_F^2}{d(m+1) - \chi_F^2} \tag{3.6.2}$$

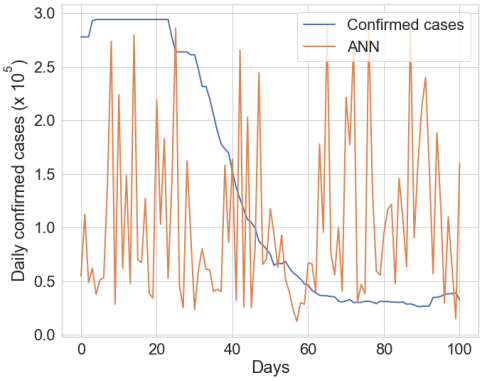
where m is number of models, d is number of datasets, and r_j is the average rank of each model.



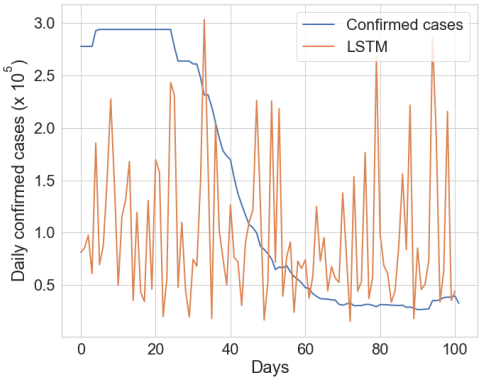
(a) USA forecast using ETS



(b) USA forecast using ARIMA



(c) USA forecast using ANN



(d) USA forecast using LSTM

Figure 3.7: Comparative forecasting results of the proposed model for the COVID-19 confirmed cases of the USA

The RMSE results included in Table 3.12 are taken from section 3.5.2 for statistical significance testing of the four compared models on five datasets. Therefore, $m = 4$, and $d = 5$. Value of average rank r_j of each j^{th} model is identified from Table 3.12. For example average rank of ARIMA model from the table is $\frac{3+3+3+3+3}{5} = 3$. Chi-square statistics result using the values is given below.

$$\chi_F^2 = \frac{12\times5}{4\times5} \times \left[(4^2 + 3^2 + 2^2 + 1^2) - \frac{4\times5^2}{4}\right]$$

$$\chi_F^2 = 15$$

The χ_F^2 follows a chi-square distribution with $m-1$ degrees of freedom.

$$F_F = \frac{(5-1)\times15}{5\times(4+1)-15} = 6$$

Now, check the critical value $q_\alpha(df1, df2)$ from the F-distribution [2] statistics. Where, α is confidence level, $df1$ is the degree of freedom across columns calculated as $(m-1) = 3$, and $df2$ is the degree of freedom across rows calculated as $(m-1)\times(d-1) = 3\times4 = 12$. The critical value at $q_{\alpha=0.05}(3, 12)$ is 3.49. The critical value is below the statistics obtained from χ_F^2 and F_F. Therefore, we reject H_0 with 95% confidence. In simple words, we have enough proof to say

[2] https://datatab.net/tutorial/f-distribution

Table 3.12: Forecasting accuracy (RMSE) of the models on five different datasets.

Dataset	RMSE (Relative Rank)			
	ETS	ARIMA	ANN	LSTM
USA	145967 (4)	143898 (3)	141847 (2)	138425 (1)
India	69662 (4)	69522 (3)	65644 (2)	64319 (1)
UK	37424 (4)	35344 (3)	32149 (2)	31263 (1)
Russia	26398 (4)	25352 (3)	24443 (2)	23831 (1)
Italy	38917 (4)	37805 (3)	37518 (2)	37114 (1)

that the compared models have statistically significant differences in the mean performance.

The same experiment can be carried out using available libraries in Python language. In this view, we conducted Friedman test using *stats.friedmanchisquare()* function of *scipy* library in python 3.9. Test result of *stats.friedmanchisquare(USA, India, UK, Russia, Turkey)* is *FriedmanchisquareResult(Friedman Statistic: 15, p-value: 0.0018)*. Since the p-value in the output is less than 0.05, we can reject the null hypothesis that the compared models perform equally.

Nemenyi test is a post-hoc test that compares the models after the null hypothesis from Friedman test is rejected. Nemenyi test identifies exactly which model differs from each other. The test makes pair-wise tests of performance. Hypothesis for Nemenyi post-hoc test is given below.

- H_0: All models perform equally.
- H_a: At least one pair of models is different.

The formula of critical difference (CD)for Nemenyi test is given in equation 3.6.3.

$$CD = q_\alpha \sqrt{\frac{m(m+1)}{6d}} \tag{3.6.3}$$

where q_α is the critical difference value from the Studentized range distribution at the significance level α divided by $\sqrt{2}$, and m is the number of groups. The value of $q_{\alpha=0.05} = 2.569$. The $q_{\alpha=0.05}$ value is taken from the online link [3]. Critical difference (CD) is calculated using the values as given below.

$CD_{\alpha=0.05} = 2.569 \times \sqrt{\frac{4\times5}{6\times5}} = 2.098$

Difference of highest and lowest average ranks from Table 3.12 is $(4-1) = 3$ which is greater than $CD = 2.098$. Hence, we can reject the null hypothesis H_0 of the Nemenyi test with 95% confidence. In simple words, we have enough proof to say that at least one pair of models is different. Pairwise difference of p-values are identified using *scikit-posthocs* library in python 3.9 which provides the posthoc_nemenyi_friedman() function for the Nemenyi test. The results

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6218068/table/pone.0206798.t009/

using Python library *scikit_posthocs.posthoc_nemenyi_friedman(dataframe)* is shown in Table 3.13.

Table 3.13: Nemenyi test p-values pairwise comparison.

Model	Nemenyi test p-values			
	ETS	ARIMA	ANN	LSTM
ETS	1.000000	0.597695	0.068212	0.001363
ARIMA	0.597695	1.000000	0.597695	0.068212
ANN	0.068212	0.597695	1.000000	0.597695
LSTM	0.001363	0.068212	0.597695	1.000000

The Table 3.13 highlights the statistical differences between the compared models. We can observe that LSTM is statistically significantly performed better than other compared models with the 95% confidence.

3.7 Analysis and Discussion

This chapter presents a comprehensive evaluation and comparative analysis of state-of-the-art forecasting models for COVID-19 prediction. Multiple datasets and diverse performance metrics were employed to assess model efficacy. Numerous aspects of the COVID-19 pandemic are analyzed using state-of-the-art methods and compared using diverse datasets spanning different timelines of the pandemic. In this chapter, it can be observed that LSTM model has outperformed the other models in most of the cases. It indicates that machine learning models are more adaptive to dynamic situations in comparison with the statistical models. Other than performance, time complexity analysis is also an importance area to discuss for comparative analysis of the models. In this view, time complexities of ARIMA and LSTM are analysed using an experimental study where ARIMA win the race but with lower performance. There are multiple factors which can impact the performance of a model. Therefore, it cannot be concluded that a model always perform best on all datasets and all metrics. The statistical significance test can help to reach a conclusion in such scenarios. In this view, statistical significance test is carried out for the compared models in this chapter.

CHAPTER 4

DEVELOPMENT OF NOVEL TIME SERIES FORECASTING MODEL FOR COVID-19 PREDICTION

[1] Fluctuations are an inherent characteristic of time series data, which can be analyzed to identify patterns and make predictions. Fuzzy time series (FTS) is more descriptive than traditional time series as it provides semantic interpretations for uncertain and fluctuating data. The forecasting efficiency of FTS models is primarily governed by three critical hyperparameters: (1) the cardinality of partitions in the universe of discourse (??), (2) the interval length of these partitions, and (3) the selected fuzzy order of the FTS model. Predicting COVID-19 cases is highly challenging due to involvement of multiple factors leading to dynamic time series data. In this context, studies such as [222, 62] have utilized FTS for COVID-19 prediction. However, there remains a need to explore additional hyperparameters of FTS to further improve forecasting accuracy for COVID-19 data.

The accuracy of a forecasting model is crucial for effective decision-making and planning. Developing a prediction model with optimized FTS hyperparameters for pandemic forecasting can significantly aid in policy formulation and controlling the spread of the disease. This chapter is motivated by the goal of designing an FTS-based COVID-19 forecasting model with optimized hyperparameters. Among the various optimization techniques proposed in the literature, the particle swarm optimization (PSO) algorithm [103] has emerged as a prominent focus of research in recent years, demonstrating significant potential across various optimization domains. Therefore, the PSO algorithm is utilized to identify optimal solutions for COVID-19 forecasting based on FTS.

This chapter presents two algorithms that integrate FTS and PSO:- 1) nested-FTS-PSO and 2) exhaustive-search-FTS-PSO. These algorithms aim to find the optimal combination of the number of partitions, the length of partition intervals in the UOD, and the fuzzy order. Two phases of the COVID-19 pandemic are considered for predictive analysis. The first phase, during 2020, was characterized by limited knowledge and lack of containment solutions. The second phase, during 2021, saw the availability of more information and the rollouts of vaccinations in many countries. The proposed forecasting approaches are evaluated against state-of-the-art models, including the classic FTS forecasting model [40], FTS-PSO [222], FB-Prophet [217], and ARIMA [230], using daily confirmed COVID-19 cases for both phases. This chapter includes the steps of FTS forecasting, an overview of the PSO algorithm, detailed descriptions of the proposed nested-FTS-PSO and

[1]The contents of this chapter are published in a research paper entitled "*Particle swarm optimization of partitions and fuzzy order for fuzzy time series forecasting of COVID-19*". Applied Soft Computing 110 (2021): 107611.

exhaustive-search-FTS-PSO algorithms, proposed forecasting framework, and the experimental results.

4.1 FTS and FLRG

Fuzzy time series (FTS) is framed using fuzzy sets. The concept of FTS was introduced in [208, 209, 238] studies.

To define a FTS, let the UOD be $U = \{u_1, u_2, ..., u_n\}$, where a fuzzy set A, specified on U is defined as given below.

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + ... + f_A(u_n)/u_n$$

where, f_A represents the membership function (MF) of A, denoted as given below.

$$f_A : U \rightarrow [0, 1]$$

The membership degree of the element u_i in the fuzzy set A is denoted by $f_A(u_i)$, where, $1 \leq i \leq n$.

Let crisp time series $Y(t), (t = ..., 0, 1, 2, ...)$, be a subset of $\mathbb{R}$. The fuzzy sets $f_i(t), (i = 1, 2, ...)$ are defined on the UOD U. Let $F(t)$ consists of $f_1(t), f_2(t), f_3(t), ...$, then, $F(t)$ is called a fuzzy time series defined on crisp time series $Y(t)$. Here, $F(t)$ is represented as a linguistic variable and $f_i(t)$ represents possible linguistic values of $F(t)$.

Relationship between linguistic variables is defined using right arrow as shown below.

If, $F(t-1) \rightarrow F(t)$, then, $F(t)$ is called derived from $F(t-1)$.

If the maximum degree of membership of $F(t)$ corresponds to the fuzzy set A_j and $F(t-1)$ corresponds to the fuzzy set A_i then the fuzzy logical relationship (FLR) between $F(t-1)$ and $F(t)$ can be represented as given below.

$$A_i \rightarrow A_j$$

where A_i and A_j are called current state and next state of the FLR respectively, and it is a first order FLR.

Similarly, if

$$F(t-m), ..., F(t-2), F(t-1) \rightarrow F(t)$$

then the FLR is called m-order.

FLRs sharing the same left-hand side can be combined into a group, referred to as a fuzzy logical relationship group (FLRG).

4.2 FTS Forecasting Steps

FTS forecasting steps follows a set of steps as described in the article [40]. The steps are explained below.

- Step-1 Defining UOD: Consider $Y(t)$ as the given historical time series dataset. Let $D_{\min}$ and $D_{\max}$ represent the minimum and maximum values of $Y(t)$, respectively. The universe of discourse, U, can then be defined as $[D_{\min} - D_1, D_{\max} + D_2]$, where D_1 and D_2 are chosen as appropriate positive values.

- Step-2 Partitioning of UOD: In this step, the universe of discourse, U, is partitioned into n intervals of equal length, denoted as $u_1, u_2, \ldots, u_n$. These intervals are defined as given below.

$$u_i = (Umin + (i-1) * L, Umin + i * L]$$

  where, $1 \leq i \leq n$, and the length of each interval $L = (Umax - Umin)/n$

- Step-3 Defining fuzzy sets: Each interval determined in Step 2 is associated with a linguistic variable to represent different regions within the universe of discourse (UOD). For n intervals, there will be n corresponding linguistic variables. A fuzzy set A_i is defined for each linguistic variable, as given below.

$$A_i = \frac{a_{i1}}{u_1} + \frac{a_{i2}}{u_2} + ... + \frac{a_{in}}{u_n}$$

$$a_{ij} = \begin{cases} 1 & j = i \\ 0.5 & j = i-1 \text{ or } j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

  Here, $a_{ij} \in [0, 1]$ and $(1 \leq i \leq n, 1 \leq j \leq n)$, and the + symbol represents the set union operator. The value a_{ij} denotes the membership grade of u_j in the fuzzy set A_i. The membership values of the fuzzy set A_i are determined based on a_{ij}.

- Step-4 Fuzzification: In this step, each historical data point is mapped to an interval represented by a linguistic value. The primary rule for assigning a linguistic value, based on the corresponding fuzzy set, is that the interval with the highest membership grade is selected. Let $Y(t)$ denote the actual time series and $F(t)$ the fuzzy time series corresponding to $Y(t)$. According to the maximum membership rule, the fuzzy set A_1 has the highest membership grade within the interval u_1.

- Step-5 Fuzzy relationships and grouping: Using the definitions from Section 4.1, a relationship is formed such that $F(t-m), \ldots, F(t-2), F(t-1) \rightarrow F(t)$, which is referred to as an m-order fuzzy logical relationship (FLR). Here, $F(t-m), \ldots, F(t-2), F(t-1)$ represent the current state, while $F(t)$ represents the next state. In these relationships, the right-hand side is unique, meaning a linguistic value cannot appear more than once on the right-hand side. Relationships with the same current state can be grouped together, forming what is known as a fuzzy logical relationship group (FLRG).

- Step-6 Defuzzification: In this step, forecasted values are computed using FLRGs. The defuzzification process can be carried out following the rules given in [40], outlined below.

  (i) If there is only one fuzzy relation in the FLRGs, such as $A_i \rightarrow A_r$, where the maximum membership of A_r corresponds to the interval u_r, and the midpoint of u_r is m_r, then the forecasted value is m_r.

  (ii) If there are multiple fuzzy relations in the FLRGs, such as $A_i \rightarrow A_{r1}$, $A_i \rightarrow A_{r2}$, ..., $A_i \rightarrow A_{rn}$, where the maximum membership of A_{r1}, A_{r2}, ..., A_{rn} corresponds to intervals u_{r1}, u_{r2}, ..., u_{rn}, with the midpoints of the intervals being m_{r1}, m_{r2}, ..., m_{rn}, then the forecasted value is calculated as $\frac{m_{r1}+m_{r2}+\cdots+m_{rn}}{n}$.

  (iii) If there is no fuzzy relation in the FLRGs defined as $A_i \rightarrow \#$, where the maximum membership degree of A_i corresponds to the interval u_i, and the midpoint of u_i is m_i, then the forecasted value is m_i.

4.3 PSO Algorithm

Particle Swarm Optimization (PSO) has been widely used for optimizing the parameters of various learning models. PSO is swarm based evolutionary algorithm introduced by Kennedy and Eberhart [103, 102]. It is inspired by the behavior of fish schooling or bird flocking, enabling it to search for optimal or near-optimal solutions to complex problems without getting trapped in local minima. In PSO, a swarm of particles explores the n-dimensional search space of an optimization problem, with each particle representing a potential solution.

The position of i^{th} particle $(i = 1, 2, ..., K)$ is denoted as $X_i = [x_i^1, , x_i^2, ..., x_i^n]$, and its velocity as $V_i = [v_i^1, v_i^2, ..., v_i^n]$. Here, K is the number of particles in the swarm. Each particle moves through the search space in search of the optimal solution. Each particle retains its personal best position, P_{best_i}, which is the best position it has encountered so far. The best position in the entire swarm is recorded as the global best position, P_{gbest}. Initially, all particles are assigned random positions in the search space. The personal best P_{best_i} of the i^{th} particle and the global best P_{gbest} are updated until the predefined maximum number of iterations, t_{max}, is reached.

The velocity and position of i^{th} particle are updated using the following equations.

$$V_i^{t+1} = w^t * V_i^t + C1 * Rand() * (P_{best\_i} - X_i^t) + C2 * Rand() * (P_{gbest} - X_i^t) \quad (4.3.1)$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad (4.3.2)$$

$$w^t = w_{max} - \frac{t * (w_{max} - w_{min})}{t_{max}} \quad (4.3.3)$$

Where w_{max} and w_{min} are predefined values for the inertia weight, and w^t is the inertia weight at the t^{th} iteration. V_i^t and X_i^t represent the current velocity and position, respectively, of i^{th} particle in the t^{th} iteration. The velocity V_i^t is constrained within the predefined range $[-Vmax, Vmax]$. $C1$ and $C2$ are the cognitive and social coefficients, respectively, also known as the acceleration coefficients. The function $Rand()$ generates a random value between 0 and 1, based on a uniform distribution. PSO algorithm can be used to find the optimum interval lengths in UOD for a given number of partitions [120, 222].

4.4 Proposed Methodology

Integration of PSO with FTS denoted as FTS-PSO substantially enhances forecasting accuracy. In this context, the PSO algorithm is employed to optimize key hyperparameters of FTS, including the number of intervals, the length of the intervals, and the fuzzy order. In FTS forecasting problems, it is required to determine the optimum number of partitions along with the optimum length of partitions, to generate better forecasting results. Length of partitions and the number of partitions are dependent variables in a FTS forecasting problem. In such a scenario, nested operations can be used to optimize the dependent variables. It can be said that finding an optimum combination of number of partitions, length of partitions, and fuzzy-order is an integrated optimization problem. To address this, we propose two novel approaches for tackling this optimization problem. The two optimization approaches for FTS hyperparameters are, namely nested-PSO and exhaustive-search-PSO. These approaches are described in the following sections.

4.4.1 Nested-FTS-PSO

The Nested-PSO approach involves two PSO loops as described below.

(i) The inner PSO layer optimizes the fuzzy order and partition interval parameters while maintaining the fixed partition count determined by the outer PSO layer, subsequently returning the computed cost metric to the outer optimization loop.

(ii) The outer PSO layer determines the optimal number of partitions by minimizing the objective function value while adhering to predefined constraints.

Algorithm 1 : Nested FTS-PSO

initialize random number of partitions n from $[n_{min}, n_{max}]$ and velocity of all particles' of outer PSO which is denoted as PSO-1.
while the maximum iterations of PSO-1 is not reached **do**
  for particle q, $(1 \leq q \leq maxParticlesOfOuterPSO)$ **do**
    set number of partitions equal to n which is passed by PSO-1, and initialize position and velocity of all particles' of inner PSO which is denoted as PSO-2.
    sort particles' position vector into ascending order.
    while the maximum iterations of PSO-2 is not reached **do**
      for particle i, $(1 \leq i \leq maxParticlesOfInnerPSO)$ **do**
        create intervals by using current position of i_{th} particle.
        define linguistic values according to all intervals.
        fuzzify time series data.
        for fuzzy-order m, $(1 \leq m \leq maxFuzzyOrder)$ **do**
          create m order fuzzy relationships and fuzzy groups.
          calculate forecasting using defuzzification.
          calculate the MSE for fuzzy order m for the i_{th} particle.
        end for
        update the personal best m_{best_i} fuzzy order and P_{best_i} position of i_{th} particle according to the calculated MSE for PSO-2.
      end for
      update the global best m_{gbest} order and P_{gbest} position among all the particles according to the calculated MSE_{gbest} for PSO-2.
      update PSO-2 inertia weight w_{pso2} according to equation-4.3.3
      for particle i, $(1 \leq i \leq maxParticlesOfInnerPSO)$ **do**
        update particle i position according to equations 4.3.1 and 4.3.2
        limit position of i_{th} particle within $[n_{min}, n_{max}]$.
      end for
    end while
    return global best combination $(m_{gbest}, MSE_{gbest}, P_{gbest})$ to PSO-1
    update the personal best n_{best_q} number of partitions of particle q in PSO-1 according to the received MSE from PSO-2.
  end for
  update the global best n_{gbest} number of partitions among all the particles in PSO-1.
  update PSO-1 inertia weight w_{pso1} according to equation-4.3.3
  for particle q, $(1 \leq q \leq maxParticlesOfOuterPSO)$ **do**
    update particle q number of partitions according to equations 4.3.1 and 4.3.2
    limit number of partitions n for a particle q within $[n_{min}, n_{max}]$.
  end for
end while
return global best combination $(n_{gbest}, m_{gbest}, P_{gbest})$

The procedure for optimizing FTS hyperparameters using the nested-PSO approach is outlined in Algorithm-1, with the steps described as follows.

- Step 1: Initialize the parameters for the outer PSO, including a random selection of the number of partitions within a specified range.
- Step 2: Begin the iteration loop for the outer PSO, which is tasked with optimizing the number of partitions.
- Step 3: Within the outer loop, iterate over the particles. The number of partitions for each particle is passed to the inner PSO.
- Step 4: Divide the Universe of Discourse (UOD) into unequal intervals based on the received number of partitions.
- Steps 5–7: The inner PSO optimizes the lengths of the partition intervals using the position equation-4.3.2 during each iteration and for each particle.
- Steps 8–14: Perform FTS forecasting to identify the optimal combination of partition intervals and fuzzy order.
- In further steps, the optimal result from the inner PSO, comprising the best intervals, best fuzzy order, and best MSE, is then passed back to the outer PSO. The outer PSO updates the number of partitions based on the received MSE using the position equation-4.3.2.
- Repeat the procedure until convergence or until the maximum iteration threshold is met..

The final output includes the optimal combination of the number of partitions, partition interval lengths, and fuzzy order.

The use of nested loops in the proposed integration of FTS-PSO results in a non-linear increase in time complexity [199]. However, by analyzing all parameters in a single run and carefully selecting an optimal range for the loop parameters, the time complexity can be reduced. To address this, a new variant referred as exhaustive-search-PSO, is proposed which offers reduced time complexity. Detail of the exhaustive-search-PSO algorithm to optimize FTS hyperparameters is provided in next section.

4.4.2 Exhaustive-Search-FTS-PSO

The proposed exhaustive-search-PSO identifies the optimal combination of three FTS hyperparameters: (1) the number of intervals, (2) the length of the intervals, and (3) the fuzzy order. This method also employs two loops, where the inner loop functions similarly to that in the nested-PSO. However, the outer loop systematically iterates over a predefined search space for the number of partitions and retains the optimal result from all iterations. Compared to nested-PSO, this approach is more time-efficient, as it explores a fixed search space rather

than relying on a swarm of particles to explore all ranges. The steps for the exhaustive-search-PSO are outlined in Algorithm-2.

In Steps 1 and 2 of Algorithm 2, the search space for the number of partitions is initialized, and the iterations are set to determine the optimal value. For each specified number of partitions, PSO is applied to optimize the length of partition intervals and the fuzzy order, using the predefined number of iterations and particles as outlined in Steps 5 and 6 of the algorithm. During each iteration, particle positions are updated using Equation-4.3.2, and FTS forecasting is performed on the training data following Steps 7 to 11. The best result from PSO is recorded for each iteration across the search space of partition numbers. The final optimal result is obtained as the combination of the optimal number of partitions, partition interval lengths, and fuzzy order.

Algorithm 2 : Exhaustive search FTS-PSO

1: initialize search space of number of partitions $[n_{min}, n_{min}+1, ..., n_{max}]$
2: **for** partitions n, $(n_{min} \leq n \leq n_{max})$ **do**
3: initialize all particles' positions X_i, velocity V_i, and PSO parameters
4: sort particles' position vector into ascending order.
5: **while** the maximum iterations is not reached **do**
6: **for** particle i, $(1 \leq i \leq maxParticles)$ **do**
7: create intervals of UOD by using particle current position.
8: define linguistic values for the intervals.
9: fuzzify time series data.
10: **for** fuzzy-order m, $(1 \leq m \leq maxFuzzyOrder)$ **do**
11: create m-order fuzzy relationships and grouping.
12: calculate forecasting values using defuzzification.
13: calculate MSE for fuzzy order m for the i_{th} particle.
14: **end for**
15: update the personal best P_{best_i} position and m_{best_i} fuzzy order of the i_{th} particle based on MSE.
16: **end for**
17: update the global best m_{gbest} fuzzy order, and P_{gbest} position among all the particles according to the calculated MSE.
18: update inertia weight w according to equation-4.3.3.
19: **for** particle i, $(1 \leq i \leq maxParticles)$ **do**
20: update position of i_{th} particle according to equations 4.3.1 and 4.3.2.
21: **end for**
22: **end while**
23: update global best n_{gbest} number of partitions
24: **end for**
25: return global best combination $(n_{gbest}, m_{gbest}, P_{gbest})$

4.5 Experimental Setup and Evaluation

Modeling datasets, experimented setup and parameters, and evaluation results are presented in following sections.

4.5.1 Modeling Datasets

In this study, 10 countries are selected, namely, France, India, Germany, Iran, Italy, Spain, Russia, Turkey, the UK, and the USA for predicting COVID-19 confirmed cases. These countries were among the most severely impacted by the pandemic globally. The COVID-19 datasets were sourced from a GitHub repository [94] maintained by CSSE, USA. Evaluation study is carried out in two phases:- (i) initial phase (COVID-19 Phase-1) (ii) evolved phase (COVID-19 Phase-2).

- In Phase-1, cumulative daily infected case data from April 1, 2020, to October 30, 2020, is used for each country.
- In Phase-2, cumulative daily infected case data from January 1, 2021, to May 15, 2021 is used for India and the USA. It has been focused on the USA and India in Phase-2 because they were among the most significantly impacted countries during this period.

4.5.2 Demonstration of optimization problem

In this section, impact of number of partition and fuzzy order parameters in FTS modeling is shown using experimental studies. Identifying optimal values of these parameters combinedly is an optimization problem which needs to be addressed effectively. In this view, an experiment is carried out to show the impact of number of partition on the FTS forecasting accuracy. Table 4.1 presents the MSE results for varying numbers of partitions in the UOD. The table includes MSE values for predictions using first-order FTS of COVID-19 confirmed cases in the USA during the first phase, with the number of partitions ranging from 46 to 55. It is evident that increasing the number of partitions improves the accuracy of FTS predictions. However, adopting large value of partitions will be diminish the concept of FTS modeling. Therefore, identifying the optimal number of partitions remains a complex challenge. Optimization algorithms, such as PSO, can be employed to determine the optimal value of number of partitions in the UOD.

Table 4.1: FTS prediction accuracy with increasing number of partitions for the USA data

FTS	Partitions in UOD	46	47	48	49	50	51	52	53	54	55
	MSE ($\times 10^6$)	23.476	24.330	23.108	21.903	21.964	23.143	22.007	21.973	22.889	22.005

An additional experiment was conducted to examine the impact of fuzzy order on FTS forecasting accuracy. This experiment focused on predicting day-level

COVID-19 confirmed cases in the USA during Phase 1. The prediction accuracy, measured by MSE, is presented in Table 4.2 for a fixed number of partitions set to 50 and fuzzy order values ranging from 1 to 5. The range of fuzzy orders aligns with that used in the study by Chen et al. [41]. As shown in the table, the prediction accuracy improves with increasing fuzzy order, consistent with the findings reported in [222].

Table 4.2: FTS accuracy pattern with increasing number of partitions in the UOD

FTS	**Fuzzy Order**	1	2	3	4	5
	MSE ($\times 10^6$)	21.964	5.028	1.273	0.351	0.299

Next, an experiment was conducted to assess the combination of the number of partitions and fuzzy orders of FTS using confirmed cases data from Phase 1 of COVID-19 in the USA. The PSO algorithm was employed to optimize the length of the partition intervals in the UOD. The prediction accuracy results are measured in terms of MSE for various combinations. The obtained results for the experiment are presented in Table 4.3. From the results, it can be observed that forecasting accuracy initially improves with an increasing number of partitions but begins to fluctuate after a certain point. A similar trend is observed as the fuzzy order increases. This suggests that the optimal combination likely exists within the given ranges.

Table 4.3: Prediction accuracy with number of partitions and fuzzy order

Approach	**UOD Partitions**	**Fuzzy Order (MSE$\times 10^6$)**				
		1	**2**	**3**	**4**	**5**
FTS-PSO	46	18.680	4.956	0.637	0.511	0.270
	47	18.688	3.831	0.573	0.321	0.382
	48	18.393	4.976	0.700	0.339	0.328
	49	20.744	3.833	1.140	0.294	0.366
	50	20.996	3.423	0.443	0.392	0.285
	51	19.322	3.907	0.800	0.326	0.334
	52	19.555	3.130	0.677	0.278	0.287
	53	18.746	5.442	0.481	0.283	0.263
	54	18.221	3.592	0.454	0.318	0.246
	55	17.745	4.333	0.567	0.394	0.249

The experimental analysis suggests that the optimal values of hyperparameters can be identified by experimenting with various combinations using optimization techniques. However, determining all hyperparameters in an integrated manner is

a complex and computationally intensive task, making it an integrated optimization problem. To tackle these challenges, two approaches, namely, nested-PSO and exhaustive-search-PSO, are proposed to optimize the FTS hyperparameters effectively.

4.5.3 Evaluation of the proposed approaches

The proposed approaches are implemented using python 3.8. Performance metrics, MSE and MAPE, are used to measures the forecasting accuracy. Numerous ranges of the number of partitions, length of the partition intervals, and fuzzy orders are experimented to identify the optimal values.

The datasets of COVID-19 confirmed cases was used to evaluate the performance of the proposed and compared models. Experiments were conducted with the number of UOD partitions ranging from 30 to 90 in Phase 1 and from 20 to 60 in Phase 2, and the fuzzy order ranging from 1 to 5 for FTS. These parameters, along with the PSO parameters, are detailed in Table 4.4. The range for the number of UOD partitions was chosen such that its maximum value remains less than half the total data points in the UOD. These parameters were applied in FTS-PSO, nested-FTS-PSO, and exhaustive-search-FTS-PSO models to perform COVID-19 forecasting.

Table 4.4: Values of hyperparameters used in the experiments.

Method	Parameter	Value
FTS	Number of Partitions (COVID-19 phase-1)	{30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90}
	Number of Partitions (COVID-19 phase-2)	{20, 25, 30, 35, 40, 45, 50, 55, 60}
	Fuzzy order range	{1, 2, 3, 4, 5}
PSO	Number of particles	30
	Maximum number of iterations	50
	Inertia weight $[w_{min}, w_{max}]$	[0.4, 0.9]
	C1, C2	2

The maximum number of iterations in PSO for the experiments was determined by analyzing various MSE convergence patterns relative to iteration counts. The convergence graph of accuracy (MSE) with iteration counts, using the training data for confirmed cases in the US, is presented in Figure 4.1. As shown, accuracy improves as the number of iterations increases. In this example, the MSE value converges to an acceptable low value at an iteration count of 50.

The time complexity of the proposed techniques is also analyzed. Since nested loops are employed to optimize all hyperparameters in a single run, the time complexity increases non-linearly [199]. The runtime of the techniques is recorded on a system with the specifications outlined in Section 1.7. Measurements have

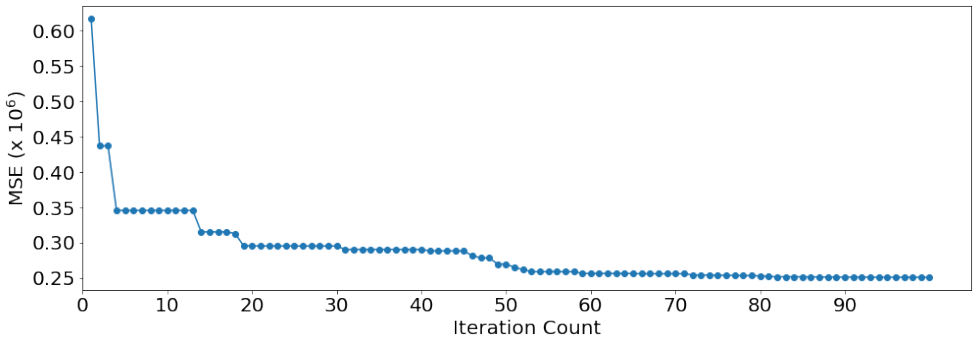


Figure 4.1: Convergence graph of PSO for COVID-19 confirmed cases time series of the USA.

Table 4.5: Measured runtime (seconds) for the proposed approaches.

Approach	Single runtime (A) in ms	Outer loop iterations (B)	Inner loop iterations (C)	Total runtime (A x B x C) in sec
Nested-PSO	37.869 ms	50	50	94.673 s
Exhaustive-search-PSO	37.869 ms	13	50	24.615 s

been taken for a fixed number of partitions (50), fuzzy orders ranging from 1 to 5, and a single particle for the Phase 1 COVID-19 data from the USA. The total runtime is estimated based on the number of iterations used in the outer and inner loops of the approaches. The results indicate that the runtime of exhaustive-search-PSO is significantly lower, due to the reduced number of iterations in its outer loop.

Predictions and comparative analysis of the proposed approaches are presented in the next section.

4.5.4 Experimental Results

The proposed approaches are compared against FB-Prophet [217], ARIMA [230], FTS [40], and FTS-PSO [120] using time series data from both phases of the COVID-19 outbreak. The datasets are split into training and testing sets, with the last 20 samples used as the testing set for Phase 1 and the last 15 samples for Phase 2. Prediction results are generated for the day-level cumulative confirmed COVID-19 cases across the 10 selected countries. Each experiment has been run 10 times, and the best result from these runs is selected as the final outcome.

COVID-19 forecasting for Phase-1 (Initial phase)

This section presents prediction results for the COVID-19 timeline spanning from April 2020 to October 2020, referred to as Phase 1. The predictions have been generated using the confirmed COVID-19 cases from the 10 selected countries, as detailed in Table 4.6. The results have been obtained using the partition and fuzzy order ranges specified in Table 4.4 for Phase 1. Prediction accuracy for the

test set is evaluated using the MAPE function across all compared models. The primary objective of this comparison is to assess the performance of the proposed approaches on diverse datasets.

Table 4.6 presents the forecasting results, with FB-Prophet predictions generated using daily seasonality. Among the compared models, FB-Prophet has shown the lowest performance. The ARIMA model, characterized by its three primary parameters, namely, order of autoregression (p), order of differencing (d), and order of moving average (q), has performed better than FB-Prophet. The optimal (p, d, q) parameters and corresponding prediction accuracy (MAPE) for ARIMA are listed in Table 4.6 for each country, aligning with findings from our earlier work [114].

Table 4.6: Forecasting Performance of the models for COVID-19 confirmed cases in Phase-1.

Country	Optimal hyperparameters for all methods and MAPE					
	FB-Prophet	**ARIMA**	**FTS**	**FTS-PSO**	**Nested-FTS-PSO**	**Exhaustive-search-FTS-PSO**
		(p, d, q)	(K, O)	(K, O)	(K, O)	(K, O)
France	18.833	(5, 1, 1) 0.766	(85, 4) 0.049970	(80, 3) 0.045644	(90, 5) 0.047976	(70, 5) 0.033117
Germany	17.195	(4, 1, 1) 0.868	(80, 4) 0.028425	(85, 5) 0.016065	(90, 4) 0.020281	(70, 5) 0.014956
India	5.162	(4, 1, 1) 0.315	(85, 3) 0.005724	(90, 2) 0.003991	(90, 5) 0.003837	(90, 5) 0.003050
Iran	5.164	(3, 1, 1) 0.025	(70, 3) 0.004722	(80, 5) 0.004520	(90, 5) 0.004662	(85, 5) 0.004417
Italy	22.589	(4, 1, 1) 0.877	(60, 5) 0.025832	(90, 4) 0.022416	(89, 5) 0.024141	(75, 5) 0.021285
Russia	8.389	(5, 1, 1) 0.778	(55, 2) 0.003989	(90, 4) 0.003450	(90, 5) 0.003731	(65, 5) 0.003314
Spain	3.407	(3, 1, 1) 0.173	(90, 5) 0.009503	(55, 4) 0.008251	(90, 5) 0.007637	(80, 5) 0.006277
Turkey	0.631	(5, 1, 1) 0.043	(75, 5) 0.005271	(80, 3) 0.004107	(90, 5) 0.008281	(75, 5) 0.003275
UK	26.075	(5, 1, 1) 0.128	(85, 3) 0.006306	(90, 4) 0.005590	(90, 5) 0.005791	(90, 5) 0.005188
US	3.236	(5, 1, 1) 0.213	(50, 5) 0.003933	(75, 2) 0.003726	(90, 4) 0.003891	(85, 5) 0.002699

For FTS forecasting, the two key parameters are the number of partitions (K) and the fuzzy order (O). The table highlights the best-performing (K, O) combinations that yielded optimal accuracy (MAPE) for FTS, FTS-PSO, nested FTS-PSO, and exhaustive FTS-PSO. In the case of FTS-PSO and the proposed techniques, PSO was employed to optimize the length of intervals for each param-

eter pair. The impact of PSO is evident in the significant accuracy improvements observed when comparing FTS with FTS-PSO in Table 4.6.

Nested FTS-PSO demonstrated moderate performance compared to the other approaches, though it exhibited signs of overfitting during the training phase, consistently following the same pattern across datasets. This suggests the need for a stopping condition before making predictions with nested FTS-PSO. Despite these limitations, the approach produced competitive results.

Exhaustive-search-FTS-PSO outperformed all other models for all the selected countries. It adapted effectively to the scenarios, delivering the most accurate forecasting results among the compared approaches.

COVID-19 forecasting for Phase-2 (Evolved phase)

Many countries experienced multiple waves of COVID-19 outbreaks. To account for this, a comparative prediction analysis of the adopted and proposed approaches are conducted for the later (evolved) phases of the pandemic. For this analysis, datasets from India and the USA are selected, as these countries were among the most severely affected during Phase-2 of COVID-19. Table 4.7 presents the prediction accuracy results for all approaches applied to both countries. The table shows that the exhaustive-search FTS-PSO consistently outperformed all other approaches on the Phase-2 COVID-19 timeline data. The pattern of performance is similar to the results in Table 4.6.

Table 4.7: Forecasting performance of the models for Phase-2 of COVID-19

Country	Optimal hyperparameters for all methods and MAPE					
	FB-Prophet	**ARIMA** (p, d, q)	**FTS** (K, O)	**FTS-PSO** (K, O)	**Nested-FTS-PSO** (K, O)	**Exhaustive-search-FTS-PSO** (K, O)
India	7.411	(0, 1, 1) 0.560	(60, 4) 0.009233	(60, 2) 0.008289	(60, 5) 0.008137	(50, 5) 0.005865
US	0.698	(5, 1, 1) 0.022	(50, 3) 0.003580	(60, 4) 0.002961	(60, 5) 0.003192	(60, 5) 0.002410

4.6 Analysis and Discussion

The hyperparameters of the FTS forecasting model include the number of partitions in the UOD, partition intervals, and fuzzy order. The prediction accuracy of any FTS forecasting model relies heavily on the proper tuning of these hyperparameters. Existing approaches in the literature have addressed FTS forecasting problems by optimizing the interval lengths and fuzzy order while typically setting the number of intervals to a fixed value.

In this chapter, the optimization of all three FTS hyperparameters is performed using variations of the PSO algorithm. Specifically, the nested-FTS-PSO and exhaustive-search-FTS-PSO algorithms are proposed to optimize the number of partitions, the length of partition intervals, and the fuzzy order. COVID-19 datasets from 10 highly affected countries are used to evaluate the performance of

the proposed models. These models are compared against ARIMA, FB-Prophet, conventional-FTS, and FTS-PSO.

Prediction results are presented for two distinct timelines: the initial (starting) phase and the evolved phase of COVID-19. The proposed approaches effectively identified the optimal combination of hyperparameters, with the exhaustive-FTS-PSO outperforming all other compared models across both timelines.

CHAPTER 5

DESIGN OF TIME SERIES FORECASTING MODEL FOR PREDICTING COVID-19 FOR MUTANT AFFECTED POPULATION

[1] The performance of classical time series forecasting models significantly declines when the data is non-stationary in nature [58, 190]. Research has shown that fuzzy time series (FTS) models are effective in handling highly uncertain data [133, 97]. A prime example of a non-stationary time series is the COVID-19 pandemic [96]. Several factors, including lockdowns, quarantines, healthcare facilities, population density, education, virus variants, vaccination efforts, and government policies, have influenced the spread dynamics of the COVID-19. The forecasting of COVID-19 cases is a challenging task for a single forecasting model due to the complexity introduced by the multiple factors.

The combination of FTS with deep learning (DL) techniques is a rapidly growing research area, as it enhances the interpretability and explainability of the system. In this direction, wang et al. [226] leveraged the benefits of the fuzzy concept in conjunction with LSTM for long-term forecasting. The study suggested that the fuzzy integration is able to address the limitations of LSTM. The hybrid model outperformed eight other state-of-the-art models, demonstrating its superiority. In another study [20], the authors integrated FTS with DL methods and proposed the use of butterfly optimization [12] to enhance the results. The inclusion of the optimization algorithm led to a significant improvement in the forecasting performance.

The LSTM has demonstrated its capability to handle nonlinear, time-varying data [130]. Higher-order FTS models offer improved forecasting accuracy compared to first-order FTS models. As a result, many recent studies focus on high-order FTS models [246, 140]. The PSO algorithm is able to find nearly optimal solutions without getting stuck in local minima [103]. A very few studies have explored the combination of high-order FTS with deep learning for forecasting of a pandemic spread. Therefore, this chapter investigates a hybrid model that integrates high-order FTS, PSO, and deep learning to address the dynamics of COVID-19 and enhance the forecasting accuracy.

5.1 FTS modeling

FTS forecasting models involve five distinct stages: (1) defining the universe of discourse (UOD), (2) determining the intervals within the UOD, (3) fuzzifying

[1]The contents of this chapter are published in research papers entitled "*Non-Stationary Fuzzy Time Series Modeling and Forecasting using Deep Learning with Swarm Optimization.*" in International Journal of Machine learning and Cybernetics. pp. 1-19, 2025. and another research paper entitled "*Epidemic Modeling using Hybrid of Time-varying SIRD, Particle Swarm Optimization, and Deep Learning*" in International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-7. IEEE, 2023

historical data, (4) establishing fuzzy logical relationships (FLRs) and organizing them into groups, and (5) optionally, defuzzification.

In stage-1, the lower (D_{min}) and upper (D_{max}) bounds of the time series are identified. The minimum and maximum values of the UOD are defined as $[D_{min} - D1, D_{max} + D2]$, where $D1$ and $D2$ are proper positive values. Stage-2 involves dividing the UOD into either equal or unequal intervals. Optimal length of intervals is crucial for achieving higher prediction accuracy. Stage-3 assigns linguistic variables to the intervals identified in Phase 2, and fuzzification of each historical data point is carried out by mapping it to the corresponding interval. For k intervals, there will be k linguistic variables. In stage-4, FLRs are established between the linguistic variables. The left-hand side of a relationship represents the present state, while the right-hand side represents the next state. These relationships may be of first order or higher. A linguistic value can appear only once on the right-hand side of a relationship. FLRs sharing identical present states are aggregated to constitute a fuzzy logical relationship group (FLRG). The order of FLRs is critical for efficient forecasting. Finally, stage-5 involves calculating the forecasted values using FLRGs, it is known as defuzzification. The crisp values can be computed using statistical methods or by a rule based conversion method.

Detailed overview of fuzzy concept and FLRs is provided in chapter 4.1. Steps of FTS forecasting are described in chapter 4.2. In this chapter, FTS is explored with deep learning models (LSTM and its variants) and optimization techniques (PSO).

5.2 LSTM and its variants for modeling FLRs

RNN contains loops between nodes which allows it to store temporal information, and information feedback for decision making. RNNs use internal memory to process the sequences of inputs but they suffer from gradient vanishing and exploding problem when dealing with long sequences [19]. To address this problem, LSTM network was developed [87]. LSTM units contain a memory cell and three gates (input, forget, output) that regulate information propagation. The memory cell preserves temporal state information throughout sequential processing. These gates control the flow of information within the network. The memory cell is responsible for retaining information over time. The input gate determines which information is added to the memory cell, while the forget gate discards information that is no longer relevant. The output gate regulates the information that is sent from the memory cell as output. This selective information flow allows the LSTM to retain important information and transmit only what is necessary through the network, enabling it to learn long-term dependencies effectively. The structure of LSTM network is shown in Figure 5.1.

Recurrent neural network (RNN) contains loops between nodes which allows it to store temporal information, and information feedback for decision making [168]. RNNs use internal memory to process the sequences of inputs but they suffer from gradient vanishing and exploding problem when dealing with long sequences [19]. To address this problem, long Short-Term Memory (LSTM) network was introduced in [87]. To address more complex tasks, several LSTM variants have

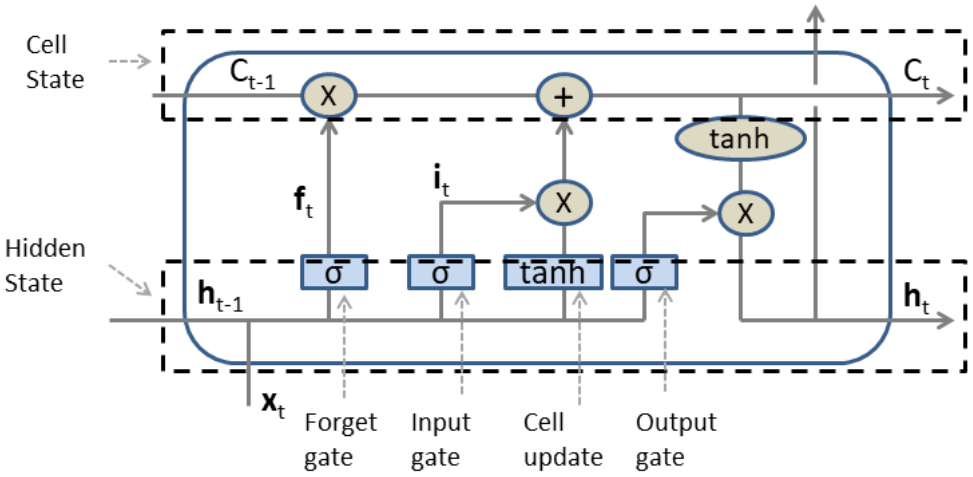


Figure 5.1: LSTM structure.

been developed, including stacked LSTM, bidirectional LSTM, convolutional LSTM, and attention-based LSTM

Researchers have developed stacked-LSTM model to improve accuracy of the prediction results. A stacked-LSTM model contains multiple LSTMs in the stacked form to process the input data one after another [64, 63]. To solve various real-world problems, different variants of LSTM have been introduced over time. Bi-directional LSTM (bi-LSTM) has been introduced to primarily deal with challenges of natural language processing. Basically, bi-LSTM is created by combining two LSTMs: one is used for forward direction, and the other one is for backward direction data flow [193]. Therefore bi-LSTM is capable of utilizing input flows in forward as well as backward directions. Convolutional neural networks (CNNs) are prominently used for image processing and pattern identification because it reduces frequency variation [214]. LSTM and CNN are complementary in their capability to solve time series problems. Ensemble of CNN and LSTM denoted as conv-LSTM which has been introduced to deal with the problems related to processing of sequential patterns or images to improve the results [49]. The addition of convolutional layers at the input helps to capture contextual information in the time series data. In artificial intelligence, the attention mechanism is a machine learning technique where the network focuses only on important and related part of the data [18, 211]. So, many researchers have experimented using LSTM with attention technique to improve the forecasting results by capturing contextual patterns in the input data [245, 2]. In this chapter, we investigate variants of LSTM model with convolutional and attention mechanisms for modeling of FLRs in FTS data.

5.3 PSO Variant

In this section, variant of PSO is described which is an improved version of the PSO described in Chapter 4.3. PSO is an evolutionary algorithm which aims to achieve an optimal or nearly optimal solution of a problem without getting

trapped into a possible local minimum solution [103, 119]. Each particle in a swarm represents a potential solution. A particle of a swarm swims through the search space of the optimization problem to find an optimal solution.

Let K be the number of particles in the swarm, and there is d-dimensional search space of an optimization problem, then the position of the $i^{th}(i = 1, 2, ..., K)$ particle is represented by $X_i = [x_i^1, , x_i^2, ..., x_i^d]$, and the velocity by $V_i = [v_i^1, v_i^2, ..., v_i^d]$. Initially, all particles are initialized with random values of the position vector in search space. Each particle swims through the search space, and keeps the personal best position P_{best_i} which has been recorded so far. Overall best position among the particles in the swarm is recorded as global best position P_{gbest}. Position and velocity of a particle are updated till the pre-defined condition is achieved or maximum iteration is reached. Following equation is used to update the inertia weight.

$$w^t = w_{max} - \frac{t * (w_{max} - w_{min})}{t_{max}} \tag{5.3.1}$$

where w^t is the inertia weight in the t^{th} iteration, w_{max} and w_{min} are pre-defined inertia weights. t_{max} is the maximum iteration count.
The acceleration coefficients $C1$ and $C2$ are adaptively adjusted based on the inertia weight w^t. To enhance local search capabilities, a larger $C1$ and a smaller $C2$ are employed, whereas a smaller $C1$ and a larger $C2$ are preferred for improved global search. This adaptive strategy is implemented by updating the acceleration coefficients according to the following equations.

$$C1^t = \frac{w^t}{w_{max}} \times (w^t + \sqrt{w^t} + 1) \tag{5.3.2}$$

$$C2^t = \frac{w_{max} - w^t}{w_{max}} \times (w^t + \sqrt{w^t} + 1) \tag{5.3.3}$$

where $C1^t$ and $C2^t$ are the acceleration coefficients in the t^{th} iteration. Their values are restricted within the range $[0.5, 2.5]$. Position and velocity of i^{th} particle are updated using the following equations.

$$V_i^{t+1} = w^t * V_i^t + C1^t * Rand() * (P_{best\_i} - X_i^t) + C2^t * Rand() * (P_{gbest} - X_i^t) \tag{5.3.4}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{5.3.5}$$

$$V_{max} = \frac{x_{max} - x_{min}}{t_{max}} \tag{5.3.6}$$

where V_i^t and X_i^t are the current velocity and current position respectively of a particle i in the t^{th} iteration. Maximum speed of a particle is estimated using equation 5.3.6 to ensure uniform speed of all dimensions. V_i^t is restricted to the range $[-V_{max}, V_{max}]$. $Rand()$ is a random function which is used to generate a value in the range [0,1] under uniform distribution. The hyperparameters (fuzzy order, length of intervals) associated with FTS models are determined using the PSO algorithm.

5.4 Proposed Methodology

Hybrid models using integration of FTS, PSO, and deep learning model are explored to handle non-stationary time series data for forecasting of the COVID-19 confirmed cases. FTS deals with uncertainties, and deep learning enhances the forecasting capability by incorporating the long term changes in the non-stationary data. PSO identifies the optimum parameters of FTS which results in improvement in the forecasting accuracy. Block diagram of the proposed framework is shown in Fig. 5.2. The framework consists of mainly four components; a.) data preprocessing, b.) FTS modeling using PSO optimization, c.) deep learning and forecasting, and d.) defuzzification and evaluation. As shown in Fig. 5.2, data preprocessing is performed by applying three method; 1. differencing, 2. weekly averaging, and 3. removal of outliers in training and testing samples.

The time series was split 70:15:15 for training, validation, and testing purposes respectively. After preprocessing, the crisp time series $Y(t)$ is obtained, which is then used for FTS modeling. The key hyperparameters for FTS include the number of fuzzy intervals, the fuzzy order, and the optimal configuration of unequal interval lengths. The effective number of intervals is determined using the average-based method proposed by Huarng [90]. PSO is employed to identify the optimal fuzzy order and the best configuration of unequal interval lengths. The data is then fuzzified through a fuzzification process. Based on the optimized fuzzy order, FLRs are constructed. Subsequently, a deep learning model is integrated to further model the fuzzified data. This model is trained and validated using the fuzzified data and the established fuzzy relationships. Forecasting is then performed on the testing set using the trained model. Finally, the forecasted fuzzy data points are converted back into crisp values through a defuzzification process. Step-by-step explanation is given in the following.

Step-1: Identify universe of discourse (UOD).

Let there be a crisp time series dataset $Y(t)$, where y_{min} and y_{max} are the minimum and maximum values in the dataset, then UOD is defined as $U = [U_{min}, U_{max}]$, where $U_{min} = y_{min} - D$, $U_{max} = y_{max} + D$, $D = 0.2 \times (y_{max} - y_{min})$. In this study, confirmed cases of the COVID-19 are counted as proper positive. So, value of U_{min} is set as 0 in the experiments whenever it is found negative. For example, if maximum value in a dataset is $D_{max} = 28$, minimum value is $D_{min} = 1$, then, identified proper positive value $D = 0.2 \times (28 - 1) = 6$, then $U = (y_{min} - D, y_{max} + D) = (1 - 6, 28 + 6) = (0, 34)$

Step-2: Determine the number of intervals.

The selection of the number of intervals is crucial for the predictive accuracy of a FTS model. An excessively large number of intervals can undermine the fundamental principles of FTS, while too few intervals may lead to reduced forecasting accuracy. So, in this proposal, an average-based method is adopted to determine the effective number of intervals [167]. In this method, the dataset $Y(t)$ is first sorted and duplicate values are removed to identify unique data points. Subsequently, one-lag differencing is applied to the sorted series to obtain an absolute difference series. The length of intervals is calculated using equation 5.4.1, and the number of intervals are find out using equation 5.4.2.

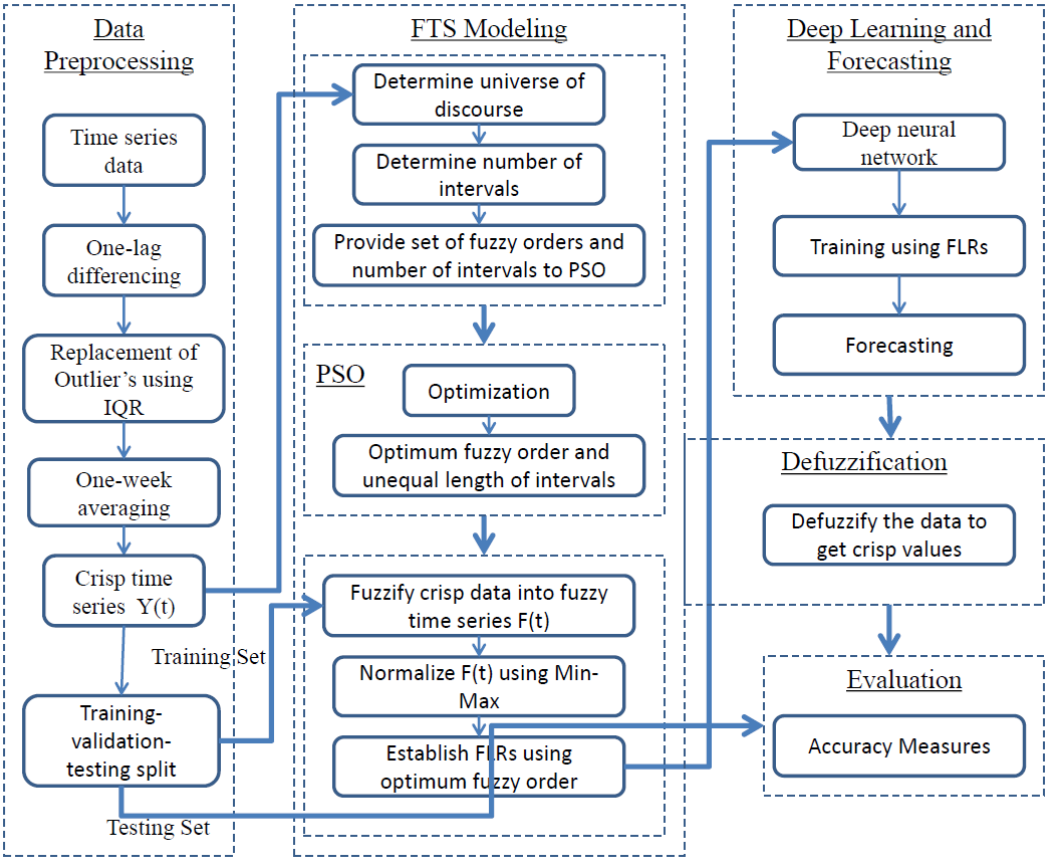


Figure 5.2: Proposed model framework.

$$interval\_len = \lfloor \frac{1}{2} \sum_{i=1}^{n} \frac{y_i}{n} \rfloor \tag{5.4.1}$$

$$\#intervals = \lceil \frac{U_{max} - U_{min}}{interval\_len} \rceil \tag{5.4.2}$$

For example, suppose a time series dataset is $Y(t) = \{1, 21, 3, 23, 3, 27, 8, 28, 12, 28\}$. After, the first lag differenced absolute series is $Y_d(t) = \{20, 18, 20, 20, 24, 19, 20, 16, 16\}$. Half of the average of $Y_d(t)$ is 9.61. By rounding $Y_d(t)$ the effective length of intervals is $interval\_len = 9$. Therefore, number of intervals will be $\#intervals = \lceil \frac{34-0}{9} \rceil = 4$.

Step-3: Determine optimum length of intervals and fuzzy order.
In this step, intervals are generated based on the number determined in Step 2. The boundary points for these intervals are created within the UOD. To obtain the optimal (unequal) interval lengths, the PSO algorithm is employed. Initially, each particle is assigned a set of boundary points as its coordinates. During each iteration, a particle's velocity is updated using Eq. 5.3.4, and its position (coordinates) is updated using Eq. 5.3.5. The resulting points are sorted in ascending order, and consecutive pairs are used to form intervals, which are then defined as fuzzy sets. Each fuzzy set is associated with a linguistic term

represented by an index value. Fuzzification of the crisp time series $Y(t)$ is performed by replacing each data point with the index of the fuzzy set to which it belongs, resulting in a fuzzy time series $F(t)$. Subsequently, fuzzy relations are established based on the selected fuzzy order, following standard FTS definitions. In this chapter, optimum fuzzy order is identified from the set of fuzzy orders $\{1, 2, 3, 4, 5\}$ using PSO. Each fuzzy order is experimented by integrating with PSO, and the best performing fuzzy order is identified. Finally, combination of optimum values of a fuzzy order and interval boundary points are returned.

Using example from step-2, 4 number of fuzzy intervals in the range of UOD having length 10 can be created as {[0, 9), [9, 19), [19, 29), [29, 34)}. Index values of the fuzzy intervals are {0, 1, 2, 3}. Assume that optimum fuzzy order for the example sample is 1.

Step-4: Time series Fuzzification and establishment of FLRs.

Using the optimal fuzzy intervals obtained in Step 3, the crisp time series $Y(t)$ is fuzzified to generate the fuzzy time series $F(t)$. In this context, a fuzzy forecast corresponds to the right-hand side value of a fuzzy relation. Each data point in $Y(t)$ is mapped to the index of its corresponding fuzzy set during fuzzification. The resulting fuzzified data is then normalized using min-max normalization, a common preprocessing step that enhances the performance and convergence speed of LSTM models. Finally, FLRs are established on the normalized data using the optimal fuzzy order identified in Step 3.

Using fuzzy intervals from the example in step-3, crisp time series $Y(t)$ is fuzzified to fuzzy time series $F(t) = \{0, 2, 0, 2, 0, 2, 0, 2, 1, 2\}$. We have applied min-max normalization on $F(t)$ in this study. But for the simplicity, here we are explaining without normalization. Using fuzzy order 1, fuzzy relations are defined on $F(t)$ as $0 \rightarrow 2, 2 \rightarrow 0, 0 \rightarrow 2, 2 \rightarrow 0, 0 \rightarrow 2, 2 \rightarrow 1, 1 \rightarrow 2$. After removing duplicates, remaining FLRs are $0 \rightarrow 2, 2 \rightarrow 0, 2 \rightarrow 1, 1 \rightarrow 2$. After grouping final FLRs are $0 \rightarrow 2, 1 \rightarrow 2, 2 \rightarrow 0, 1$

Step-5: Forecasting using deep learning.

In this step, FLRs are divided into training, validation, and testing sets. In this study, variants of LSTM deep learning model with attention and convolutional mechanisms are experimented. The model is trained on the training samples, results are validated on the validation set, and forecasting is performed on the test samples.

Step-6: Defuzzification.

Defuzzification is the process of converting fuzzified data back into crisp values. In Step-5, each fuzzy forecast is first denormalized and then rounded to the nearest integer, which serves as the index of a corresponding fuzzy set. The midpoint of the identified fuzzy set is taken as the final forecasted (crisp) value. If the rounded index exceeds the total number of intervals, the forecast is assigned the midpoint of the last fuzzy set. Conversely, if the index is less than zero, the forecast is set to the midpoint of the first fuzzy set. This process results in the generation of the defuzzified data series.

5.5 Experimental setup

The performance of Particle Swarm Optimization (PSO) is significantly influenced by its acceleration coefficients: the cognitive coefficient (C1) and the social coefficient (C2). The cognitive component (C1) encourages local search by guiding particles toward their own previously best-known positions, a process referred to as exploitation. In contrast, the social component (C2) promotes global search by steering particles toward the global best position, known as exploration. An excessively high C1 combined with a low C2 may result in excessive exploitation, potentially causing premature convergence to local optima. Conversely, a low C1 and high C2 may lead to excessive exploration, resulting in slow convergence and suboptimal solution refinement. Therefore, achieving an appropriate balance between exploitation and exploration through careful tuning of C1 and C2 is essential. In this chapter, these coefficients are dynamically adjusted during runtime using the equations provided in Section5.3.

An experiment is carried out to identify the number of iterations required for the PSO convergence on the adopted datasets. Fig. 5.3 illustrates the convergence of the PSO algorithm for the adopted datasets. PSO convergence is carried out with number of particles as 50, number of partitions for the datasets as 100, and fuzzy order as 2. Inertia weight is updated using equation 5.3.1, velocity of particles is updated using equation 5.3.4, and acceleration coefficients are updated using the equations 5.3.2 and 5.3.3 in each iteration. We can observe that PSO is able to converge in approximately 200 iterations for each country as shown in Fig. 5.3. Therefore, we set the maximum iterations for PSO as 200 in each experiment, and the values of other parameters are restricted within the range as given in Table 6.9.

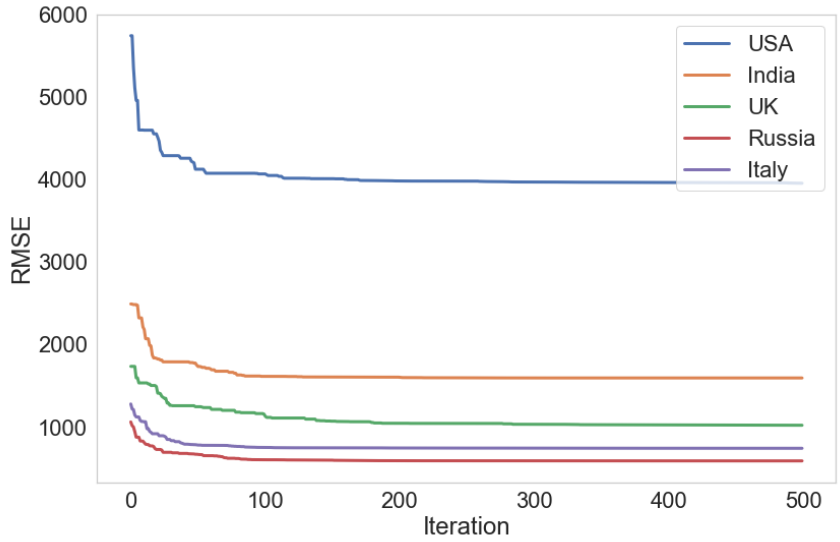


Figure 5.3: Convergence graph of PSO for the adopted datasets.

Table 5.1: PSO hyperparameters set in the experiments

PSO Parameter	Value
Number of particles	50
Maximum number of iterations	200
Inertia weight $[w_{min}, w_{max}]$	[0.4, 0.9]
C1, C2	[0.5, 2.5]

In FTS modeling of non-stationary data, the number of partitions plays a critical role in how effectively the UOD is segmented. A higher number of partitions can capture finer variations in the data, potentially enhancing forecast accuracy, but at the cost of increased model complexity. Conversely, too few partitions may oversimplify the data, leading to underfitting. Therefore, selecting an optimal number of partitions is essential to balance model complexity, computational efficiency, and forecasting performance. Likewise, the size of the fuzzy intervals significantly influences the fuzzification process. Properly sized intervals contribute to the model's robustness by effectively managing uncertainty and imprecision inherent in time series data. Thus, both the number of partitions and the size of fuzzy intervals are crucial factors in developing an effective FTS model capable of handling uncertainty. Therefore, in this study, the number of fuzzy intervals is determined using the proven average-based methodology. Combination of unequal length of intervals and fuzzy order of FTS are optimized using the PSO. Optimized number of fuzzy intervals and optimized fuzzy order for each dataset is given in Table 5.2. Further, the fuzzification of crisp time series is performed using the identified optimum parameters to generate FTS as explained in Step-4 in Section 5.4. FLRs are established using the optimum fuzzy order for each dataset.

Table 5.2: Optimized values of fuzzy intervals and fuzzy order for each dataset

COVID-19 Dataset	#Optimized Fuzzy Intervals	Optimized value of Fuzzy Order
USA	36	5
India	17	4
UK	12	3
Russia	10	3
Italy y	9	4

DL modeling using LSTM can achieve higher accuracy, flexibility, and robustness in time series forecasting. In DL modeling, the choice of optimizer, learning rate, and loss function significantly impacts the performance of the model, convergence, and ability to generalize. Adam optimizer combines the advantages of adaptive learning rate and momentum. It is widely used due to its robust

performance and efficiency. Attention mechanisms have shown great promise in time series forecasting. Instead of processing the entire input sequence as a whole, attention mechanism indicates their relevance to the current step in the output sequence.

Hyperparameters for the deep learning models are shown in Table 6.10. We have experimented with activation functions, namely, relu, tanh, linear, and sigmoid in LSTM model, out of which the linear activation function is found appropriate for the adopted time series datasets. We kept the values of the model parameters same for all the compared models. In stacked-LSTM model, one more layer of LSTM is used with 20 hidden units. In conv-LSTM model, two 1D convolutional layers are used. First layer with 60 filters, 4 kernels, single stride, and relu activation function. Second layer with 60 filters, 4 kernels, single stride, and tanh activation function. In attention-LSTM and attention-bi-LSTM model, Bahdanau [18] attention mechanism has been used.

Table 5.3: Parameters set for LSTM and its variants

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Loss function	Huber
Evaluation metrics	MSE
Hidden units	50
Dense units	1
Epochs	1000

The proposed method is evaluated against two statistical models, ETS [232] and ARIMA [85], two machine learning models, ANN [157] and LSTM [6], and a fuzzy-based hybrid model, Panigrahi-FTSF-LSTM [167]. The ARIMA and ETS models are implemented using the open-source Statsmodels Python library, while the ANN and LSTM models are built using the Keras library in TensorFlow. The Panigrahi-FTSF-LSTM model is implemented in Python 3.9.

The ETS model comprises three key components: 1) trend (T), 2) seasonal (S), and 3) an error term (E). These components influence how the model handles errors, trends, and time. The error term determines how the nearest prior periods are modeled, with statistical information criteria used to select the best fit among auto, additive, or multiplicative error specifications. The trend term defines how trends are addressed, allowing options for additive, multiplicative, or no-trend corrections. When the damped_trend option is enabled, it reduces the influence of recent trends. For daily data with a monthly cycle, the seasonal_periods parameter is set to 12.

In this analysis, extensive experiments are conducted with the ETS model to determine its optimal parameters for various datasets. An example of the model's

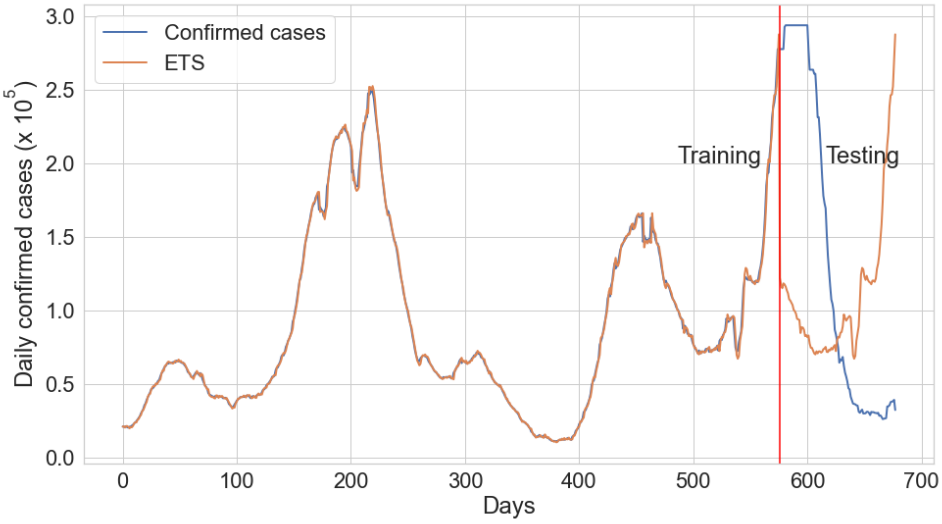


Figure 5.4: ETS(AAdA) model forecasting results on COVID-19 daily infections in the USA.

training and testing with optimal parameters for the USA time series is shown in Fig. 5.4, where the training and testing time series are separated by a vertical red line. The identified optimal parameters are specified as: ETSModel(error='add', trend='add', seasonal='add', damped_trend=True, seasonal_periods=12). This configuration is denoted as ETS(AAdA), where all components are additive, and recent trend is dampened. The ETS model with these settings is compared against the proposed methodology across different datasets and evaluation metrics in the next section.

The ARIMA(p, d, q) model integrates three components: auto-regressive (AR), differencing (d), and moving average (MA). Differencing is applied to remove trends or seasonality, ensuring the time series becomes stationary. In this analysis, Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) graphs are used to determine the AR parameter (p) and MA parameter (q) of the model. Based on these analyses, the optimal ARIMA order for the COVID-19 time series of the USA is identified as (2,1,0). The training and testing results of this configuration are shown in Fig. 5.5. Optimal ARIMA orders are similarly determined for other datasets and compared against the proposed methodology using various metrics, as discussed in the next section.

The ANN and LSTM models are evaluated using the parameters listed in Table 6.10. Additionally, various activation functions, including ReLU, tanh, linear, and sigmoid, have been tested. Among these, the linear activation function is found to be the most suitable for the adopted time series datasets.

5.5.1 Experimental Results

This study conducts a comprehensive performance comparison between the proposed hybrid forecasting framework and five contemporary state-of-the-art benchmark models. The training set is used for model training, the validation set

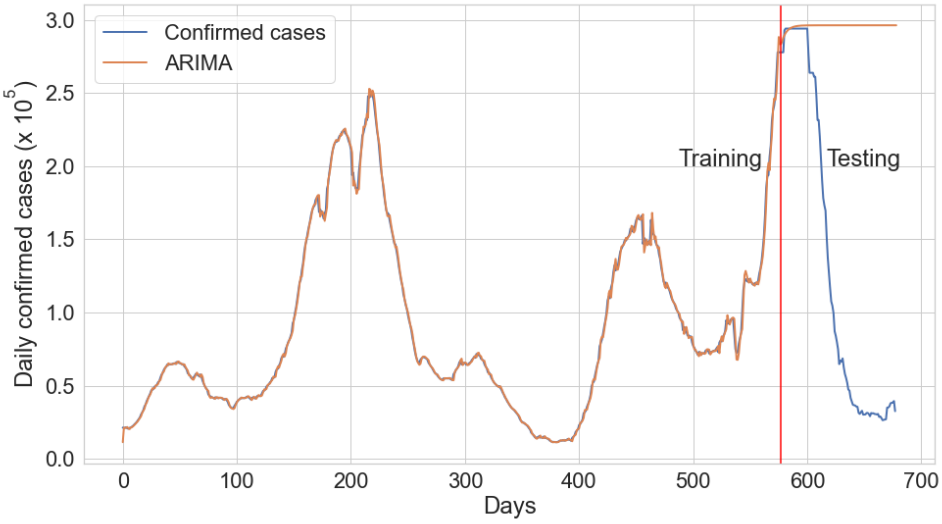


Figure 5.5: ARIMA(2, 1, 0) model forecasting results on COVID-19 daily infections in the USA.

for parameter tuning, and the testing set for forecasting. The primary focus is to address the challenge of non-stationarity in time series data. To evaluate adaptability to non-stationary dynamics, the models are first tested without applying any preprocessing to the time series data, with the results presented in Section 5.5.1. Since some forecasting models perform effectively only on stationary time series, preprocessing techniques are applied to standardize the datasets and provide a consistent baseline for all models. Following this, additional experiments are conducted on the preprocessed datasets, and the corresponding results are presented in Section 5.5.1. All models are evaluated on testing samples using the performance metrics MAE, RMSE, RRSE, sMAPE, and MASE.

Models evaluation without data preprocessing

This section presents the forecasting results for COVID-19 confirmed cases in the USA without applying preprocessing to the time series data. The performance outcomes for the testing samples are summarized in Table 5.4. The reported values represent the mean and standard deviation over three runs for each model across all performance metrics. Statistical models show no variation in standard deviation across multiple runs with fixed hyperparameters, whereas machine learning models dynamically learn patterns and adjust hyperparameters during runtime, resulting in noticeable variations in performance. The performance ranking, from lowest to highest, follows the order of statistical models, machine learning models, and hybrid fuzzy models. Notably, machine learning models demonstrated performance comparable to the proposed hybrid models. Among all the models, the hybrid approach combining FTS, PSO, and attention-Bi-LSTM delivered the best overall performance.

Table 5.4: Forecasting accuracy (mean + standard deviation) of the models without preprocessing of the USA dataset (best values are shown in bold).

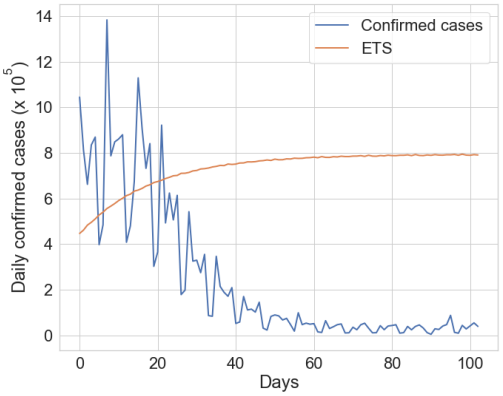
Model	USA				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(MAdA)	510612 ± 0	474277 ± 0	1.464 ± 0	123.34 ± 0	132.88 ± 0
ARIMA(4,1,0)	322914 ± 0	389853 ± 0	1.238 ± 0	111.08 ± 0	165.391 ± 0
ANN	226450 ± 424	356206 ± 410	1.152 ± 0.18	108.66 ± 8	1.893 ± 0.64
LSTM	213629 ± 412	351019 ± 384	1.135 ± 0.15	99.62 ± 8	1.646 ± 0.52
Panigrahi-FTSF-LSTM	212642 ± 374	349544 ± 292	1.153 ± 0.16	102.74 ± 6	1.627 ± 0.34
FTS+PSO +LSTM	209206 ± 357	343899 ± 284	1.119 ± 0.14	99.83 ± 7	1.625 ± 0.28
FTS+PSO +stacked-LSTM	206641 ± 344	342092 ± 261	1.108 ± 0.13	97.26 ± 6	1.615 ± 0.22
FTS+PSO +Bi-LSTM	205772 ± 341	341209 ± 244	1.115 ± 0.13	96.14 ± 5	1.591 ± 0.19
FTS+PSO +Conv-LSTM	202588 ± 336	341972 ± 228	1.199 ± 0.12	95.95 ± 5	1.593 ± 0.21
FTS+PSO +Attention-LSTM	202673 ± 322	341483 ± 214	1.114 ± 0.11	95.54 ± 4	1.589 ± 0.18
FTS+PSO +Attention-Bi-LSTM	**201567 ± 302**	**340929 ± 144**	**1.104 ± 0.11**	**94.34 ± 4**	**1.559 ± 0.16**

The prediction results are visualized in Fig. 5.6. The models are trained on the training samples, and the testing results are generated using the trained hyperparameters. The forecasting results for the ETS(MAdA) model are shown in Fig. 5.6a. The ETS(MAdA) model performed best with a multiplicative error term, additive seasonal and trend components, and dampened recent trends.

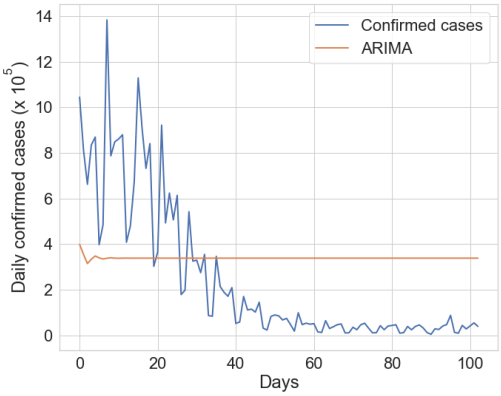
Fig. 5.6b illustrates the forecasting results of the ARIMA model, which achieved its best fit with an order of (4,1,0). Testing results are generated using ARIMA(4,1,0), showing effectiveness of the model in handling non-stationary data.

The ANN and LSTM machine learning models are trained on the training samples and fine-tuned using the validation set to capture time series patterns. These models have demonstrated significant improvements in forecasting accuracy, as shown in Fig. 5.6c and Fig. 5.6d. Notably, LSTM has outperformed both ANN and the statistical models on the USA dataset.

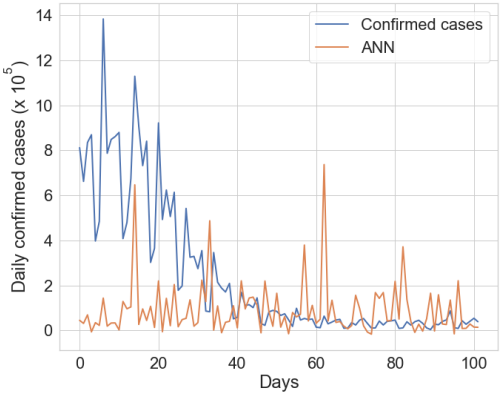
The forecasting results of the Panigrahi-FTSF-LSTM model are displayed in Fig. 5.6e, while the results of the proposed hybrid models, integrating FTS, PSO,



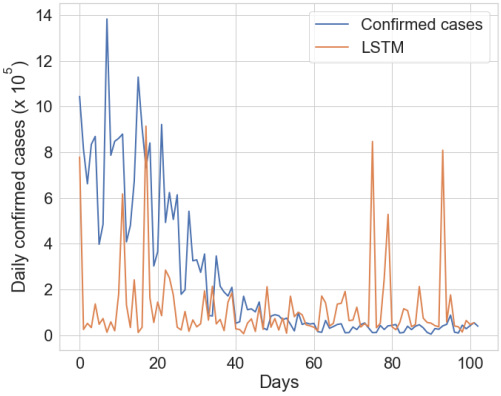
(a) ETS(AAdD) forecast testing results.



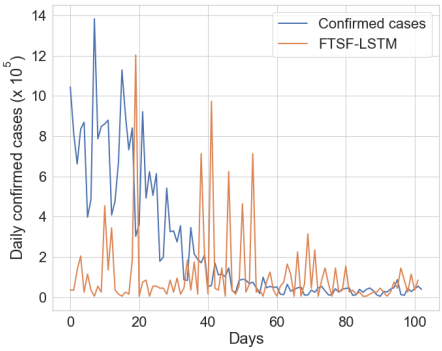
(b) ARIMA(2,1,0) forecast testing results.



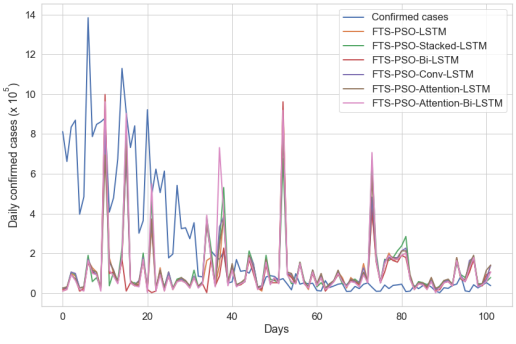
(c) ANN forecast testing results.



(d) LSTM forecast testing results.



(e) Panigrahi-FTSF+LSTM model forecast testing results.



(f) Proposed hybrid variants forecast testing results.

Figure 5.6: Testing results of forecasts using the proposed and compared models for COVID-19 confirmed cases without preprocessing time series of the USA.

and LSTM variants, are presented in Fig. 5.6f. The fuzzy-based machine learning models demonstrated superior accuracy compared to traditional models.

Models evaluation on preprocessed data

This section assesses the forecasting performance of all models using preprocessed time series datasets to establish a consistent baseline for comparison.

Evaluation on the USA dataset: The experiment detailed in Section 5.5.1 is repeated here using the preprocessed dataset for the USA. The improvement in model performance after preprocessing is illustrated in Fig. 5.7 and summarized in Table 5.5.

The forecasting results for all models on the testing samples of the USA dataset are shown in Fig. 5.7. It is evident that statistical models struggle to handle the multiple waves of COVID-19. In contrast, machine learning models exhibit significant improvement, further enhanced by the integration of fuzzy logic and optimization techniques.

Table 5.5 compares performance of the models using various metrics on the preprocessed USA dataset. The results indicate that all models benefit from data preprocessing, leading to improved forecasting accuracy. However, statistical models, such as ETS(AAdA) and ARIMA(2,1,0), despite being well-fitted on the training data, failed to handle dynamic changes in the time series. This limitation causes them to converge toward the mean for the testing samples, as reflected in Fig. 5.7.

The hybrid models, particularly those combining LSTM (augmented with attention and convolutional layers), fuzzy logic, and PSO, demonstrate significant performance improvements. While the impact of data preprocessing on hybrid models is less pronounced, it still contributes to better results, underscoring the importance of data preprocessing in enhancing model performance.

Table 5.5: Forecasting accuracy (mean + standard deviation) of models on preprocessed dataset of the USA (best values are shown in bold).

Model	USA				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	123115 ± 0	140967 ± 0	1.261 ± 0	94.74 ± 0	30.24 ± 0
ARIMA(2,1,0)	157951 ± 0	193898 ± 0	1.735 ± 0	90.55 ± 0	140.63 ± 0
ANN	113179 ± 364	141847 ± 396	1.273 ± 0.19	90.37 ± 8	1.52 ± 0.28
LSTM	128375 ± 328	138425 ± 384	1.252 ± 0.17	88.12 ± 6	1.62 ± 0.29
Panigrahi-FTSF-LSTM	115439 ± 284	134054 ± 289	1.263 ± 0.15	86.13 ± 5	1.48 ± 0.31
FTS+PSO +LSTM	108509 ± 262	138872 ± 274	1.246 ± 0.14	86.51 ± 4	1.46 ± 0.24

Continued on next page

Table 5.5 – continued from previous page

Model	USA				
	MAE	RMSE	RRSE	sMAPE	MASE
FTS+PSO +stacked-LSTM	107632 ± 253	137718 ± 259	1.251 ± 0.14	86.46 ± 4	1.44 ± 0.22
FTS+PSO +Bi-LSTM	107610 ± 241	137632 ± 256	1.245 ± 0.12	85.63 ± 5	1.41 ± 0.21
FTS+PSO +Conv-LSTM	107344 ± 233	137428 ± 244	1.247 ± 0.13	87.56 ± 4	1.45 ± 0.19
FTS+PSO +Attention-LSTM	107198 ± 224	137205 ± 237	1.242 ± 0.11	85.18 ± 4	1.42 ± 0.18
FTS+PSO +Attention-Bi-LSTM	**106230 ± 218**	**136938 ± 227**	**1.147 ± 0.11**	**84.41 ± 3**	**1.41 ± 0.16**

Evaluation on India dataset:The superiority of a model cannot be judged solely based on its performance on one or two datasets. Hence, five diverse datasets are used to evaluate all the models. The forecasting performance of the models on the test samples of the India dataset is given in Table 5.6. The results demonstrate that the hybrid model combining FTS, PSO, and attention Bi-LSTM outperformed all other models across all metrics. It is important to use multiple performance metrics, as excelling in one metric does not necessarily indicate superior performance across others. Each metric provides a unique perspective, contributing to a more comprehensive evaluation of model performance.

Table 5.6: Forecasting accuracy (mean + standard deviation) of three runs of the models on the India dataset (best values are shown in bold).

Model	India				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(MAdM)	60731 ± 0	65662 ± 0	1.249 ± 0	113.66 ± 0	42.41 ± 0
ARIMA(6,1,0)	61255 ± 0	69522 ± 0	1.217 ± 0	122.25 ± 0	115.33 ± 0
ANN	59067 ± 367	65644 ± 423	1.215 ± 0.41	104.05 ± 4	1.48 ± 0.14
LSTM	58280 ± 349	64319 ± 439	1.222 ± 0.36	106.92 ± 5	1.34 ± 0.15
Panigrahi FTSF-LSTM	57388 ± 369	65547 ± 419	1.234 ± 0.35	109.61 ± 7	1.32 ± 0.18
FTS+PSO +LSTM	55282 ± 356	63841 ± 398	1.182 ± 0.33	98.22 ± 6	1.29 ± 0.17
FTS+PSO +stacked-LSTM	54211 ± 344	63281 ± 381	1.164 ± 0.32	96.66 ± 6	1.28 ± 0.16

Continued on next page

Table 5.6 – continued from previous page					
Model	India				
	MAE	RMSE	RRSE	sMAPE	MASE
FTS+PSO +Bi-LSTM	54157 ± 339	63014 ± 377	1.159 ± 0.31	94.52 ± 5	1.26 ± 0.15
FTS+PSO +Conv-LSTM	53456 ± 331	62607 ± 367	**1.129 ± 0.18**	94.56 ± 4	1.24 ± 0.14
FTS+PSO +Attention-LSTM	53232 ± 328	62506 ± 359	1.139 ± 0.23	95.02 ± 5	1.23 ± 0.15
FTS+PSO +Attention-Bi-LSTM	**53044 ± 318**	**62597 ± 356**	1.135 ± 0.17	**94.13 ± 4**	**1.22 ± 0.16**

Evaluation on UK dataset: The forecasting performance of the models on the COVID-19 confirmed cases from the UK is presented in Table 5.7. The multiple waves in dataset of the UK makes it an ideal example of a non-stationary time series dataset. As shown in the table, hybrid models consistently outperform traditional models. Notably, the hybrid model combining FTS, PSO, and Attention Bi-LSTM has demonstrated strong performance across all metrics except sMAPE. The multiple waves present in the test dataset have negatively impacted the performance of traditional models, highlighting their limitations to deal with dynamics in time series data. This indicates that the proposed hybrid model is better equipped to handle complex, multi-wave time series data effectively.

Table 5.7: Forecasting accuracy (mean + standard deviation) of the models on the UK dataset (best values are shown in bold).

Model	UK				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	30953 ± 0	37424 ± 0	1.836 ± 0	90.065 ± 0	26.917 ± 0
ARIMA(4,1,30)	25802 ± 0	35344 ± 0	1.341 ± 0	92.769 ± 0	13.735 ± 0
ANN	24588 ± 297	32149 ± 368	2.623 ± 0.14	82.20 ± 13	1.53 ± 0.12
LSTM	24374 ± 285	31263 ± 345	2.417 ± 0.15	81.915 ± 14	1.49 ± 0.15
Panigrahi FTSF-LSTM	22925 ± 277	30941 ± 329	2.484 ± 0.18	85.43 ± 11	1.51 ± 0.11
FTS+PSO +LSTM	21820 ± 256	30875 ± 364	2.496 ± 0.19	78.53 ± 12	1.37 ± 0.09
FTS+PSO +stacked-LSTM	21426 ± 242	30574 ± 391	2.481 ± 0.17	77.69 ± 11	1.34 ± 0.12

Continued on next page

Table 5.7 – continued from previous page

Model	UK				
	MAE	RMSE	RRSE	sMAPE	MASE
FTS+PSO +Bi-LSTM	21820 ± 281	30772 ± 383	2.496 ± 0.16	79.53 ± 13	1.32 ± 0.08
FTS+PSO +Conv-LSTM	20914 ± 231	29583 ± 359	2.481 ± 0.14	**74.59 ± 14**	1.29 ± 0.09
FTS+PSO +Attention-LSTM	20823 ± 218	28876 ± 365	2.496 ± 0.15	76.53 ± 13	1.26 ± 0.11
FTS+PSO +Attention-Bi-LSTM	**20584 ± 212**	**28470 ± 354**	**2.329 ± 0.16**	75.089 ± 12	**1.25 ± 0.09**

Evaluation on Russia dataset: The models are further evaluated using the time series data of COVID-19 confirmed cases in Russia. The forecasting performance results for the testing samples are presented in Table 5.8 for each model. In this evaluation, the hybrid of FTS, PSO, and attention Bi-LSTM model outperformed all others on the MAE, RRSE, and sMAPE metrics. The hybrid of FTS, PSO, and convolution LSTM model delivered the best performance on RMSE, while the hybrid of FTS, PSO, and attention LSTM model excelled on the MASE metric.

Table 5.8: Forecasting accuracy (mean + standard deviation) of the models on Russia dataset (best values are shown in bold).

Model	Russia				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	20935 ± 0	22398 ± 0	2.797 ± 0	61.45 ± 0	27.39 ± 0
ARIMA(5,1,10)	23236 ± 0	25352 ± 0	3.171 ± 0	72.79 ± 0	110.607 ± 0
ANN	20356 ± 158	24443 ± 185	1.802 ± 0.29	76.89 ± 5	1.82 ± 0.14
LSTM	20267 ± 169	23831 ± 193	1.766 ± 0.24	75.45 ± 4	1.67 ± 0.11
Panigrahi FTSF-LSTM	19863 ± 149	22634 ± 176	1.716 ± 0.21	73.22 ± 4	1.56 ± 0.09
FTS+PSO +LSTM	19481 ± 147	21514 ± 173	1.648 ± 0.19	62.17 ± 3	1.49 ± 0.12
FTS+PSO +stacked-LSTM	19339 ± 159	21246 ± 169	1.655 ± 0.18	61.51 ± 5	1.47 ± 0.11
FTS+PSO +Bi-LSTM	19457 ± 128	21209 ± 171	1.648 ± 0.15	62.11 ± 3	1.46 ± 0.09

Continued on next page

Table 5.8 – continued from previous page					
Model	Russia				
	MAE	RMSE	RRSE	sMAPE	MASE
FTS+PSO +Conv-LSTM	19288 ± 123	**21121 ± 154**	1.643 ± 0.16	61.15 ± 4	1.44 ± 0.13
FTS+PSO +Attention-LSTM	19293 ± 124	21325 ± 165	1.631 ± 0.15	60.18 ± 3	**1.41 ± 0.12**
FTS+PSO +Attention-Bi-LSTM	**19085 ± 118**	21213 ± 163	**1.629 ± 0.15**	**60.16 ± 3**	1.43 ± 0.12

Evaluation on Italy dataset: The models are further evaluated using the Italy dataset to add another dimension to the comparative analysis. After being trained on the training samples and fine-tuned on the validation samples, forecasting is conducted on the testing samples. The forecasting results for the testing samples are shown in Table 5.9. The results indicate that the hybrid of FTS, PSO, and attention Bi-LSTM model outperformed all other models across all metrics.

Table 5.9: Forecasting accuracy (mean + standard deviation) of the models on the Italy dataset (best values are shown in bold).

Model	Italy				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS(AAdA)	36716 ± 0	38917 ± 0	9.436 ± 0	130.83 ± 0	75.06 ± 0
ARIMA(1, 1, 0)	35073 ± 0	37805 ± 0	7.832 ± 0	97.21 ± 0	52.767 ± 0
ANN	34352 ± 348	37518 ± 397	6.67 ± 1.09	84.62 ± 6	2.39 ± 0.16
LSTM	34012 ± 334	37114 ± 389	6.51 ± 1.12	89.13 ± 8	2.27 ± 0.15
Panigrahi FTSF-LSTM	33527 ± 322	35959 ± 366	6.54 ± 1.18	83.45 ± 7	2.36 ± 0.13
FTS+PSO +LSTM	33462 ± 321	35183 ± 362	5.73 ± 1.14	73.32 ± 6	2.23 ± 0.17
FTS+PSO +stacked-LSTM	33536 ± 316	34643 ± 367	5.53 ± 1.12	72.89 ± 5	2.17 ± 0.14
FTS+PSO +Bi-LSTM	33387 ± 314	34399 ± 363	5.05 ± 1.11	74.18 ± 5	2.14 ± 0.12
FTS+PSO +Conv-LSTM	32529 ± 316	33826 ± 345	5.16 ± 1.12	66.61 ± 5	2.12 ± 0.11

Continued on next page

Table 5.9 – continued from previous page					
Model	Italy				
	MAE	RMSE	RRSE	sMAPE	MASE
FTS+PSO +Attention-LSTM	32468 ± 311	33425 ± 316	4.96 ± 1.11	62.28 ± 4	1.11 ± 0.11
FTS+PSO +Attention-Bi-LSTM	**32113 ± 308**	**33192 ± 314**	**4.25 ± 1.02**	**61.15 ± 4**	**1.09 ± 0.08**

This study carried out extensive experiments to evaluate the models across multiple datasets and performance metrics. The forecasting capabilities were tested on both raw and preprocessed data, emphasizing the importance of preprocessing. The results indicated that the models demonstrate varied behavior depending on the metric used, with performance differences generally being minimal. Consequently, a statistical analysis of the models' performance is provided in Section 5.5.2 to determine if any model is significantly superior to the others.

Analysis of computational complexity

A hybrid model improves performance but at the cost of increased computational complexity. Therefore, it is important to analyze the computation time of all models during training. The computation time for all models on the USA dataset is provided in Table 5.10, with the measured time shown in seconds. Hybrid models accumulate the training time of the integrated methods, resulting in longer computation times compared to the baseline models, which are faster. However, in an era where computing resources are readily available, model accuracy takes precedence. As a result, hybrid models are widely adopted despite their higher computational cost.

Table 5.10: Training computation time (sec) of the models for USA dataset

Model	Training time (sec)	PSO computation time (sec)	Total computation time (sec)	Relative Rank
ETS(AAdA)	5	-	5	2
ARIMA(2,1,0)	1	-	1	1
ANN	1226	-	1226	3
LSTM	2382	-	2382	5
Panigrahi-FTSF-LSTM	2361	-	2361	4
FTS+PSO +LSTM	2121	3141	5262	6
FTS+PSO +stacked-LSTM	3192	3141	6333	11

Continued on next page

Table 5.10 – continued from previous page

Model	Training time (sec)	PSO computation time (sec)	Total computation time (sec)	Relative Rank
FTS+PSO +Bi-LSTM	2808	3141	5949	8
FTS+PSO +Conv-LSTM	2920	3141	6061	10
FTS+PSO +Attention-LSTM	2178	3141	5319	7
FTS+PSO +Attention-Bi-LSTM	2902	3141	6043	9

5.5.2 Statistical significance test

Comparing the performance of forecasting models on a single dataset is relatively straightforward, but it is essential to assess whether performance of a model remains consistently significant across different scenarios. To address this, this section presents a statistical significance test of the proposed model in comparison with other models. The evaluation is based on COVID-19 infections data from five highly affected countries namely, the USA, India, the UK, Russia, and Italy. The relative and average rankings of all models based on the RMSE metric across these five datasets are presented in Table 5.11.

Table 5.11: Relative ranking of the forecasting models using RMSE performance metric on the adopted datasets)

Model	Relative Rank on RMSE Performance					
	USA	India	UK	Russia	Italy	Avg. Rank (R)
ETS	10	10	11	10	11	10.4
ARIMA	11	11	10	11	10	10.6
ANN	7	9	9	9	9	8.6
LSTM	8	7	8	8	8	7.8
Panigrahi-FTFS-LSTM	9	8	7	7	7	7.6
FTS+PSO +LSTM	6	6	5	6	6	5.8
FTS+PSO +stacked-LSTM	5	5	6	5	5	5.2
FTS+PSO +Bi-LSTM	4	4	4	4	4	4
FTS+PSO +Conv-LSTM	3	3	3	1	3	2.6

Continued on next page

Table 5.11 – continued from previous page

Model	Relative Rank on RMSE measure					
	USA	India	UK	Russia	Italy	Avg. Rank (R)
FTS+PSO +Attention-LSTM	2	2	2	3	2	2.2
FTS+PSO +Attention-Bi-LSTM	1	1	1	2	1	1.2
$\chi_F^2 = 18.36, F_F = 13.89$, critical value at $\alpha = 0.05$ is 2.7642, so, H_0 rejected with 95% confidence						

Table 5.11 presents the relative performance ranks of the models based on the RMSE metric across all five datasets. These ranks are derived from the RMSE performance on the different datasets discussed in the previous section. The table also includes the average ranking for each model across the five datasets, with the best-performing model receiving a rank of 1. To determine whether one model is truly better than the others, statistical significance tests can be used to support or reject such a claim (also known as a hypothesis) based on sample data. A claim is considered valid if it is statistically significant at a predetermined significance level. The threshold for this significance is set by the p-value, which must be chosen before the test is conducted. Traditionally, a p-value of 1% or 5% is used, and in this analysis, the p-value is set at 5%.

In the statistical analysis, the Nemenyi test [158] is employed to assess the significance of the performance differences between the models. The Nemenyi post-hoc test is used to identify which specific models have different means. This test conducts pairwise comparisons of model performance. The hypothesis for the Nemenyi test is outlined below.

- H_0: There is no difference between any two models.
- H_a: At least one pair of models is different.

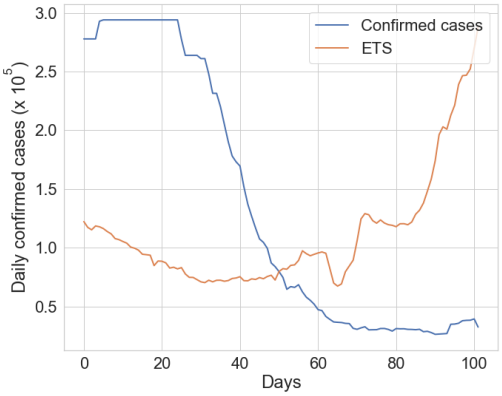
The formula of critical difference (CD) for Nemenyi test is given in equation 5.5.1.

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{5.5.1}$$

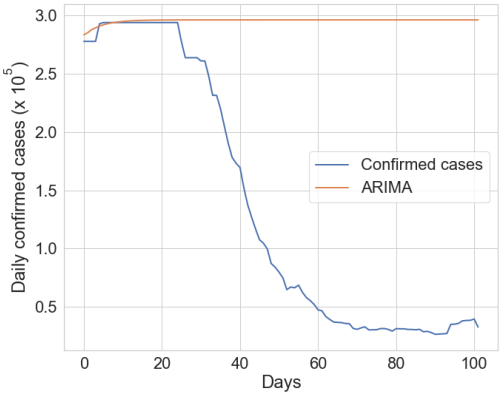
where q_α is the critical difference value from the Studentized range distribution at the significance level α divided by $\sqrt{2}$, and k is the number of groups. In our case $k = 11$, hence $q_{\alpha=0.05} = 1.782$. The value is taken from the online link [2]. Critical difference (CD) is calculated using the values as given below.

$CD_{\alpha=0.05} = 1.782 \times \sqrt{\frac{11\times12}{6\times5}} = 3.732$

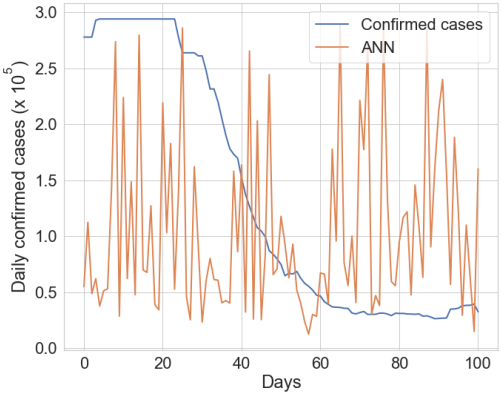
[2] https://statisticsbyjim.com/hypothesis-testing/t-distribution-table/



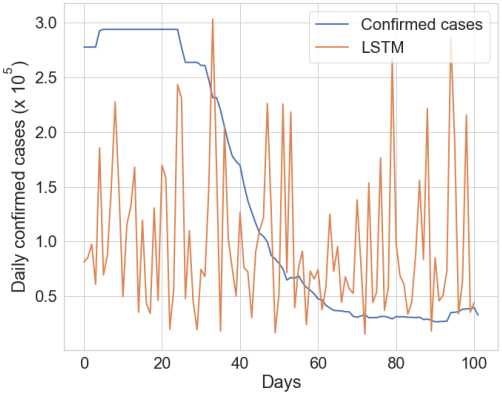
(a) ETS(AAdA) forecast testing results on preprocessed data of the USA.



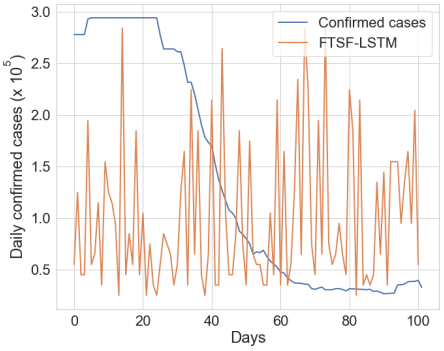
(b) ARIMA(2,1,0) forecast testing results on preprocessed data of the USA.



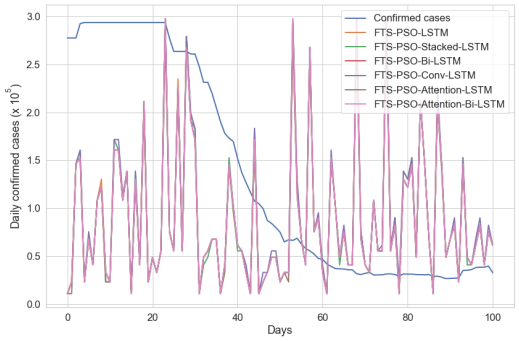
(c) ANN forecast testing results on preprocessed data of the USA.



(d) LSTM forecast testing results on preprocessed data of the USA.



(e) Panigrahi-FTST+LSTM forecast testing results on preprocessed data of the USA.



(f) Proposed hybrid variants forecast testing results on preprocessed data of the USA.

Figure 5.7: Comparative forecasting results of the models for preprocessed time series of COVID-19 confirmed cases of the USA.

The difference between the highest and lowest average ranks of the models from Table 5.11 is $(10.6 - 1.2) = 9.4$, which is greater than the critical difference (CD) of 3.7322. Therefore, the null hypothesis H_0 of the Nemenyi test is rejected with 95% confidence. In other words, there is sufficient evidence to conclude that at least one pair of models differs from the others.

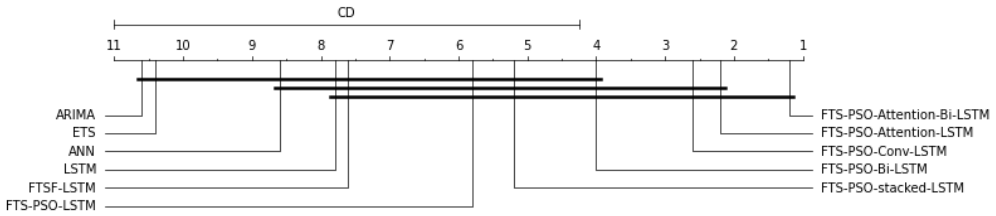


Figure 5.8: Critical difference plot showing the results of Nemenyi test for the compared models after forecasting on test data.

The pairwise comparison of all models using the critical difference is graphically analyzed in Fig. 5.8. The average ranks of the models are plotted along the top line of the figure, with the best ranks appearing on the right side of the axis. Models that are not significantly different at $\alpha = 0.05$ are connected by a bold solid line. This figure not only highlights the statistical differences between the models but also provides insight into the most effective models. It is evident that hybrid of FTS, PSO, and attention Bi-LSTM significantly outperforms the ETS, ARIMA, and ANN models.

5.6 Epidemiological model based study for mutant affected population

Numerous research studies have evaluated state-of-the-art epidemiological compartmental models using data from the COVID-19 pandemic [142, 100]. These models have been further enhanced by incorporating additional compartments and leveraging various machine learning and optimization techniques to improve disease modeling [55, 25]. A standard epidemiological model may not fully capture the complexities of a pandemic like COVID-19. Predicting the progression of a pandemic is crucial for governments to plan strategies to control disease spread. Therefore, a hybrid epidemiological model may be an effective alternative. Hybrid models combine the strengths of multiple models or algorithms to achieve better performance.

Epidemiological compartmental models are essential tools for analyzing various aspects of epidemic spread and assessing the effectiveness of public health interventions. Therefore, a hybrid epidemiological compartmental model is proposed integrating SIRD, PSO, and stacked-LSTM approaches. The proposed hybrid model is compared with independent stacked-LSTM and SIRD models. This chapter provides an overview of the SIRD and LSTM models, along with a detailed explanation of the proposed methodology and the experimental results.

5.7 SIRD Compartmental Model

The Susceptible-Infected-Recovered-Deceased (SIRD) epidemiological model is a widely used framework for analyzing the spread of infectious diseases. It helps in understanding the dynamics of disease transmission within a population and can predict the future trajectory of an epidemic under various scenarios. For this reason, the SIRD model has been chosen to model the spread of the COVID-19 pandemic. The model categorizes the population into four compartments, namely, susceptible (S), infected (I), recovered (R), and deceased (D). The state transitions between these compartments is illustrated in Figure 5.9.

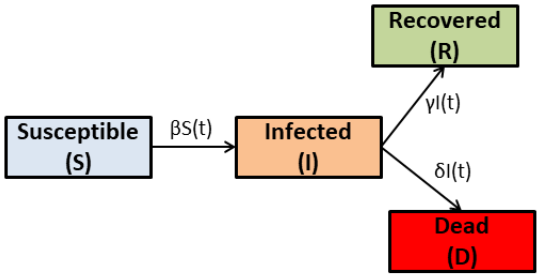


Figure 5.9: State diagram of SIRD model.

Transitions between the model compartments are governed by ordinary differential equations (ODEs). ?? for the SIRD epidemiological model are given below.

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N} \tag{5.7.1}$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma I(t) - \delta I(t) \tag{5.7.2}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{5.7.3}$$

$$\frac{dD(t)}{dt} = \delta I(t) \tag{5.7.4}$$

$$N = S(t) + I(t) + R(t) + D(t) \tag{5.7.5}$$

Here, $S(t), I(t), R(t)$, and $D(t)$ represent the numbers of susceptible, infected, recovered, and deceased individuals at time t, respectively. The parameters β, γ, δ denote the infection rate, recovery rate, and fatality rate. The model is constructed as a positive system, therefore all variables remain positive for $t > 0$ if initialized with non-negative values at $t = 0$. Human births and natural deaths are not the considered in the model [132]. The sum of all compartments remains constant, totaling equal to the population N. The SIRD model is a foundation for more complex models, making it an essential tool for understanding infectious diseases. In this chapter, the SIRD model is used to develop a hybrid forecasting model.

5.8 Proposed hybrid methodology

Hybrid modeling combines multiple modeling approaches such as epidemiological models, machine learning and optimization techniques to leverage their respective strengths and achieve improved performance in solving complex problems. In the context of epidemiology, hybrid modeling is used to improve disease spread prediction results. For example, a hybrid epidemiological model integrated with machine learning can take advantages of the predictive power and adaptability of the ML approaches. The hybrid concept is based on combination of three approaches, namely, epidemiological model, machine learning, and computational intelligence technique.

In this chapter, PSO algorithm is adopted because it is capable of providing a solution without entrapping into a local minima. In PSO, each particle navigates through the search space of the optimization problem to locate the optimal solution. Where, each particle within the swarm represents a potential final optimized solution. In the methodology, the PSO algorithm is utilized to determine the time-varying optimal values of the SIRD model parameters by fitting them to time series data of the COVID-19 cases.

The LSTM deep learning network is used to deal with the long-term dependencies in the time series data. The LSTM neural network consists of a memory cell, and three gates *viz.* an input gate, a forget gate, and an output gate. Information in the LSTM network is regulated by these gates and the memory cell. The stacked-LSTM model consists of multiple LSTM layers arranged in a stacked configuration, allowing the input data to be processed sequentially through each layer [64, 63]. In this chapter, the stacked-LSTM is utilized to perform forecasting of the SIRD model parameters obtained after optimization using PSO algorithm.

In this chapter, a hybrid epidemiological compartmental model is proposed integrating SIRD, PSO, and stacked-LSTM to improve the COVID-19 forecasting results [116]. Initial estimates of the SIRD model parameters are insufficient to reflect real-world scenarios, such as the dynamic nature of COVID-19 time series data. To address this, the model parameters are updated weekly to capture variations caused by new waves or policy changes. The model parameters are optimized using PSO algorithm. The optimized parameters are obtained for the training dataset. For the testing dataset, stacked-LSTM is used to identify the model parameters for future predictions. Stacked-LSTM is trained on the optimized parameters of SIRD for training samples, and parameter forecasting is performed for the testing samples. These predicted parameters are then input into the SIRD model to generate forecasts for the time series data.

In the modeling process, SIRD model parameters are optimized weekly using PSO by fitted on real values of the COVID-19 cases. The model compartments $S(0), I(0), R(0)$, and $D(0)$ are initialized using real values at $t = 0$ which is start date in the adopted time series dataset. Model states $S(t), I(t), R(t)$, and $D(t)$ define values at time at time t, where, $t > 0$. The weekly optimized SIRD model parameters are fed into the stacked-LSTM neural network to train and forecast the parameters for upcoming four weeks (28 days). The forecasted parameters are used in the SIRD model to get the evolution trend of the cases for next 28 days.

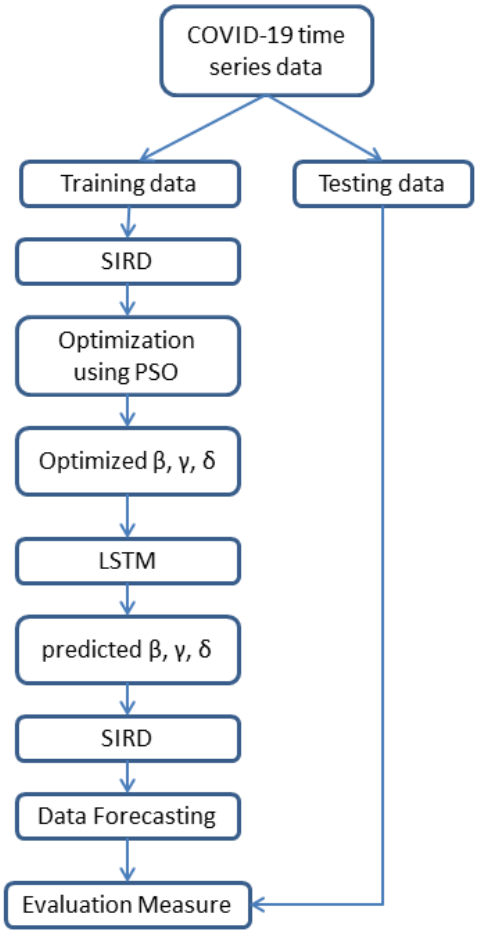


Figure 5.10: Proposed hybrid model framework for epidemiological modeling and forecasting.

The proposed hybrid epidemiological model is illustrated in Figure 5.10. The modeling framework is based on a set of sequential steps. The steps are explained below.

1. Data preprocessing: the COVID-19 time series data exhibits significant fluctuations. Weekly averaging is applied to smooth the data and reduce the fluctuations.

2. Train-test splitting: the preprocessed data is divided into training and testing samples. The last 28 days (four weeks) samples are kept for testing. In case of a highly dynamic pandemic such as COVID-19, long-term forecasting may not be realistic.

3. SIRD modeling: SIRD compartmental model is used for the COVID-19 pandemic evolution. The model is explained in section 5.7.

4. Model parameter optimization: The SIRD model parameters are estimated periodically. PSO algorithm is used to optimize the SIRD model parameters (β, γ, δ). Initially, the values of infection rate (β), recovery rate (γ), and fatality rate (δ) are randomly generated between 0 and 1. These values are updated iteratively using the PSO algorithm by fitted on real values of the COVID-19 cases using the SIRD model. Data fitting of $S(t), I(t), R(t), D(t)$ values are evaluated with respect to Mean Squared Error (MSE). Optimum values of the parameters (β, γ, δ) are returned by the PSO. The periodicity for the parameters estimation is set as weekly in the experimented model.

5. Model parameter forecasting: The stacked-LSTM is used to forecast the value of the SIRD parameters. The weekly optimized parameters (β, γ, δ) are used to train the LSTM model. Further, the LSTM is used to perform forecasting for four weeks (28 days).

6. Pandemic evolution: In this step, the SIRD model is used to determine the final forecasted values. The parameters forecasted by the stacked-LSTM are input into the SIRD model. Four weeks forecasted values of β, γ, and δ are applied to the SIRD model to simulate the progression of the COVID-19 pandemic over the next 28 days.

The proposed hybrid time-varying epidemiological model integrating stacked-LSTM and PSO optimization is evaluated using COVID-19 datasets. Modeling datasets and experimental results are presented in the following sections.

5.8.1 Modeling Datasets

COVID-19 cases time series data from three countries, namely, the USA, UK, and India are used to evaluate the proposed model. The data is publicly available at github repository [93]. Time series of the datasets is selected involving recent two major waves of the COVID-19 spread. The detail of the adopted time series for each country is given in Table 5.12. Constant population is considered in the SIRD model for the USA as 330 Million, India as 1100 Million, and the UK as 60 Million.

Table 5.12: Timeline of COVID-19 time series used in the experiments

Dataset	Start-Date	End-Date
USA	July 1, 2021	April 15, 2022
India	January 1, 2021	March 30, 2022
UK	June 1, 2021	April 30, 2022

5.8.2 Experimental setup

The proposed hybrid model is implemented using Python 3.9. Its forecasting results are compared with the results obtained from standard stacked-LSTM and the SIRD-PSO hybrid models. The PSO parameters are configured as shown in Table 5.13. PSO algorithm was able to converge in around 200 iterations for the adopted datasets. So, maximum number of iterations are set as 200 in each experiments for the optimization. Root Mean Square Error (RMSE) performance metric is used to compare the experimented models.

Table 5.13: PSO parameters configured in the experiments

PSO Parameter	Value
Number of particles	30
Maximum number of iterations	200
Inertia weight $[w_{min}, w_{max}]$	[0.4, 0.9]
C1 = C2	2

The configured parameters for the LSTM are presented in Table 5.14. In the stacked-LSTM model, two LSTM layers are employed: the first layer with 20 hidden units and the second layer with 30 hidden units.

Table 5.14: LSTM parameters configured in the experiments

Parameter	Value
Activation Function	linear
Optimizer	Adam
Learning rate	0.005
Loss function	Huber
Evaluation metrics	MSE
Epochs	500

5.8.3 Experimental results

The proposed model is experimented for the evolution study of the COVID-19 confirmed cases from the USA, UK, and India. The experiments are carried out in two ways, 1) COVID-19 evolution is carried out using the proposed model, and 2) Forecasting performance comparison of the proposed model with stacked-LSTM and SIRD-PSO models in graphical and tabular form.

COVID-19 evolution study using the proposed hybrid model

COVID-19 evolution results using the proposed model for training and testing of all the three countries are shown in Figure 5.11. A vertical line is used to separate the training and forecasting results. Evolution results of the proposed model for infected cases of the COVID-19 from the USA, India, and the UK are shown in Figures 5.11a, 5.11b, and 5.11c, respectively.

Figure 5.11 illustrates that the proposed model successfully fits the training samples of all the selected countries. The tuned parameters of the model are utilized for forecasting, and the corresponding results are also displayed in the figure. A red vertical line separates the training and forecasting outcomes. In the next section, the forecasting performance of the proposed model is compared with state-of-the-art approaches using the testing samples of the datasets.

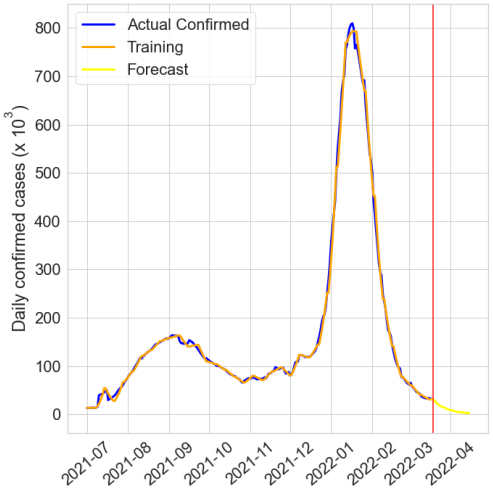
Comparative analysis of the models

A comparative analysis of the proposed model with other state-of-the-art approaches is presented in Fig. 5.12 using the testing samples from each dataset. The figure displays the forecasting results for daily cumulative COVID-19 infection cases over the testing period. Forecasting is limited to the next four weeks (28 days) to maintain realistic predictions, particularly in the context of dynamic pandemics like COVID-19.

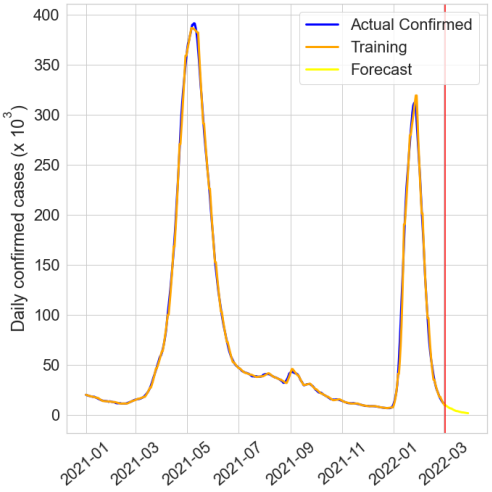
Figures 6.9a, 6.9b, and 6.9c show forecasting results for the USA using the proposed model and compared models. Figures 6.9d, 6.9e, and 6.9f show forecasting results for India using the three models. Figures 5.12g, 5.12h, and 5.12i show forecasting results for the UK using the three models. From these figures, it is evident that standalone LSTM does not perform effectively on the testing samples. However, incorporating the PSO optimization technique with LSTM significantly improves forecasting results across all datasets. The proposed hybrid model, which integrates time-varying SIRD, PSO, and LSTM, outperforms the other models on the testing samples of the selected datasets.

Table 5.15: Forecasting accuracy of the proposed model and compared models for the COVID-19 infected cases from three countries

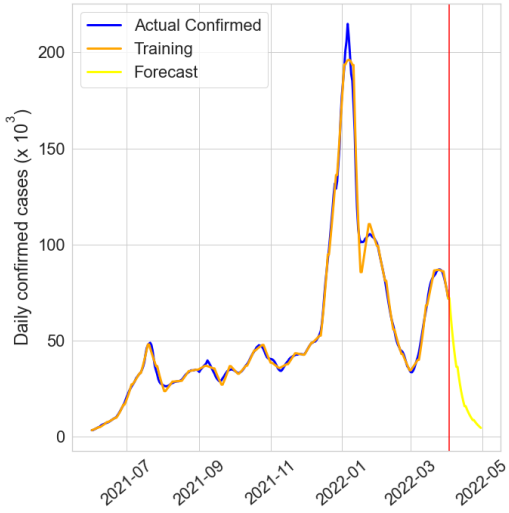
Model	RMSE		
	USA	India	UK
Stacked-LSTM	27836	3814	37479
SIRD + PSO	25816	2399	26582
Proposed model (SIRD + PSO + LSTM)	**23528**	**1887**	**24629**



(a) Training and forecasting results of the proposed hybrid model for the USA

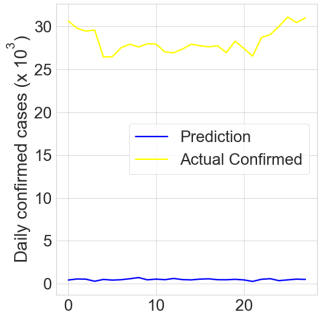


(b) Training and forecasting results of the proposed hybrid model for India

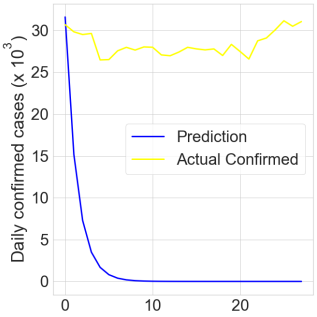


(c) Training and forecasting results of the proposed hybrid model for the UK

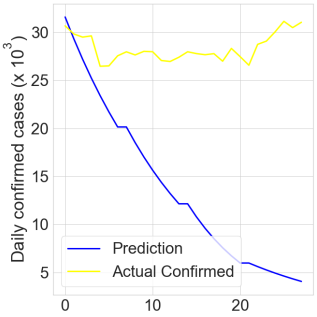
Figure 5.11: COVID-19 infected cases evolution study using the proposed hybrid model for the adopted country.



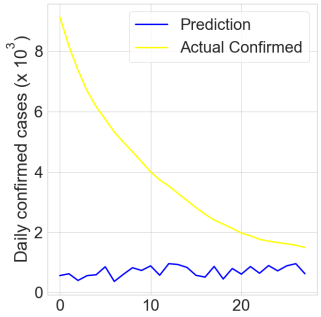
(a) USA forecast using LSTM for testing samples.



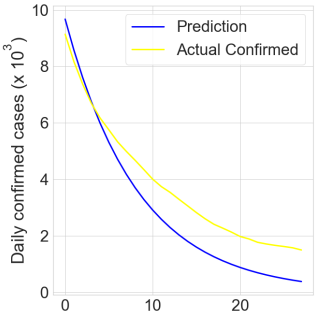
(b) USA forecast using hybrid of SIRD and PSO.



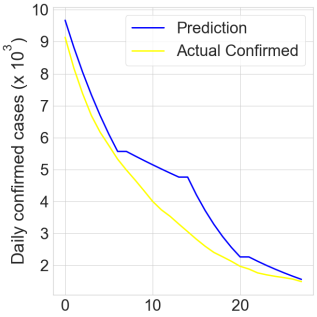
(c) USA forecast using the proposed hybrid model.



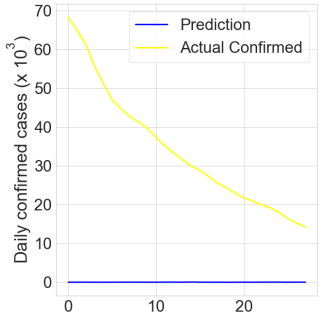
(d) India forecast using LSTM for testing samples.



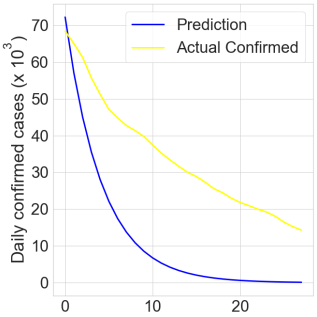
(e) India forecast using hybrid of SIRD and PSO.



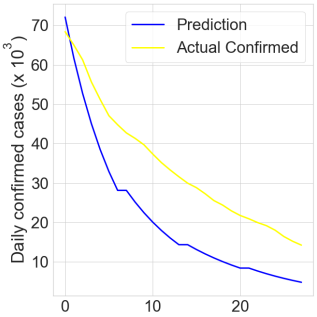
(f) India forecast using the proposed hybrid model.



(g) UK forecast using LSTM for testing samples.



(h) UK forecast using hybrid of SIRD and PSO.



(i) UK forecast using the proposed hybrid model.

Figure 5.12: Forecasting of the COVID-19 infected cases using the proposed model and compared models for the testing samples of the adopted country.

The simplicity and adaptability of the SIRD model make it a valuable tool for understanding and managing infectious disease outbreaks. LSTM is uniquely designed to handle sequential data, making it well-suited for time series forecasting, where past observations play a crucial role in predicting future values. It effectively manages noisy data by identifying underlying trends and discarding irrelevant information, thereby enhancing prediction accuracy. On the other hand, PSO serves as a powerful tool for hyperparameter optimization in forecasting models. Its computational efficiency and simplicity make it a preferred choice for improving time series forecasting without adding significant complexity. By optimizing model parameters, PSO further enhances the accuracy and performance of models like LSTM.

The graphical representation provides a relative comparison of the results. To precisely evaluate the forecasting accuracy of the models, statistical measures are utilized. Accordingly, the forecasting performance is assessed using the RMSE metric, as presented in Table 5.15, with the best results highlighted in bold. The table reveals that the hybrid model combining SIRD, PSO, and stacked-LSTM consistently outperformed all other models across all datasets. The combination of SIRD and PSO outperformed the standalone stacked-LSTM model. This demonstrates that integrating optimization techniques and deep learning with an epidemiological model significantly enhances forecasting performance.

5.9 Analysis and Discussion

This chapter investigates a variety of hybrid forecasting models for predicting COVID-19 cases, combining the strengths of FTS, contextual deep learning, and hyperparameter optimization using the PSO algorithm. FTS forecasting effectively handles the vagueness and imprecision inherent in non-stationary data, while the PSO algorithm optimizes the fuzzy order and interval lengths of the FTS model. FLR modeling and forecasting are performed using deep learning techniques. The proposed model is evaluated using COVID-19 confirmed case data from the USA, India, the UK, Russia, and Italy. Its performance is benchmarked against state-of-the-art methods, demonstrating that the hybrid model FTS-PSO-attention-Bi-LSTM statistically outperforms the other models. Several factors can influence model performance, making it impossible to claim that a model always performs best across all datasets. Therefore, a statistical significance test is necessary to draw conclusions. In this chapter, the hybrid model combining FTS, PSO, and Bi-LSTM with attention outperforms the others, as confirmed by the Nemenyi significance test. The proposed hybrid forecasting model successfully captures contextual patterns in the time series data and leverages past observations to predict future data points.

Further, an epidemiological compartmental model based study is carried out because these models are suitable for modeling the spread of epidemics. Literature shows that the SIRD epidemiological model effectively represents the progression of an epidemic in stationary scenarios, but it struggles to capture the time-varying trends. Therefore, it is not able to deal with the spread of the COVID-19 pandemic. To address this limitation, this study proposes a hybrid model that leverages the

strengths of PSO algorithm and stacked-LSTM to handle the dynamic nature of epidemics. The model has been evaluated using COVID-19 datasets from three different countries and compared with other state-of-the-art approaches. The proposed hybrid model consistently outperformed all other models across the adopted datasets. These results demonstrate the ability of the model to accurately project epidemic spread patterns. It can provide valuable insights for governments to take proactive measures.

CHAPTER 6

DESIGN OF NOVEL EPIDEMIOLOGICAL MODEL INCORPORATING LOCKDOWNS, MOBILITY RESTRICTIONS, AND VACCINATION FOR COVID-19 PREDICTION

[1] Throughout history, numerous pandemics and epidemics have emerged, significantly impacting humanity. Identifying the symptoms and behavior of a new virus takes time, and governments often face challenges in controlling its spread during the early stages of its evolution. COVID-19 stands out as one of the most severe pandemics in history. The SARS-CoV-2 virus has profoundly affected human health, leading to increased hospitalizations and fatalities [83]. Initially, key non-pharmaceutical methods (social distancing, self-isolation, contact tracing, quarantines, mandatory mask usage, lockdowns, and restrictions on travel) are adopted to curb the spread of virus [187, 114]. Over time, restrictions were gradually eased in a phased manner. Scientists worked tirelessly to develop vaccines to combat the disease, and several vaccines were eventually approved and distributed in stages [122]. Although COVID-19 is no longer classified as a global health emergency [126], it has raised critical questions for future pandemic preparedness. These include evaluating the effectiveness of restriction policies, vaccine efficacy and distribution strategies, and the role of epidemic modeling in policy formulation.

A primary measure to contain an epidemic is the implementation of a complete lockdown. However, prolonged lockdowns can have severe economic repercussions for both governments and individuals, making them an undesirable long-term solution [47]. As a result, partial or intermittent lockdowns are often recommended as more viable options to limit virus transmission while minimizing disruptions to essential services. During the COVID-19 pandemic, the SARS-CoV-2 virus mutated into several variants of concern [79]. These variants posed significant challenges due to their high transmission and mortality rates [37]. Numerous strategies and policies were employed to control the spread of COVID-19. Among these, vaccination emerged as one of the most effective solutions. However, developing a new vaccine requires time, and the emergence of new variants of a virus can compromise vaccine efficacy [154].

[1]The contents of this chapter are published in "*Novel Epidemiological Model Leveraging Resource-Optimized Containment Strategy with Hybrid Deep Learning and PSO Based Hyperparameter Optimization.*" in International Conference on Advancement in Communication and Computing Technology (INOACC), IEEE, April, 2025. and
in "*Swarm-Optimized Time-varying Epidemiological Model Incorporating Multi-dose Vaccinations, Vaccine Efficacy and Restriction Policies.*" in International Journal of Annals of Operations Research, 2025

Since the outbreak of COVID-19, researchers have proposed numerous compartmental epidemic models. These models typically divide a population into compartments to analyze the impact of various parameters on epidemic dynamics. One widely used baseline model for COVID-19 is the SEIRD epidemic model [50], where S, E, I, R, and D represent Susceptible, Exposed, Infected, Recovered, and Dead individuals, respectively. Although many studies have focused on modeling the spread of COVID-19, few have considered the combined effects of restriction policies, vaccine efficacy, and vaccination doses. To address this gap, this study introduces a novel epidemiological model that integrates these factors along with time-varying parameters. The proposed model, named SEIRDPV, extends the SEIRD framework by incorporating new compartments for multi-dose vaccinations, protected individuals, and immunized susceptibles. Additional parameters such as restriction policies, vaccine efficacy, and vaccination doses are explicitly included. To optimize the time-varying parameters of the model, the PSO algorithm is employed, as it provides efficient solutions with minimal computational cost [119]. The SEIRDPV model is evaluated using two case studies of the COVID-19 outbreaks in India and the USA, covering the first year of vaccination rollouts: January 16, 2021, to January 15, 2022 (India), and December 14, 2020, to December 13, 2021 (USA). Furthermore, the proposed model is compared with five state-of-the-art models: ETS [232], 2. ARIMA [230], 3. ANN [157], 4. LSTM [6], and 5. SEIRD [50]. The models are evaluated and compared using using five different performance metrics:- MAE, RMSE, RRSE, sMAPE, and MASE. The proposed methodology is described in the following sections.

6.1 Proposed 10-Compartments Epidemiological Model

At the onset of the COVID-19 pandemic, governments implemented complete lockdowns and travel restrictions to slow and halt the transmission of the virus. These restrictive measures were introduced in response to the limited availability of medical resources and the absence of vaccines [107]. The development of a COVID-19 vaccine was anticipated as a key solution to controlling the disease. Consequently, efforts to develop vaccines were rapidly accelerated [176]. Currently, several single-dose or double-dose based vaccines have been developed and approved for use in most countries worldwide. Table 6.1 presents a list of vaccines approved for adults during the initial phase of the pandemic. Vaccination programs have been implemented based on the prescribed number of doses and the required intervals between consecutive shots. In this study, vaccination parameters have been incorporated in alignment with the vaccines approved in the selected country for the experiments.

This study introduces a novel compartmental epidemiological model by extending the SEIRD framework with additional compartments for multi-dose vaccinations, protected individuals, and immunized susceptibles. The proposed model, referred to as the Susceptible-Exposed-Infected-Recovered-Deceased-Protected-Vaccinated (SEIRDPV) model, comprises ten compartments: Susceptible (S),

Table 6.1: Type of vaccines and efficacy with number of doses.

Vaccine	1st Dose Efficacy (%)	2nd Dose Efficacy (%)	Interval between 1st and 2nd Dose	Immunity Time (days)	Reference
Oxford/ AstraZeneca (Covishield)	57-74	70-90	12-16 Weeks	14	[21, 171]
Covaxin	78	78-95	4-6 weeks	14	[21, 60, 171]
Pfizer/ BioNTech	80	95	14	7	[72, 173]
Johnson & Johnson	85	-	-	14	[159]
Sputnik V	83-88	85.6-95.2	21	14	[202, 181]
Moderna	82	94.1	4-8 weeks	14	[72]

Exposed (E), Infected (I), Recovered (R), Deceased (D), Not-fully-vaccinated (V_{nf}), Fully-vaccinated (V_f), Booster (B), Protected (P), and Immunized susceptible (S'). These compartments are designed to reflect the dynamics of the post-vaccination phase of the COVID-19 pandemic. In the ?? model, illustrated in Fig. 6.1, the S compartment represents the population susceptible to infection. The Exposed (E) compartment represents individuals who have contracted the virus but remain asymptomatic, while the Infected (I) compartment contains confirmed positive cases exhibiting symptoms. Infected individuals may either recover, transitioning to the R compartment, or succumb to the disease, moving to the D compartment. Recovered individuals acquire natural immunity and are added to the P (Protected) compartment. Additionally, vaccinated individuals gain immunity after receiving the required doses and/or booster shots and are also added to the P compartment. However, depending on vaccine efficacy, some vaccinated individuals may still contract the infection and are moved to the E compartment. The state-transition diagram of the SEIRDPV model, showing the interactions and transitions between compartments, is provided in Fig. 6.1.

Most approved vaccines follow a one- or two-dose regimen to achieve maximum protection. To account for this, the proposed model introduces two new compartments: not-fully-vaccinated (e_{nf}) and fully-vaccinated (e_f). For single-dose vaccines, vaccinated individuals are directly added to the fully-vaccinated (e_f) compartment. Each vaccine type has specific efficacy and administration delays. Vaccines may not provide complete (100%) protection, even after all recommended doses. To capture this, the model designates the efficacy of not-fully-vaccinated individuals as e_{nf} and that of fully-vaccinated and booster recipients as e_f. After completing all required doses, an individual acquires immunity and transitions

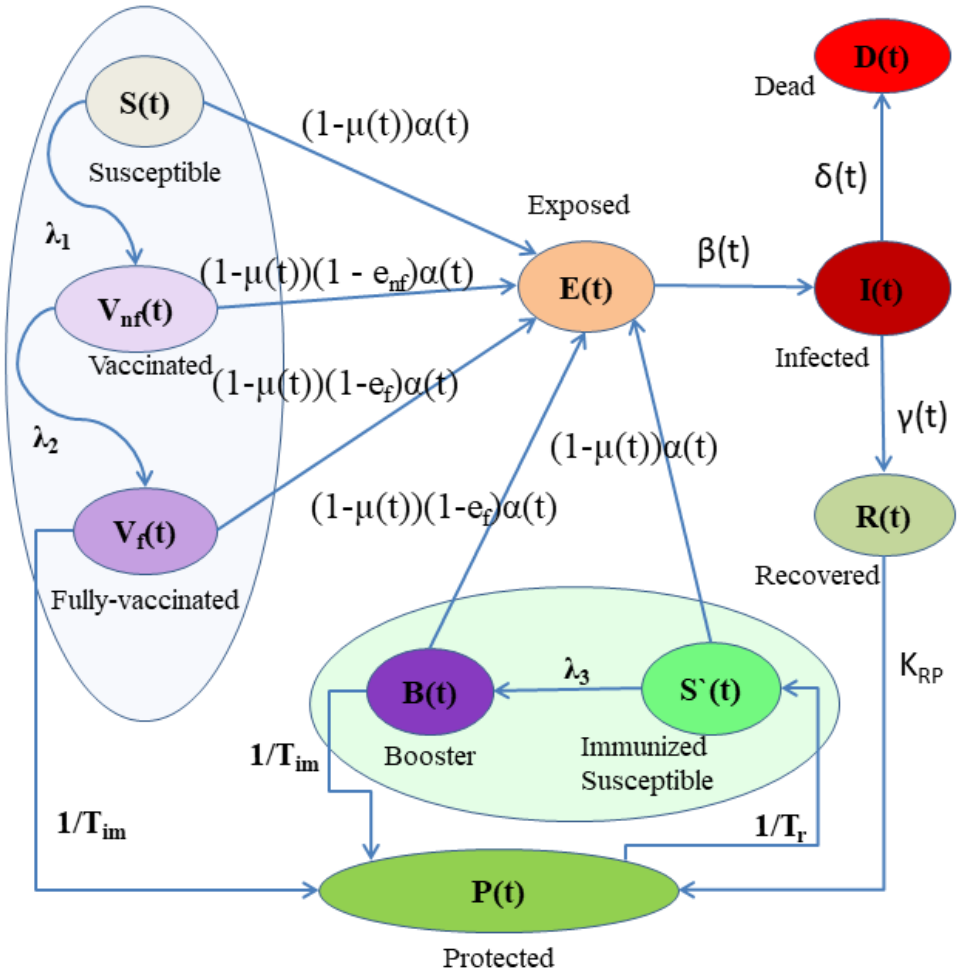


Figure 6.1: State-transition diagram of the proposed SEIRDPV model

to the protected (P) compartment after a delay of T_{im} days, representing the time required to develop immunity post-vaccination. Individuals in the P compartment retain immunity for a finite duration, T_r days, after which they move to the immunized susceptible (S') compartment. Those in S' compartment have higher immunity than individuals in the susceptible (S) compartment. To reflect governmental efforts to control disease spread, the model incorporates a restriction parameter (μ). This parameter accounts for the impact of restriction policies, with its value estimated based on factors such as the emergence of new variants or government advisories and restrictions implemented during specific periods. For simplicity, natural births and deaths are excluded from the model.

The proposed model is a bi-linear, positive system where all variables remain positive for time $t > 0$, provided they are initialized with non-negative values at $t = 0$. The dynamics of the system are described using ordinary differential equations (ODEs) derived from the state-transition diagram shown in Fig. 6.1. The ODEs are given below.

$$\dot{S} = -(1-\mu(t))\alpha(t)\frac{I(t)}{N(t)}S(t) - \lambda_1(t)S(t) \tag{6.1.1}$$

$$\dot{V_{nf}} = \lambda_1(t)S(t) - (1-\mu(t))(1-e_{nf})\alpha(t)V_{nf}(t) - \lambda_2(t)V_{nf}(t) \tag{6.1.2}$$

$$\dot{V_f} = \lambda_2(t)V_{nf}(t) - (1-\mu(t))(1-e_f)\alpha(t)V_f(t) - T_{im}^{-1}V_f(t) \tag{6.1.3}$$

$$\begin{aligned}\dot{E} = (1-\mu(t))\alpha(t)\frac{I(t)}{N(t)}S(t) + (1-\mu(t))(1-e_{nf})\alpha(t)V_{nf}(t) \\ + (1-\mu(t))(1-e_f)\alpha(t)V_f(t) + (1-\mu(t))\alpha(t)S'(t) \\ + (1-\mu(t))(1-e_f)\alpha(t)B(t) - \beta(t)E(t)\end{aligned} \tag{6.1.4}$$

$$\dot{P} = T_{im}^{-1}(V_f(t) + B(t)) + K_{RP}R(t) - T_r^{-1}P(t) \tag{6.1.5}$$

$$\dot{S'} = T_r^{-1}P(t) - (1-\mu(t))\alpha(t)S'(t) - \lambda_3(t)S'(t) \tag{6.1.6}$$

$$\dot{B} = \lambda_3(t)S'(t) - (1-\mu(t))(1-e_f)\alpha(t)B(t) - T_{im}^{-1}B(t) \tag{6.1.7}$$

$$\dot{I} = \beta(t)E(t) - \gamma(t)I(t) - \delta(t)I(t) \tag{6.1.8}$$

$$\dot{R} = \gamma(t)I(t) - K_{RP}R(t) \tag{6.1.9}$$

$$\dot{D} = \delta(t)I(t) \tag{6.1.10}$$

In the above equations, state variables at event time t are $S(t)$: susceptible individuals, $E(t)$: Exposed individuals, $I(t)$: Infected individuals, $R(t)$: recovered individuals, $D(t)$: dead individuals, $V_{nf}(t)$: not-fully-vaccinated individuals, $V_f(t)$: fully-vaccinated individuals, $B(t)$: Booster taken individuals, $P(t)$: Protected individuals, and $S'(t)$: Immunized susceptibles. For the mass conservation property, the summation of all the state variables is constant and equal to the total population N at time t, as given in equation 6.4.7.

$$N = S(t) + E(t) + I(t) + R(t) + D(t) + V_{nf}(t) + V_f(t) + B(t) + P(t) + S'(t) \tag{6.1.11}$$

The parameters used in the proposed model are described below.

- μ: The restriction parameter, representing the level of government-imposed restrictions during a specific timespan. It ranges from 0 to 1, where 0 indicates no restrictions (fully open) and 1 represents a complete lockdown.
- T_r: The duration of vaccinal and natural immunity after vaccination or recovery, set to 240 days (8 months) based on findings from [57].
- T_{im}: The time required to acquire immunity after receiving the full vaccination or booster dose. This duration is set to 14 days, as reported in [173].
- e_{nf}: The efficacy of the first dose of a vaccine. The required number of doses and corresponding efficacies for various vaccines are detailed in Table 6.1.
- e_f: The efficacy of the fully vaccinated or booster dose of a vaccine, as specified in Table 6.1.
- $\alpha(t)$, $\beta(t)$, $\gamma(t)$, $\delta(t)$, $\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$: These represent the exposed rate, infection rate, recovery rate, death rate, first-dose vaccination rate, second-dose vaccination rate, and booster-dose vaccination rate at time t, respectively.
- K_{RP}: The transition rate from the recovered compartment to the protected compartment. This parameter ranges from 0 to 1. In this model, it is set to 1, assuming that all recovered individuals acquire immunity.

The structure of the proposed compartmental epidemiological modeling and forecasting framework is illustrated in Fig. 6.2. The framework comprises several main components: data preprocessing, epidemic modeling with time-varying parameters, parameter optimization, forecasting, and evaluation metrics. Each component is distinctly outlined using dashed borders for clarity. Thick arrows represent transitions between the main components, while thin arrows indicate transitions within individual processing units or algorithms. The key stages of the proposed framework are described below.

- Data preprocessing: To reduce fluctuations, daily cumulative COVID-19 cases are averaged weekly. The Interquartile Range (IQR) method is applied to detect and remove outliers [224]. Min-max normalization is performed on the time series data [129]. The dataset is split into training and testing sets to evaluate the model on unseen data. A 30-day forecasting time span is chosen, as predicting an epidemic beyond one month may not be realistic due to dynamic scenarios.
- Model parameter estimation: The program is initialized with the following model parameters. (1) Most relevant events are identified to model the restriction policies. (2) The value of the restriction parameter $\mu(t)$ is estimated based on the event at time t. (3) The value of vaccination related parameters are set according to the approved vaccine and vaccination

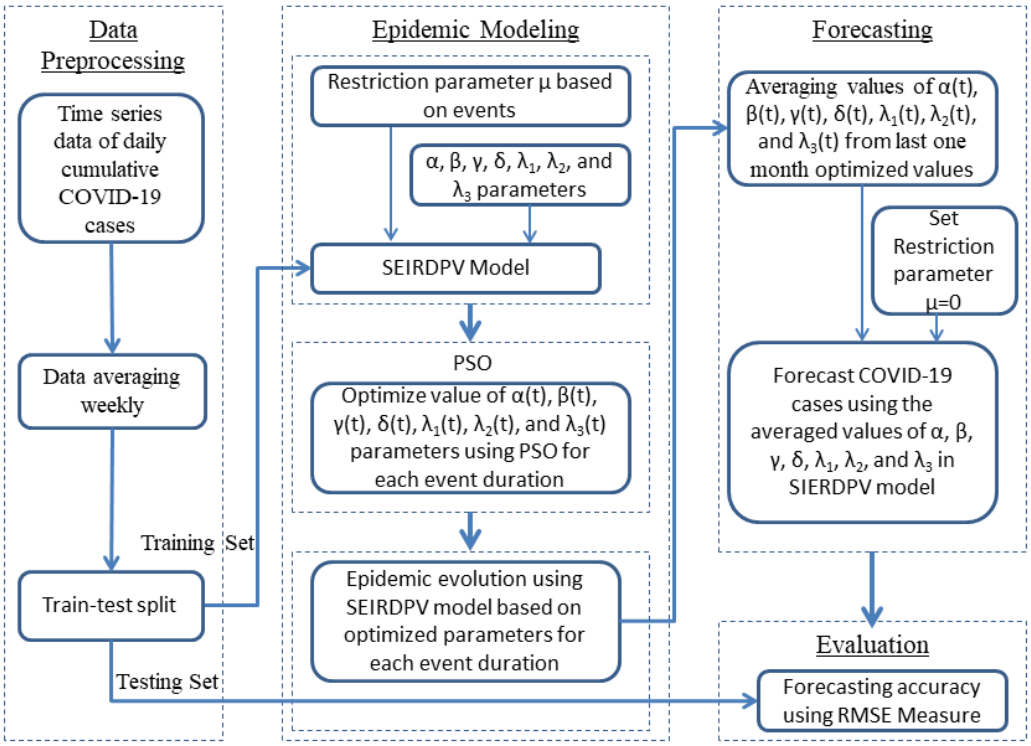


Figure 6.2: Proposed SEIRDPV epidemiological compartmental model forecasting framework.

programs in a country. (4) The value of the parameters $\alpha(t)$, $\beta(t)$, $\gamma(t)$, $\delta(t)$, $\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$ are generated randomly between 0 and 1 for duration t.

- Parameter optimization: Time-varying model parameters are optimized using Particle Swarm Optimization (PSO) by fitting the SEIRDPV model to a country's epidemic data. Optimization is performed for each event duration t, as parameters typically vary with governmental policies. Epidemic modeling is conducted using the optimized parameters.

- Forecasting: Forecasting for future days is performed using the average of the optimized parameters from the past month. Since future events are unknown, the restriction parameter $\mu(t)$ is varied between 0 and 1 to assess its impact on prediction results.

6.2 Experimental Setup and Dataset

The proposed model is implemented using Python 3.9. The COVID-19 time series data for India and the USA is sourced from the publicly available GitHub repository hosted by Our World in Data [92]. For each country, the starting point of the time series is set as the first date a vaccine was administered. The model is evaluated using COVID-19 time series data for confirmed, recovered, and death cases during the periods of January 16, 2021, to January 15, 2022 (India), and

December 14, 2020, to December 13, 2021 (USA), corresponding to the first year of vaccination rollouts in these countries. A one-year time frame is selected to encompass seasonal variations and their effects on the pandemic's progression, as well as the impact of immunization. The vaccination data includes information on not-fully vaccinated, fully vaccinated, and booster doses. Data preprocessing is conducted using three techniques.

1. Handling outliers with the Interquartile Range (IQR) method [224].
2. Weekly averaging of the time series data to mitigate fluctuations.
3. Min-max normalization of the time series data [129].

Table 6.2: PSO parameter settings

PSO Parameter	Value
Number of particles	100
Maximum number of iterations	1000
Inertia weight $[w_{min}, w_{max}]$	[0.4, 0.9]
Acceleration constants C1 = C2	2

The model parameters are optimized using the bio-inspired PSO algorithm. PSO effectively searches for an optimal or near-optimal solution to an optimization problem with minimal computational effort and avoids being trapped in local minima. Convergence analysis of the algorithm is conducted using the India and USA datasets for the proposed model. The optimal convergence parameters identified for PSO are presented in Table 6.2. Optimization of the proposed model parameters $\alpha(t)$, $\beta(t)$, $\gamma(t)$, $\delta(t)$, $\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$ is performed using PSO. The proposed SEIRDPV model is evaluated through two case studies involving India and the USA. Its performance is compared against five state-of-the-art methods: 1. ETS, 2. ARIMA, 3. ANN, 4. LSTM, and 5. SEIRD. Forecasting is limited to the next 30 days, as predicting an epidemic beyond one month may not be realistic due to the dynamic nature of COVID-19 transmission. The forecasting accuracy of the models is assessed using the metrics MAE, RMSE, RRSE, sMAPE, and MASE.

6.3 COVID-19 Case Studies and Evaluation

The proposed model is evaluated through case studies conducted for two countries, India and the USA, over the periods January 16, 2021, to January 15, 2022 (India), and December 14, 2020, to December 13, 2021 (USA), corresponding to the first year of vaccination program rollouts in each country. Detailed information on country-specific model parameters, significant events, optimization processes, and epidemic evolution for each case study is provided in the following subsections.

6.3.1 The USA Case Study

The mass vaccination campaign in the USA commenced on December 14, 2020 [218]. A one-year time series dataset, spanning from December 14, 2020, to December 13, 2021, is utilized for epidemic modeling. The Pfizer/BioNTech vaccine, which follows a two-dose regimen for optimal protection against COVID-19, was the primary vaccine administered in the USA. The vaccination dataset includes information on partially vaccinated, fully vaccinated, and booster dose recipients. Consequently, the experiments incorporate two-dose and booster-based vaccination compartments for modeling the epidemic in the USA. Additionally, the USA government implemented various restrictions over time to mitigate the spread of COVID-19, while individuals also self-imposed restrictions based on news and updates about the pandemic.

Table 6.3: Events and corresponding values of the restriction parameter ($\mu(t)$) for the USA during COVID-19.

Date	Event	Restriction Parameter ($\mu(t)$)
14 Dec 2020	COVID-19 vaccination started.	0.40
19 Dec 2020	Case of SARS-Cov-2 Alpha variant (B.1.1.7) was reported.	0.50
21 Jan 2021	Strategy is released for the pandemic preparedness and COVID-19 Response.	0.40
25 Jan 2021	SARS-Cov-2 Gamma variant (P.1) is detected.	0.45
28 Jan 2021	SARS-Cov-2 Beta variant (B.1.351) is detected.	0.35
27 Feb, 2021	Johnson & Johnson vaccine is approved for the vaccination.	0.30
02 Mar 2021	Full reopening is announced in Texas and Mississippi.	0.35
12 Apr 2021	SARS-Cov-2 Delta variant (B.1.617) detected.	0.30
19 Apr 2021	Vaccination eligibility is extended for individuals aged 16 and older.	0.15
13 May 2021	Compulsory facial mask restriction lifted for fully-vaccinated individuals in the most of the states.	0.20
01 Aug 2021	The USA passed 35 million cases of COVID-19.	0.10
08 Sep 2021	The USA passed 40 million cases of COVID-19.	0.15
01 Oct 2021	The USA passed 700000 deaths from COVID-19.	0.05
26 Nov 2021	Restrictions imposed on travel from South Africa and other impacted African countries due to the emergence of new variant called Omicron.	0.015
01 Dec 2021	SARS-Cov-2 Omicron variant is detected.	0.005

Table 6.3 highlights the most significant and impactful events of the COVID-19 pandemic in the USA. It includes the event timeline, descriptions, and the corresponding values of the restriction parameter ($\mu(t)$). These events are integrated

Table 6.4: Optimized value of time-varying SEIRDPV model parameters for each event period in the USA

From	To	α	β	γ	δ	λ_1	λ_2	λ_3
14 Dec 2020	20 Jan 2021	0.122330 ± 0.252120	0.878986 ± 0.219206	0.116691 ± 0.281240	0.013704 ± 0.000566	0.000136 ± 0.000081	0.026148 ± 0.005518	0 ± 0.0
21 Jan 2021	24 Jan 2021	0.768627 ± 0.047841	0.949413 ± 0.158665	0.989540 ± 0.031568	0.027281 ± 0.004910	0.000709 ± 0.000027	0.069829 ± 0.001815	0.000057 ± 0.000089
25 Jan 2021	27 Jan 2021	0.764900 ± 0.018356	0.422812 ± 0.024887	0.935090 ± 0.005400	0.025477 ± 0.002007	0.000831 ± 0.000009	0.113097 ± 0.000841	0.0000023 ± 0.000001
28 Jan 2021	26 Feb 2021	0.111997 ± 0.117288	0.222336 ± 0.203734	0.463980 ± 0.393917	0.025389 ± 0.000987	0.000299 ± 0.000113	0.087365 ± 0.014506	0.000002 ± 0.000001
27 Feb 2021	01 Mar 2021	0.119428 ± 0.122090	0.429410 ± 0.299722	0.885433 ± 0.123618	0.047845 ± 0.030713	0.000721 ± 0.000162	0.093794 ± 0.020173	0.000005 ± 0.000002
02 Mar 2021	11 Apr 2021	0.089126 ± 0.025931	0.567221 ± 0.412537	0.882409 ± 0.309044	0.020943 ± 0.000472	0.000526 ± 0.000044	0.070163 ± 0.002728	0.000006 ± 0.000001
12 Apr 2021	18 Apr 2021	0.088926 ± 0.001134	0.551343 ± 0.351726	0.992240 ± 0.003762	0.000082 ± 0.000007	0.000096 ± 0.000002	0.055500 ± 0.000249	0.000005 ± 0.000001
19 Apr 2021	12 May 2021	0.008293 ± 0.022420	0.144112 ± 0.300794	0.208086 ± 0.261582	0.010159 ± 0.003589	0.000227 ± 0.000045	0.104869 ± 0.007999	0.000005 ± 0.000001
13 May 2021	31 Jul 2021	0.001103 ± 0.003121	0.098863 ± 0.298350	0.013948 ± 0.038088	0.015776 ± 0.003295	0.000067 ± 0.000002	0.065487 ± 0.000809	0.000005 ± 0.000000
01 Aug 2021	07 Sep 2021	0.423800 ± 0.004361	0.772143 ± 0.267669	0.978698 ± 0.002155	0.006800 ± 0.000724	0.000394 ± 0.000005	0.160627 ± 0.000843	0.035341 ± 0.000143
08 Sep 2021	30 Sep 2021	0.474457 ± 0.005868	0.817573 ± 0.240166	0.972795 ± 0.007175	0.016193 ± 0.000446	0.000377 ± 0.000003	0.299483 ± 0.001479	0.072784 ± 0.000358
01 Oct 2021	25 Nov 2021	0.202823 ± 0.000767	0.819453 ± 0.238506	0.918534 ± 0.000551	0.017076 ± 0.000147	0.000436 ± 0.000004	0.354430 ± 0.000614	0.468067 ± 0.000573
26 Nov 2021	30 Nov 2021	0.221244 ± 0.002452	0.840013 ± 0.212505	0.998452 ± 0.001452	0.001746 ± 0.001308	0.000229 ± 0.000003	0.249734 ± 0.001708	0.251235 ± 0.002541
01 Dec 2021	13 Dec 2021	0.268168 ± 0.136830	0.806566 ± 0.303577	0.999879 ± 0.000008	0.012930 ± 0.001232	0.00092 ± 0.000201	0.657796 ± 0.120188	0.530186 ± 0.093933

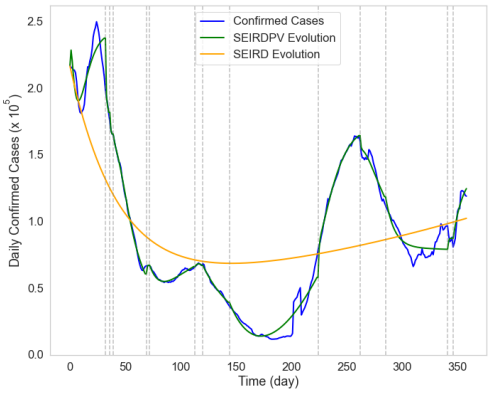
into the epidemiological modeling using both the SEIRD model and the proposed SEIRDPV model to simulate the evolution of COVID-19. Each event directly influenced movement of the population, leading to variations in model parameters [59]. The restriction parameter $\mu(t)$ is incorporated into the epidemiological modeling of the USA. The events listed are sourced from references [228, 229].

For the starting point of $(t = 0)$ on December 14, 2020, the values of the SEIRDPV model state variables for the USA are set as follows.
$N = 3.3 \times 10^8$, $I(0) = 216846$, $R(0) = 214151$, $D(0) = 315706, P(0) = 214814, V_{nf}(0) = 165826$, $V_f(0) = 662, B(0) = 0, S'(0) = 0$, $S(0) = N - I(0) - R(0) - D(0) - V_{nf}(0) - V_f(0) - B(0) - S'(0), E(0) = \frac{I(0)}{N} S(0)$.

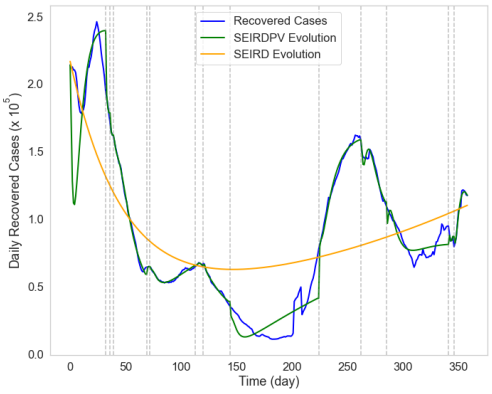
The value of PSO parameters are set as mentioned in Table 6.2. Modeling of the COVID-19 spread is started from the initial state conditions. SEIRDPV model parameters $\alpha(t)$, $\beta(t)$, $\gamma(t)$, $\delta(t)$, $\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$ are optimized using the PSO algorithm for each event time t given in Table 6.3. Each experiment is run 10 times to ensure the reliability of the optimized parameters. The mean values of the model parameters from 10 runs are given in Table 6.4 along with the standard deviation of the values. In the table, initially, value of $\lambda_3(t)$ is 0 because booster dose programs were started after completing first and second dose vaccination programs. It can be observed that the deviation in the optimized values is minimal, indicating the stability of the proposed model. The evolution of confirmed, recovered, and death cases of COVID-19 in the USA is simulated using the mean values of the optimized parameters, as presented in Table 6.4.

The evolution and forecasting results of COVID-19 cases in the USA using the proposed SEIRDPV model, along with a comparison to the state-of-the-art SEIRD model, are shown in Figure 6.3. The epidemiological evolution is performed using time series data of COVID-19 cases during the vaccination programs in the USA. Forecasting for the next 30 days is carried out using the optimized parameters of the SEIRDPV model, which are estimated based on the last month's data of the COVID-19 time series. Vertical dotted lines in the figures indicate the occurrence points of identified events. The SEIRDPV model parameters are re-estimated at the start of each event duration t, and the corresponding value of the restriction parameter $(\mu(t))$ is applied. Figure 6.3a shows the evolution of confirmed COVID-19 cases in the USA, with a comparison of the forecasting results between the SEIRD and the proposed SEIRDPV models. The results demonstrate that the time-varying parameters in the SEIRDPV model help it to closely match the real values of confirmed cases, while the SEIRD model, with parameters optimized on the training dataset, fails to effectively model the transmission dynamics of COVID-19. This highlights the importance of periodically estimating the parameters to account for dynamic conditions in epidemiological modeling.

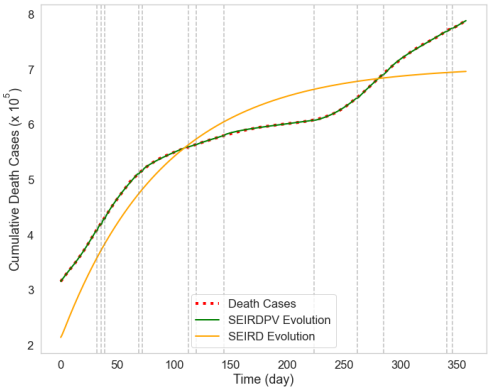
The evolution of recovered cases in the USA is depicted in Figure 6.3b. The correlation between recovered and confirmed cases is evident from the results shown in the figure. The proposed SEIRDPV model performs better in predicting recovered cases compared to the SEIRD model, as illustrated in the figure. Similarly, the evolution of death cases in the USA is presented in Figure 6.3c. The cumulative death cases in the SEIRDPV model closely align with the actual



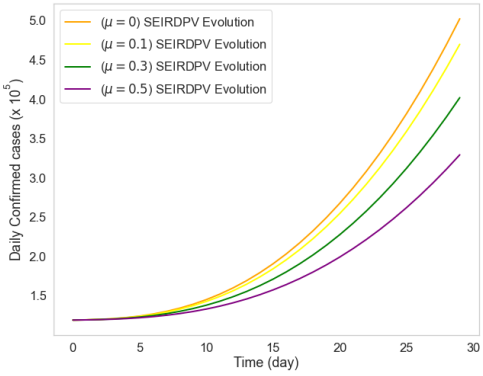
(a) Evolution of COVID-19 confirmed cases of the USA using SEIRD and proposed SEIRDPV model.



(b) Evolution of COVID-19 recovered cases of the USA using SEIRD and proposed SEIRDPV model.



(c) Evolution of COVID-19 death cases of the USA using SEIRD and proposed SEIRDPV model.



(d) Forecasting of confirmed cases of the USA for upcoming 30 days using SEIRDPV model with different values of the restriction parameter (μ).

Figure 6.3: Evolution and forecasting results of the proposed SEIRDPV model and comparison with the SEIRD epidemiological model on the USA dataset

values from the time series data. To distinguish the model's evolution from the actual time series, the evolution trend is shown with weekly data points. While the SEIRD model also demonstrates good performance in modeling death cases, the SEIRDPV model outperforms it.

Figure 6.3d presents the forecast of COVID-19 confirmed cases for the upcoming month using the SEIRDPV model for the USA. Forecasting is conducted with various values of the restriction parameter ($\mu(t)$) from the set $\{0, 0.1, 0.3, 0.5\}$, as shown in the figure. These different values of the restriction parameter are tested to examine the impact of restriction policies on the pandemic's spread. The results reveal that increasing the restriction parameter reduces the number of confirmed cases, but it does not necessarily result in the complete mitigation of the disease. Therefore, the study suggests that imposing more restrictions is not the definitive or ultimate solution in combating a pandemic like COVID-19.

6.3.2 India Case Study

For the India case study, COVID-19 time series data is collected starting from the initiation of the vaccination campaign on January 16, 2021. The study uses data from this period, covering one year from January 16, 2021, to January 15, 2022. The Covishield vaccine, which follows a two-dose regimen for maximum protection, was primarily administered in India. As a result, a two-dose vaccination compartmental model, similar to the one used for the USA, is applied in the analysis for India.

Table 6.5: Events and corresponding values of the restriction parameter ($\mu(t)$) in India.

Date	Event	Restriction Parameter ($\mu(t)$)
16 Jan 2021	First phase of vaccination started. This phase covered frontline workers and health workers.	0.20
01 Feb 2021	Restrictions Unlock 9.0.	0.15
01 Mar 2021	Restrictions Unlock 10.0.	0.10
01 Apr 2021	Partial lockdown imposed in some states and vaccination eligibility was extended to all residents over the age of 45. Restrictions Unlock 11.0 in other states.	0.15
08 Apr 2021	Vaccination started for age over 18.	0.10
01 May 2021	Restrictions Unlock 12.0.	0.08
01 Jun 2021	Restrictions Unlock 13.0.	0.07
22 Jun 2021	SARS-Cov-2 Delta plus variant detected.	0.10
01 July 2021	Restrictions Unlock 14.0.	0.09
01 Aug 2021	Restrictions Unlock 15.0.	0.08
01 Sep 2021	Restrictions Unlock 16.0.	0.07
01 Oct 2021	Restrictions Unlock 17.0.	0.06
01 Nov 2021	Restrictions Unlock 18.0.	0.05

Continued on next page

Table 6.5 – continued from previous page

Date	Event	Restriction Parameter ($\mu(t)$)
02 Dec 2021	SARS-Cov-2 Omicron variant detected and Restrictions Unlock 19.0.	0.04
01 Jan 2022	Restrictions Unlock 20.0.	0.02

Restriction policies and vaccination programs have significantly influenced the spread pattern of COVID-19 in India. Several key events that occurred during the adopted COVID-19 time series period in India are identified. For each event, the time, description, and corresponding value of the restriction parameter ($\mu(t)$) are listed in Table 6.5 for India. These intervention events are incorporated into the proposed model to accurately reflect the real-world scenarios in epidemiological modeling. The model parameters are re-estimated for each event duration t using PSO-based optimization. The events included in the study are sourced from the links [227, 86].

For the starting point of $(t = 0)$ on January 16, 2021, the values of the model state variables for India are set as follows.
$N = 1.3\times10^9$, $I(0) = 13964$, $R(0) = 13804$, $D(0) = 152565$, $P(0) = 13804$, $V_{nf}(0) = 198656$, $V_f(0) = 0$, $B(0) = 0$, $S'(0) = 0$, $S(0) = N - I(0) - R(0) - D(0) - V_{nf}(0) - V_f(0) - B(0)$, $E(0) = \frac{I(0)}{N}S(0)$.

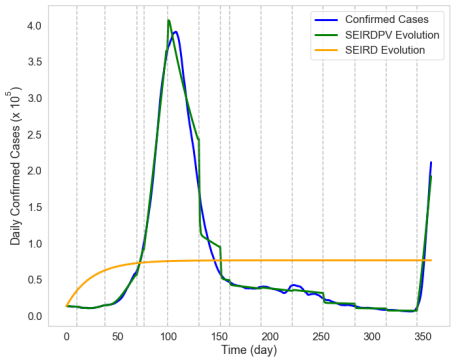
The model parameters are optimized using PSO to match the spread pattern of COVID-19 cases in India. The optimized parameter values for each event duration t in India are shown in Table 6.6. These values represent the mean of 10 runs of the proposed model for each event duration. The table also includes the average values and standard deviations of the parameters from the 10 runs for each event duration. These averaged parameter values are then used in the SEIRDPV model for COVID-19 epidemic modeling.

Figure 6.4 presents the evolution and forecasting results of COVID-19 cases in India using the proposed SEIRDPV model, alongside a comparison with the state-of-the-art SEIRD model. The dataset from India captures a significant wave of COVID-19 (the Delta variant) and the early phase of the subsequent wave (the Omicron variant). The evolution of confirmed COVID-19 cases in India is shown in Figure 6.4a. The results demonstrate that the SEIRDPV model, with its time-varying parameters, achieves close alignment with observed transmission patterns, whereas the SEIRD model proves inadequate in capturing the true spread dynamics.

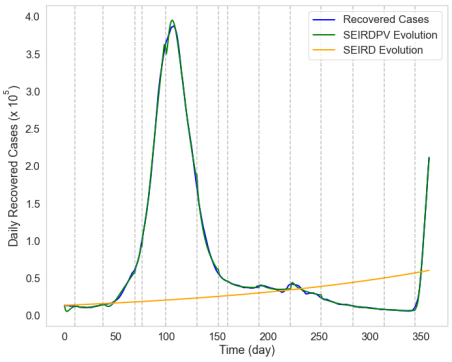
Figure 6.4b illustrates the evolution of COVID-19 recovered cases in India using both the SEIRD and the proposed SEIRDPV model. The proposed model outperforms the SEIRD model. Similarly, Figure 6.4c shows the evolution of COVID-19 death cases, where the SEIRDPV model closely matches the actual values. The proposed model demonstrates better performance compared to the SEIRD model.

Table 6.6: Optimized value of parameters for each event period in India

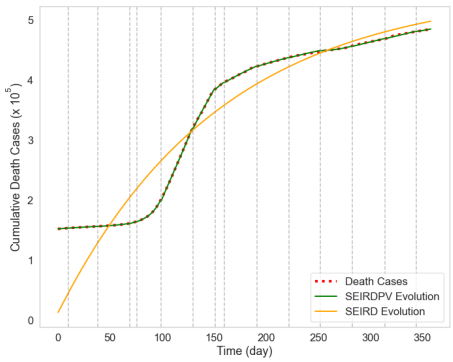
From	To	α	β	γ	δ	λ_1	λ_2	λ_3
16 Jan 2021	31 Jan 2021	0.200000 ± 0.000000	0.727273 ± 0.000000	0.303030 ± 0.000000	0.000021 ± 0.000016	0.000017 ± 0.000002	0 ± 0.0	0 ± 0.0
01 Feb 2021	28 Feb 2021	0.181200 ± 0.030330	0.533410 ± 0.312490	0.426890 ± 0.200677	0.001318 ± 0.003757	0.000012 ± 0.000002	0 ± 0.0	0 ± 0.0
01 Mar 2021	31 Mar 2021	0.108239 ± 0.004708	0.091587 ± 0.013260	0.992942 ± 0.008827	0.011302 ± 0.002542	0.000094 ± 0.000001	0 ± 0.0	0 ± 0.0
01 Apr 2021	07 Apr 2021	0.137605 ± 0.002481	0.118732 ± 0.015180	0.999777 ± 0.000164	0.000915 ± 0.002482	0.000340 ± 0.000001	0.021178 ± 0.000062	0 ± 0.0
08 Apr 2021	30 Apr 2021	0.326703 ± 0.006036	0.099052 ± 0.004883	0.917752 ± 0.015578	0.005335 ± 0.001826	0.000182 ± 0.000004	0.044076 ± 0.000380	0 ± 0.0
01 May 2021	31 May 2021	0.326315 ± 0.002923	0.154295 ± 0.006756	0.976473 ± 0.002098	0.012669 ± 0.001371	0.000143 ± 0.000001	0.054019 ± 0.000378	0 ± 0.0
01 Jun 2021	21 Jun 2021	0.045022 ± 0.003088	0.091948 ± 0.004183	0.859968 ± 0.012244	0.030866 ± 0.002206	0.000099 ± 0.000002	0.012584 ± 0.000102	0 ± 0.0
22 Jun 2021	30 Jun 2021	0.999908 ± 0.000093	0.006467 ± 0.000377	0.933687 ± 0.072023	0.052441 ± 0.019647	0.001388 ± 0.000001	0.055336 ± 0.000327	0 ± 0.0
01 Jul 2021	31 Jul 2021	0.969843 ± 0.030715	0.001187 ± 0.000031	0.967975 ± 0.019573	0.023752 ± 0.008359	0.000985 ± 0.000030	0.150275 ± 0.003481	0 ± 0.0
01 Aug 2021	31 Aug 2021	0.222649 ± 0.266898	0.000682 ± 0.000092	0.925986 ± 0.095263	0.010389 ± 0.003909	0.000422 ± 0.000342	0.045105 ± 0.020509	0 ± 0.0
01 Sep 2021	30 Sep 2021	0.000073 ± 0.000053	0.000610 ± 0.000090	0.975826 ± 0.020078	0.011636 ± 0.001688	0.000178 ± 0.000002	0.047521 ± 0.000087	0 ± 0.0
01 Oct 2021	31 Oct 2021	0.038326 ± 0.079104	0.000311 ± 0.000047	0.952723 ± 0.039576	0.012949 ± 0.001136	0.000169 ± 0.000107	0.076588 ± 0.020008	0 ± 0.0
01 Nov 2021	30 Nov 2021	0.000617 ± 0.001504	0.000194 ± 0.000029	0.958419 ± 0.043285	0.038901 ± 0.013716	0.000267 ± 0.000003	0.160291 ± 0.000819	0 ± 0.0
02 Dec 2021	31 Dec 2021	0.000099 ± 0.000035	0.000138 ± 0.000021	0.950000 ± 0.041268	0.047941 ± 0.015921	0.000259 ± 0.000003	0.186651 ± 0.000301	0 ± 0.0
01 Jan 2022	15 Jan 2022	0.274265 ± 0.001810	0.000389 ± 0.000521	0.092841 ± 0.293586	0.005353 ± 0.007698	0.000887 ± 0.000009	0.117150 ± 0.000389	0.008660 ± 0.000009



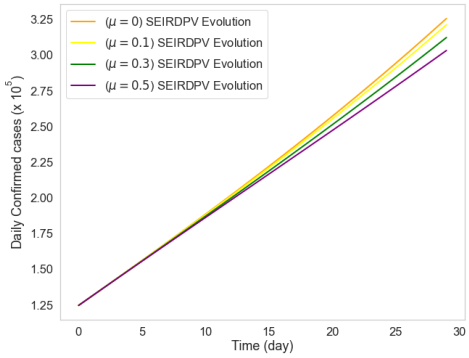
(a) Evolution of the COVID-19 confirmed cases of India using SEIRD and proposed SEIRDPV model.



(b) Evolution of the COVID-19 recovered cases of India using SEIRD and proposed SEIRDPV model.



(c) Evolution of the COVID-19 death cases of India using SEIRD and proposed SEIRDPV model.



(d) Forecasting of India confirmed cases using optimized SEIRDPV model parameters and different values of the restriction parameter.

Figure 6.4: Evolution and forecasting results of the proposed SEIRDPV model and comparison with the SEIRD epidemiological model on India dataset.

Additionally, the proposed SEIRDPV model is used to forecast COVID-19 confirmed cases in India for the next 30 days. The forecasting is based on the optimized parameters from the previous month's time series data. The results of the 30-day forecast for confirmed cases are shown in Fig. 6.4d. To assess the impact of the uncertain restriction parameter (μ) in the future, forecasts are made with varying values of μ, as shown in Fig. 6.4d. The results indicate that increasing the value of μ does not halt the pandemic's spread; rather, it only delays the timeline of the spread.

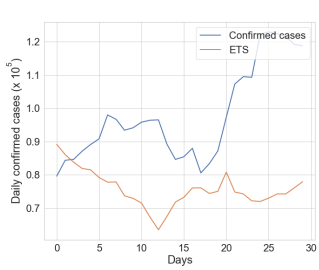
6.3.3 Empirical Analysis of the Models

We perform a rigorous comparative assessment using five evaluation metrics to analyze the predictive performance of our SEIRDPV model relative to ETS, ARIMA, ANN, LSTM, and SEIRD approaches. The performance metrics used for evaluation include RMSE, MAE, RRSE, sMAPE, and MASE. Short-term (30-day) forecasting is performed for the comparative analysis, taking into account the dynamic nature of the pandemic, as long-term forecasting may not be reliable in such cases. The models are trained on the training dataset, and their forecasting performance is evaluated on a testing dataset (30 days). The evaluation focuses on COVID-19 confirmed cases. The SEIRD and SEIRDPV models are assessed with a restriction parameter μ set to 0.2, assuming no major government restrictions during the forecasting period. The ETS, ARIMA, ANN, and LSTM models are trained on confirmed COVID-19 cases. The comparative forecasting results for India and the USA are presented in the following subsections.

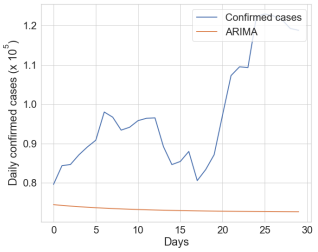
Model comparison on the USA dataset

The forecasting results for the USA are shown in Fig. 6.5. The performance of the ETS model for forecasting the next 30 days is displayed in Fig. 6.5a. It can be seen that ETS fails to capture the actual fluctuations, as it relies on smoothing data points. The ARIMA model also does not perform well in forecasting COVID-19 confirmed cases in the USA, as shown in Fig. 6.5b. In contrast, both the ANN and LSTM models successfully learned the dynamics of COVID-19 confirmed cases from the testing data and adapted to the evolving trends, as demonstrated in Fig. 6.5c and Fig. 6.5d. The SEIRD and proposed SEIRDPV models, based on statistical evolution with hyperparameters, require regular updates to account for the changing transmission dynamics of the pandemic. Additionally, these models are capable of incorporating restriction parameters, multi-dose vaccination programs, and the evolution of various related aspects.

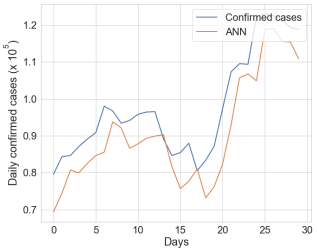
The forecasting results of the models are compared using various performance metrics, as shown in Table 6.7, with the best performance values highlighted in bold. From the table, it is evident that all models exhibit varying levels of performance across different metrics. The SEIRDPV model outperforms the other models in terms of MAE and sMAPE, while the ANN model leads in RMSE, sMAPE, and MASE metrics. These findings suggest that combining epidemiological models with machine learning techniques can be effective in addressing dynamic scenarios and uncovering hidden patterns in time series data.



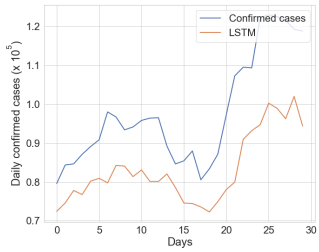
(a) Forecasting using ETS



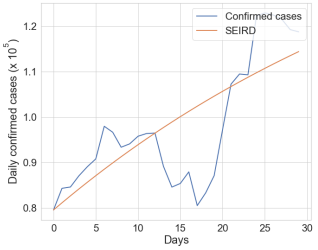
(b) Forecasting using ARIMA



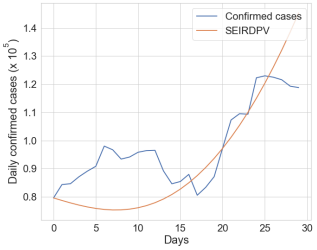
(c) Forecasting using ANN



(d) Forecasting using LSTM



(e) Forecasting using SEIRD



(f) Forecasting using proposed SEIRDPV

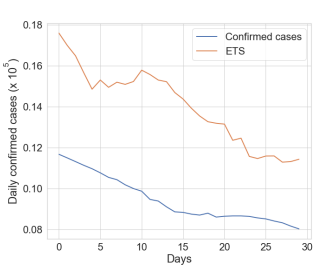
Figure 6.5: Forecasting results of the models for testing samples of the COVID-19 confirmed cases of the USA

Table 6.7: Forecasting accuracy of the models on the USA dataset using different measures (best values are shown in bold)

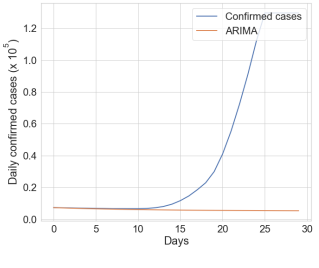
Model	USA				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS [232]	22964	27521	1.988	25.62	10.85
ARIMA [230]	24719	28489	2.058	27.83	396.36
ANN [157]	7146	**8300**	**0.60**	7.81	**1.69**
LSTM [6]	14705	16031	1.16	15.80	5.26
SEIRD [50]	10312	12778	0.92	11.12	4.02
SEIRDPV	**7134**	9457	0.68	**7.26**	5.93

Model comparison on India dataset

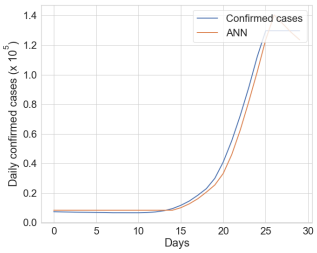
The performance of a model can vary across different datasets. In this context, the proposed model is compared with five state-of-the-art models using the COVID-19 confirmed cases dataset from India. The forecasting results of these models on the India dataset are shown in Fig. 6.6, which align with the results from the USA dataset. As seen in 6.6b, the ARIMA model does not perform well on the India dataset either. The ETS model performs relatively well across all five performance metrics, although the ANN model outperforms the others. As shown in Fig. 6.6c, the ANN model achieves the best results. The SEIRD model fails to match the testing time series, as shown in Fig. 6.6e, while the SEIRDPV model closely matches the testing time series of confirmed cases, as shown in Fig. 6.6f. While graphical results do not clearly identify a superior model, five performance metrics are used to determine the better model.



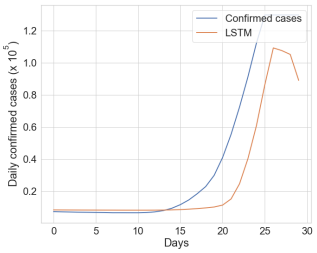
(a) Forecasting using ETS



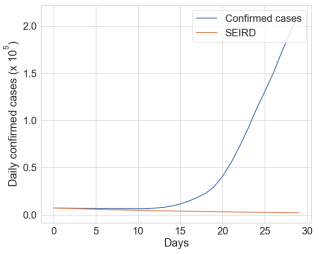
(b) Forecasting using ARIMA



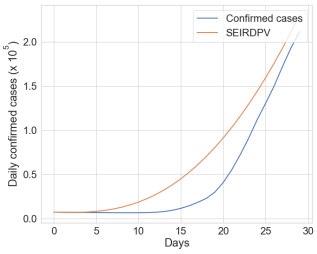
(c) Forecasting using ANN



(d) Forecasting using LSTM



(e) Forecasting using SEIRD



(f) Forecasting using SEIRDPV

Figure 6.6: Forecasting results of the models for testing samples of the COVID-19 confirmed cases of India

The forecasting accuracy results of the compared models on the testing dataset are presented in Table 6.8, with the best values highlighted in bold. From the table, it is evident that the ANN model outperforms all other models across all performance metrics. The next best-performing model is LSTM. These results suggest that machine learning models are better suited to capturing the dynamics in data that reflect multiple waves of the pandemic (e.g., Delta, Omicron). In contrast, epidemiological models like SEIRD can only model disease evolution based on preset hyperparameters. However, an epidemiological model like the proposed SEIRDPV can effectively capture real-time dynamics if the hyperparameters are

Table 6.8: Forecasting accuracy of the models on India dataset using different measures (best values are shown in bold)

Model	India				
	MAE	RMSE	RRSE	sMAPE	MASE
ETS [232]	4558	4675	4.27	38.56	13.66
ARIMA [230]	34903	59295	1.24	81.66	514.66
ANN [157]	**3621**	**4902**	**0.11**	**14.37**	**0.71**
LSTM [6]	14724	22924	0.48	39.27	3.52
SEIRD [50]	43534	77018	1.23	96.82	250.21
SEIRDPV	22077	27950	0.45	107.55	2.86

periodically re-estimated or updated based on specific events.

6.4 A study on resource-optimized restrictions

Epidemiological models are essential tools for estimating the transmission dynamics of infectious diseases. Among these, the Susceptible-Exposed-Infected-Recovered-Dead (SEIRD) model has been extensively studied for disease modeling [50]. However, the COVID-19 pandemic has introduced several additional complexities, such as asymptomatic infections, quarantine protocols, hospitalization, ICU admissions, and vaccination efforts. To better capture these dynamics, many researchers have extended the SEIRD framework to incorporate these factors and evaluate their effects on outbreak progression [116]. In this study, we extend the SEIRD model by adding a new compartment for hospitalization (H), resulting in a SEIHRD model that includes seven distinct compartments. Given that disease dynamics evolve over time, relying on static model parameters may not effectively reflect future conditions. Therefore, this study emphasizes the importance of using time-varying parameters that can adapt to shifts in disease transmission and intervention strategies.

A variety of optimization algorithms are available to identify optimal solutions with minimal computational effort. Among these, PSO has demonstrated high effectiveness [119]. In this study, a recent variant of the PSO algorithm is utilized to optimize the parameters of the SEIHRD model, which includes five hyperparameters that must be estimated concurrently. To facilitate this process, a multi-variable deep learning model based on LSTM networks is developed. This deep learning model is applied to the weekly optimized hyperparameters to generate short-term forecasts over a four-week horizon, recognizing that long-term forecasting of infectious diseases may not be reliable. The model is trained on historical hyperparameter estimates to produce accurate short-term projections. These forecasted hyperparameters are then input into the SEIHRD model to simulate disease transmission patterns. As part of the epidemiological modeling, a containment measure is derived using the ratio of available hospital beds to hospitalized cases. This measure, which ranges from 0 to 1, can be instrumental in shaping containment strategies, although the formulation of specific policy recommendations falls outside the scope of this study. COVID-19 time series data

from two countries—India and the United States is used to train and evaluate the proposed model. Forecasting performance is assessed using statistical metrics and compared against benchmark models.

6.4.1 Proposed Epidemiological Compartmental Model

In this study, the traditional SEIRD model is extended by adding a new compartment to represent hospital admission capacity. The resulting model, known as the ?? model, consists of six compartments: Susceptible (S), Exposed (E), Infected (I), Hospitalized (H), Recovered (R), and Deceased (D). The model structure is illustrated in Fig. 6.7. The Susceptible (S) group includes individuals at risk of infection. The Exposed (E) compartment contains individuals who have encountered the virus but are not yet symptomatic. The Infected (I) compartment represents confirmed cases. Infected individuals may either be admitted to the Hospitalized (H) compartment, recover (R), or die (D). Those who recover eventually return to the Susceptible (S) group after their temporary immunity period expires.

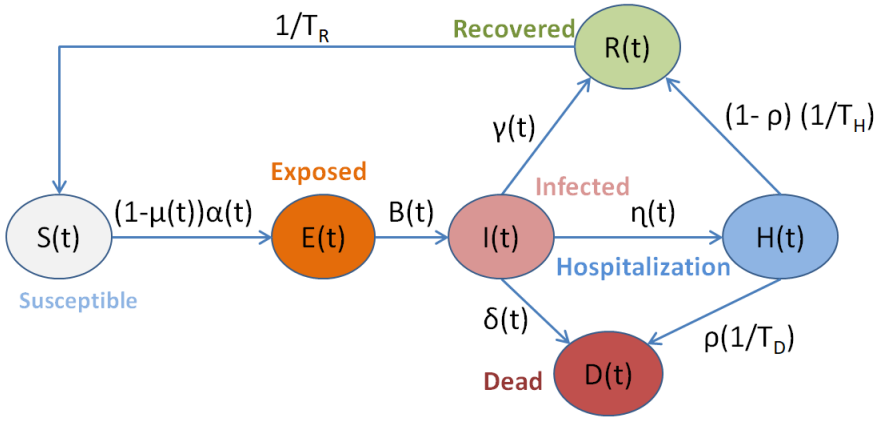


Figure 6.7: Proposed SEIHRD model state-transition diagram.

The proposed model is a bilinear and positive system, which guarantees that all state variables remain non-negative for $t > 0$, given that their initial values are non-negative at $t = 0$. The corresponding ordinary differential equations (ODEs) of the system, derived from the state-transition diagram shown in Fig. 6.7, are provided below.

$$\dot{S}(t) = -(1-\mu(t))\alpha(t)\frac{I(t)}{N(t)}S(t) + \frac{1}{T_R}R(t) \tag{6.4.1}$$

$$\dot{E}(t) = (1-\mu(t))\alpha(t)\frac{I(t)}{N(t)}S(t) - \beta(t)E(t) \tag{6.4.2}$$

$$\dot{I}(t) = \beta(t)E(t) - \gamma(t)I(t) - \delta(t)I(t) - \eta(t)I(t) \tag{6.4.3}$$

$$\dot{H}(t) = \eta(t)I(t) - \rho H(t)\frac{1}{T_D} - (1-\rho)\frac{1}{T_H}H(t) \tag{6.4.4}$$

$$\dot{R}(t) = \gamma(t)I(t) + (1-\rho)\frac{1}{T_H}H(t) - \frac{1}{T_R}R(t) \tag{6.4.5}$$

$$\dot{D}(t) = \delta(t)I(t) + \rho H(t)\frac{1}{T_D} \tag{6.4.6}$$

The above equations preserve the mass conservation of state variables, ensuring that their sum remains equal to the total population N at time t, as expressed in Equation 6.4.7.

$$N = S(t) + E(t) + I(t) + H(t) + R(t) + D(t) \tag{6.4.7}$$

The variables used in equations of the proposed model are described below.

- Restriction parameter (μ): It is determined by the restriction policies of government during a specific time period and is assigned a value between 0 and 1, where 0 represents no-restrictions and 1 indicates a complete lockdown.
- Immunity period (T_R): The number of days of immunity after recovery from the decease. It is set to 240 days (8 months) based on the study in [57].
- Hospital admission period (T_H): Average time period from admission to discharge of an infected individual. It is set to 16 days.
- Death probability (ρ): Probability of death of an infected individual after hospital admission. It is set to 0.15.
- Death period (T_D): It is an average time period from admission to death of an infected individual. It is set to 16 days.
- $\alpha(t)$, $\beta(t)$, $\gamma(t)$, $\eta(t)$, and $\delta(t)$ are the exposed rate, infection rate, recovery rate, admission rate, and death rate at time t, respectively. These are estimated weekly.

Resource-Optimized Restriction Parameter Estimation

This study investigates the optimal utilization of hospital beds by estimating a restriction parameter designed to keep the number of infections within the limits of available hospital capacity. A country's hospitalization capacity is quantified as the number of hospital beds per thousand individuals. Based on this metric, the maximum hospitalization capacity can be expressed by Equation 6.4.8.

$$H_{max} = \frac{N}{1000} \times hospital\_beds\_per\_thousand \tag{6.4.8}$$

During a pandemic, a portion of hospital beds must be reserved. Thus, a practical capacity (set point) is determined by Equation 6.4.9. In this study, 15% beds are reserved for other patients in the pandemic modeling.

$$H_{sp} = H_{max} \times (1 - hospital\_beds\_for\_other) \tag{6.4.9}$$

The value of restriction parameter (μ) at time t is determined using Equation 6.4.9 and model parameters as given in Equation 6.4.10.

$$\mu(t) = 1 - \frac{H_{sp} - H(t) - \eta(t)I(t)}{H_{sp} - H(t)} \tag{6.4.10}$$

6.4.2 Forecasting Framework

Forecasting framework based on the proposed compartmental epidemiological model is shown in Fig. 6.8. The main components of the framework are described below.

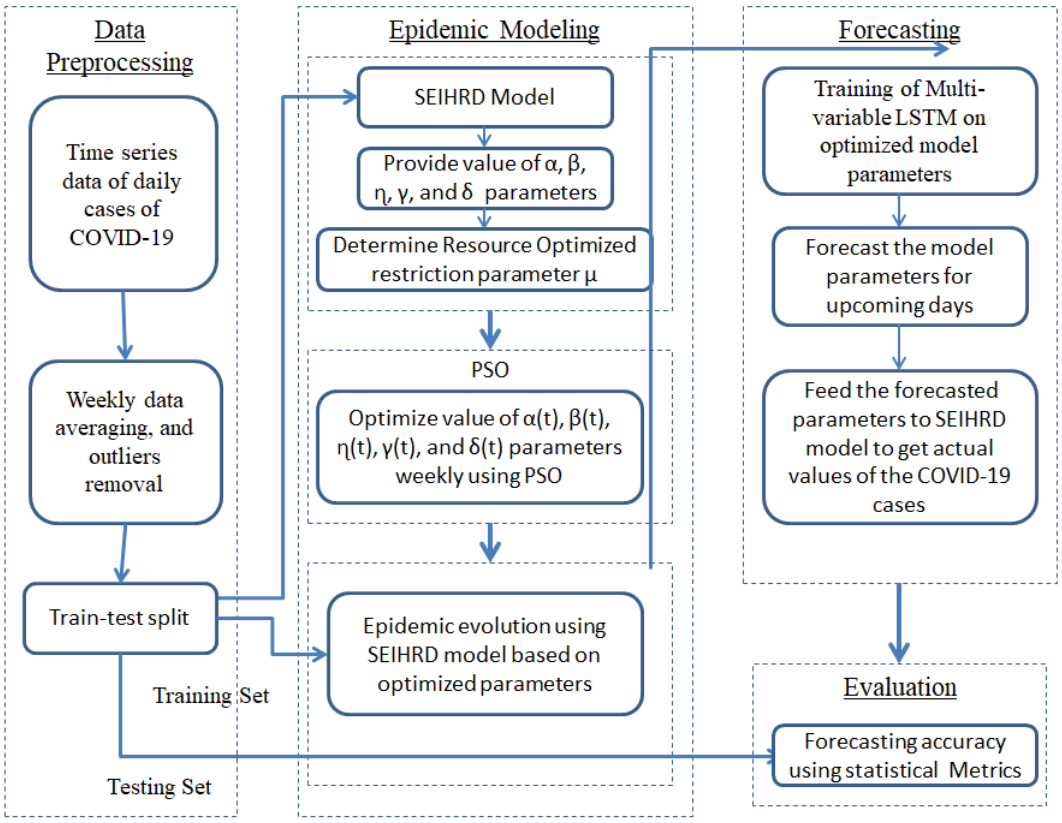


Figure 6.8: Forecasting framework for the proposed SEIHRD model.

1. Data pre-processing: Two techniques are applied to pre-process the data:- (i) Weekly averaging of the time series of COVID-19 cases to smooth the data. (ii) Interquartile Range (IQR) method is used to remove outliers from the data [224].

2. Training-testing data: The pre-processed data is split into train and test samples. The last 28 days samples are kept for testing. More than one month forecasting may not be realistic in case of highly contangious diseases.

3. SEIHRD modeling: The proposed model is used for COVID-19 pandemic evolution. The model parameters $S0, E0, I0, H0, R0, D0$ are initialized using

the real values at $t = 0$, and weekly $S(t), E(t), I(t), H(t), R(t), D(t)$ using values at time $t > 0$.

4. Resource-Optimized containment parameter estimation: Resource constraint based restriction parameter ($\mu(t)$) is estimated weekly using equations described in Section 6.4.1.

5. PSO optimization: The SEIHRD model parameters (α, β, γ, η, δ) are optimized at weekly interval using the PSO algorithm. The optimal values of the model parameters are returned by the PSO.

6. Deep learning: A multi-variable-LSTM is Incorporated to forecast the value of SEIHRD model parameters. The weekly optimized parameters (α, β, γ, η, δ) are used to train the LSTM model, and perform forecasting for four weeks.

7. Pandemic evolution and forecasting: In this step, Multi-variable-LSTM forecasted model parameters are fed into the SEIHRD model to show evolution of the COVID-19 pandemic.

Datasets

COVID-19 case data from the United States and India are used to evaluate the proposed model. The dataset, which is publicly available, is sourced from the GitHub repository maintained by Our World in Data [93]. The evaluation is based on reported cases from April 1, 2020, to September 30, 2020.

PSO, LSTM, and Model Variables Configurations

In the experiments, the value of $hospital\_beds\_per\_thousand = 2.77$ and population $N = 3.3 \times 10^8$ are set for the USA. Similarly, $hospital\_beds\_per\_thousand = 0.53$ and $N = 1.4 \times 10^9$ are set for India.

PSO algorithm is used to optimize the proposed model hyperparameters. The PSO parameters are configured as shown in Table 6.9.

Table 6.9: Configured parameters of PSO algorithm.

PSO Parameter	Value
Number of particles	50
Maximum number of iterations	300
Inertia weight $[w_{min}, w_{max}]$	[0.4, 0.9]
C1 = C2	[0.5, 2.5]

Multi-variable-LSTM is used for deep learning of the SEIHRD model parameters. Configured values of LSTM parameters are shown in Table 6.10.

Table 6.10: Configured values of LSTM parameters.

Parameter	Value
Activation Function	linear
Optimizer	Adam
Hidden units	20
Dense units	4
Learning rate	0.005
Loss function	Huber
Evaluation metrics	MSE
Epochs	500

6.4.3 Experimental Results

Based on the configured parameters of the model, results of COVID-19 pandemic spread are generated for the USA and India.

Resource Optimized Restriction Level result

Table 6.11 presents the optimized restriction levels for the United States, alongside the actual restrictions implemented by the country without accounting for the optimal use of healthcare resources. A similar analysis has been conducted for the COVID-19 pandemic in India. The results from both countries are subsequently used for forecasting COVID-19 case trends.

Forecasting results without resource-constraint

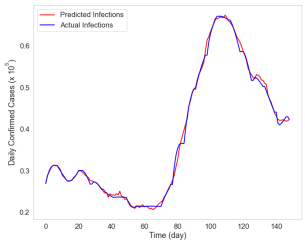
This section presents forecasting results without incorporating the optimal utilization of hospitalization capacity of a country. Figure 6.9 illustrates the forecasts for infected cases, hospitalizations, and COVID-19-related deaths for both India and the USA. The results demonstrate that the proposed model accurately predicts realistic trends in infections and fatalities for both countries. In terms of hospitalizations, the model suggests efficient utilization of hospital beds in the USA, while India shows relatively lower hospitalization numbers. Furthermore, the model reveals signs of overutilization of hospital resources in India.

Forecasting results with resource-constraint

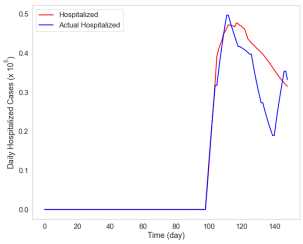
This section presents the forecasting results of the proposed model, incorporating the estimation of restriction parameters based on healthcare resource limitations. Research on COVID-19 indicates that complete lockdowns are not the most effective strategy for disease containment, as they primarily delay transmission rather than prevent it. A more practical approach involves the efficient management of medical resources until a vaccine becomes widely available. Accordingly, restriction levels should be determined in alignment with the capacity of healthcare facilities to avoid both overuse and underuse of hospital resources. Figure 6.10 displays the

Table 6.11: Optimized and direct restriction level for the USA dataset.

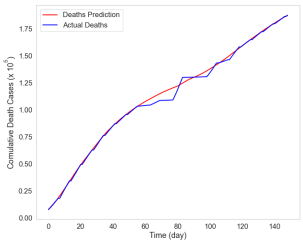
Weeks	USA	
	Optimized Restriction Level	Direct Restriction Level
0	0.00126	0.7
1	0.00146	0.65
2	0.00128	0.6
3	0.00141	0.55
4	0.00129	0.5
5	0.00117	0.45
6	0.00112	0.4
7	0.00109	0.35
8	0.00101	0.3
9	0.00098	0.25
10	0.00105	0.2
11	0.00129	0.15
12	0.00186	0.1
13	0.00234	0.1
14	0.00287	0.1
15	0.00328	0.1
16	0.00327	0.1
17	0.00298	0.1
18	0.00259	0.1
19	0.00252	0.1
20	0.00212	0.1
21	0.00205	0.1



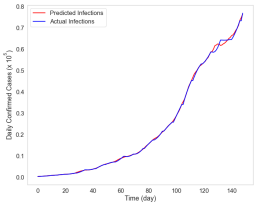
(a) Forecasting result of infected cases in USA.



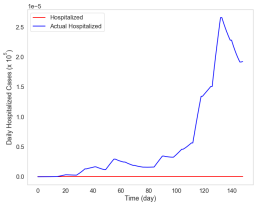
(b) Forecasting result of hospitalized cases in USA.



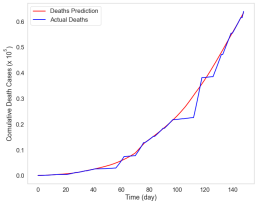
(c) Forecasting result of death cases in USA.



(d) Forecasting result of infected cases in India.



(e) Forecasting result of hospitalized cases in India.



(f) Forecasting result of death cases in India.

Figure 6.9: Forecasting results without resource-optimization in India and the USA

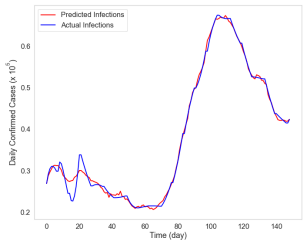
forecasted numbers of infections, hospitalizations, and deaths for India and the USA. The results show a marked improvement in forecasting accuracy compared to those generated without incorporating resource constraints. However, for India, the model still indicates an overutilization of healthcare resources in terms of hospitalizations. It shows need to increase the *hospital_beds_per_thousand* in India.

Forecasting Performance

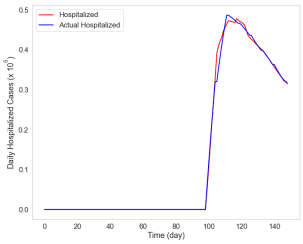
This section evaluates the performance of the proposed model using various statistical metrics. Table 6.12 summarizes the forecasting accuracy for infected cases in India and the USA, assessed using four metrics namely, MAE, RMSE, RRSE, and MAPE. The model demonstrates better performance on the USA dataset, exhibiting lower MAPE and improved RRSE values. Comparable results are observed for the forecasts of deaths and hospitalizations.

Table 6.12: Forecasting accuracy of the proposed model on infected cases using different metrics

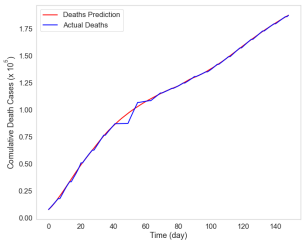
Dataset	MAE	RMSE	RRSE	MAPE
USA	23745	57227	1.24	14
India	27229	37784	1.85	18



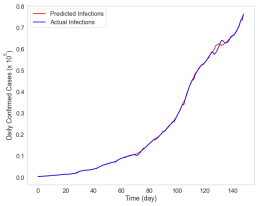
(a) Infected cases forecasting with restriction-optimization for USA.



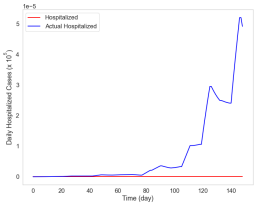
(b) Hospitalized cases forecasting with restriction-optimization for USA.



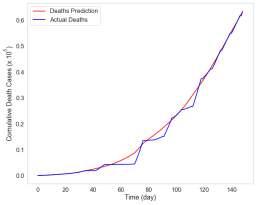
(c) Death cases forecasting with restriction-optimization for USA.



(d) Infected cases forecasting with restriction-optimization for India.



(e) Hospitalized cases forecasting with restriction-optimization for India.



(f) Death cases forecasting with restriction-optimization for India.

Figure 6.10: Forecasting of COVID-19 cases with restriction-optimization for India and the USA.

6.5 Analysis and Discussion

The proposed epidemiological model is designed to simulate the COVID-19 epidemic across different countries. It adapts to varying national contexts by learning from the evolution of an epidemic based on model parameters. The model can accommodate single-dose or multi-dose vaccination data, with the inclusion of booster dose compartments essential for long-term epidemics like COVID-19. Additionally, government-imposed restriction policies play a crucial role in controlling the spread of the epidemic, and such measures must be incorporated into the model to reflect real-world scenarios. To capture the impact of these measures, the model integrates a restriction parameter. The spread of an epidemic is influenced by both the restriction policies and the virus dynamics, making periodic estimation of the model parameters critical. In the proposed approach, the parameters are re-estimated for each event duration.

Although the model incorporates many real-world factors, achieving a perfect match with actual pandemic scenarios is challenging due to the large number of variables involved, such as virus dynamics, environmental conditions, healthcare infrastructure, government responses, education, human behavior, and regional factors. Model selection is another key aspect of epidemic evolution modeling. In this study, an epidemiological model is designed with ten compartments that account for vaccination trends, as well as protected and immunized-susceptible populations. To optimize the model parameters efficiently, PSO algorithm is adopted. The selection of an appropriate optimization algorithm is a critical

decision, and further research is needed to explore the use of other bio-inspired algorithms for optimization. To evaluate the model, case studies of two countries heavily affected by COVID-19 are conducted. These studies indicate that while increased restrictions can slow the spread of an epidemic, they cannot fully control it. Developing effective vaccines is one promising solution for eradicating large-scale pandemics like COVID-19.

It is important to note that evaluating a model based on a single dataset and performance metric may not provide a comprehensive understanding of its effectiveness. To address this, the proposed model has been assessed using datasets from both India and the USA and compared it with five state-of-the-art models across five different performance metrics. This comparative study has offered a broader perspective on the performance of a model. The results suggest that machine learning models are better at capturing the dynamic fluctuations of pandemic data, especially during multiple waves, something that is challenging for traditional epidemiological models. Therefore, combining epidemiological and machine learning models is beneficial to improve ability to learn from and adapt to evolving time series data. Additionally, periodic re-estimation of epidemiological parameters and the integration of relevant factors are essential to effectively model the dynamic nature of an epidemic.

Further, in this chapter a novel epidemiological compartmental model has been proposed that introduces hospitalization as a new compartment. Time-varying parameters are incorporated to account for dynamic conditions, and the model parameters are optimized using the PSO algorithm. A resource-constraint-based restriction parameter is integrated into the model to optimize hospital bed capacity within a country. Additionally, a multi-variable LSTM based deep learning approach is employed to train and forecast the model parameters. The proposed model is applied to COVID-19 time series data from India and the USA, effectively capturing the real-world progression of the disease. It highlights both over-utilization and under-utilization of hospital resources. Future research can further enhance the model by integrating alternative optimization techniques, deep learning approaches, and diverse datasets.

CHAPTER 7

CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT

Time series forecasting models are highly valuable for predicting future outcomes based on historical data. These models enable preventive measures to mitigate undesirable outcomes. Different types of time series data can be utilized for predictions, and in the context of epidemics, such predictions can be instrumental in combating diseases. A recent example is the COVID-19 pandemic, which claimed millions of lives and underscored the importance of analyzing and forecasting epidemic evolution. This thesis has examined widely used and well-known methodologies related to COVID-19 time series modeling and forecasting, and presented two new proposals 1) machine-learning based (FTS-PSO-attention-Bi-LSTM), and 2) novel epidemiological model incorporating restriction policies and vaccination (SEIRDPV) for accurate time series forecasting of COVID-19 confirmed, death, and recovered cases in a timeline spanning multiple COVID-19 waves attributed to multiple mutations of the SARS-CoV-2 virus.. Trends in infection cases, active cases, recoveries, deaths, and hospitalizations have been analyzed. An overview of SARS-CoV-2 variants and COVID-19 vaccines has been presented to provide insights into the dynamics of pandemic.

Research outcomes of this thesis are summarized in Table 7.1 and further elucidated below. In initial studies, state-of-the-art forecasting models, including ETS, ARIMA, FB-Prophet, ANN, LSTM, and its variants, are evaluated using COVID-19 datasets from highly affected countries such as the USA, Spain, Italy, France, Germany, Russia, Iran, the UK, Turkey, and India. The analysis begins with a comparison of ARIMA and FB-Prophet models based on error metrics such as MAE, RMSE, RRSE, and MAPE, where ARIMA outperforms FB-Prophet. In a subsequent study, ARIMA and LSTM models are compared using the USA dataset, demonstrating that while both adapt to the multi-wave nature of the pandemic, LSTM outperforms ARIMA in modeling long-term and short-term dependencies in non-linear time series data. Further, ETS, ANN, ARIMA, and LSTM models are evaluated across five datasets containing multiple COVID-19 waves, with ANN and LSTM models showing strong robustness and effectiveness in capturing temporal dependencies. The findings highlight the superior ability of the LSTM to handle multi-wave time series data. Given the dynamic nature of COVID-19, influenced by factors like lockdowns, variants, vaccination programs, and governmental policies, FTS models have been explored due to their effectiveness in handling uncertainties. Two novel optimization algorithms, nested-PSO and exhaustive-search-PSO, have been proposed to optimize FTS models by determining partition parameters and fuzzy orders. These algorithms have outperformed state-of-the-art FTS models, with the exhaustive-search-PSO-based FTS model achieving the best forecasting performance on evolving COVID-19 timelines.

Table 7.1: Summary of the research findings presented in this thesis.

Chapter	Proposed Methodology	Dataset Used	Comparison Methods	Performance Matrics	Novelties
Chapter 3	COVID-19 forecasting of infected, active, recovered, and death cases during initial phase of outbreak.	Time series data of USA, India, UK, Spain, Italy, France, Germany, Russia, Iran, and Turkey from March 1, 2020, to May 20, 2020	ARIMA and FB-Prophet	MAE, MAPE, RMSE, RRSE	ARIMA outperformed FB-Prophet on all datasets
Chapter 3	COVID-19 forecasting of confirmed cases having more than one wave.	Time series data of the USA from April 1, 2020, to March 31, 2021	ARIMA and LSTM	MAE, MAPE, RMSE, Correlation	LSTM outperformed ARIMA
Chapter 3	COVID-19 forecasting of confirmed cases having multiple waves.	Time series data of USA, India, UK, Russia, and Italy from June 1, 2020, to April 15, 2022	ETS, ANN, ARIMA and LSTM	MAE, RMSE, RRSE, sMAPE, MASE	LSTM outperformed other compared models.
Chapter 4	COVID-19 FTS forecasting using novel nested-FTS-PSO and exhaustive-search-FTS-PSO Methodologies for initial phase (Phase-1).	Time series data of USA, India, UK, Spain, Italy, France, Germany, Russia, Iran, and Turkey from April 2020 to October 2020	ARIMA, FB-Prophet, FTS, FTS-PSO, Nested-FTS-PSO, and Exhaustive-search-FTS-PSO	MSE, MAPE	Proposed Methodology exhaustive-search-FTS-PSO outperformed other compared models.
Chapter 4	COVID-19 FTS forecasting using novel nested-FTS-PSO and exhaustive-search-FTS-PSO Methodologies for evolved phase (Phase-2).	Time series data of USA and India from January 1, 2021, to May 15, 2021	ARIMA, FB-Prophet, FTS, FTS-PSO, Nested-FTS-PSO, and Exhaustive-search-FTS-PSO	MSE, MAPE	Proposed Methodology exhaustive-search-FTS-PSO outperformed other compared models.

Continued on next page

Table 7.1 – continued from previous page

Chapter	Proposed Methodology	Dataset Used	Comparison Methods	Performance Metrics	Novelties
Chapter 5	COVID-19 FTS forecasting using novel hybrid model using FTS, PSO, and deep learning techniques.	Time series data of USA, UK, India, Russia and Italy from June 1, 2020 to April 15, 2022 having multiple waves of outbreak	ETS, ANN, ARIMA, LSTM, FTSF-LSTM, FTS+PSO +LSTM, FTS+PSO +stacked-LSTM, FTS+PSO +Bi-LSTM, FTS+PSO +Conv-LSTM, FTS+PSO +Attention-LSTM, FTS+PSO +Attention-Bi-LSTM	MAE, RMSE, RRSE, sMAPE, MASE	Performance compared with or without data preprocessing where hybrid of FTS, PSO, and Attention-Bi-LSTM outperformed other compared models.
Chapter 5	Evolution of COVID-19 mutant affected population using proposed novel hybrid epidemiological model integrating SIRD, PSO, and stacked-LSTM.	Time series data with two major waves for the USA from July 1, 2021 to April 15, 2022, for India from January 1, 2021 to March 30, 2022, and for the UK from June 1, 2021 to April 30, 2022.	Stacked-LSTM, SIRD+PSO, Proposed (SIRD+PSO + stacked-LSTM)	RMSE	Proposed hybrid deep learning epidemiological model outperformed the compared models on all the datasets.

Continued on next page

Table 7.1 – continued from previous page

Chapter	Proposed Methodology	Dataset Used	Comparison Methods	Performance Metrics	Novelties
Chapter 6	Evolution of COVID-19 using novel 10-compartment SEIRDPV epidemiological model incorporating restriction parameter and multi-dose vaccination.	Time series data for confirmed, recovered, and death cases during Jan 16, 2021, to Jan 15, 2022 (India), and Dec 14, 2020, to Dec 13, 2021 (USA), corresponding to the first year of vaccination rollouts in these countries.	ETS, ARIMA, ANN, LSTM, SEIRD, Proposed-SEIRDPV	MAE, RMSE, RRSE, sMAPE, MASE	The proposed SEIRDPV effectively captures real-time dynamics if the hyperparameters are periodically re-estimated or updated based on specific events.
Chapter 6	Estimation of resource-optimized restriction parameter by adding hospitalizations compartment with SEIRD model to represent hospital admission capacity.	Time series data of the USA and India from April 1, 2020, to September 30, 2020 (evolving phase of the COVID-19 pandemic)	Actual hospitalized cases	MAE, RMSE, RRSE, MAPE	The proposed SEIHRD model investigates the optimal utilization of hospital beds by estimating a restriction parameter designed to keep the number of hospital admissions within the limits of available hospital beds.

To enhance forecasting accuracy in dynamic conditions, a hybrid FTS-deep learning model has been developed, integrating FTS with PSO and attention-Bi-LSTM. The proposed hybrid model consistently outperforms other models across datasets and performance metrics. Additionally, a hybrid model combining the SIRD epidemiological model, PSO optimization, and stacked-LSTM has been proposed to address the dynamic nature of epidemics. This model has been

evaluated on COVID-19 datasets from three countries which has outperformed all compared models.

Furthermore, a novel ten-compartment epidemiological model, SEIRDPV, is introduced for COVID-19 modeling. This model extends the traditional SEIRD model by incorporating multi-dose vaccinations, protected individuals, and immunized susceptibles. It adapts to country-specific vaccination programs by varying parameters for vaccine doses and efficacy, and it accounts for government restriction policies through a restriction parameter. The SEIRDPV model has closely matched actual pandemic trends in the USA and India and outperformed models like ETS, ARIMA, ANN, LSTM, and SEIRD in forecasting accuracy. The study concludes that while restrictions can delay disease spread, estimating time-varying parameters is essential for modeling dynamic epidemics effectively. Additionally, a study is carried for resource-optimization parameter estimation incorporating hospital beds capacity compartment into SEIRD model. The proposed model has been evaluated using COVID-19 cases from the USA and India. Results from the study reveal patterns of both over-utilization and under-utilization of hospital resources, demonstrating potential of the proposed model for informing data-driven healthcare policy and capacity planning.

This research has proposed several novel models for time series forecasting and epidemiological modeling. Time series forecasting models have the potential to drive changes in the society by improving decision-making and resource allocation across various sectors. In this view, the proposed models in this thesis can be used in fields such as economics, healthcare, and public policy, which can lead to significant positive societal outcomes. Future work could improve forecasting accuracy by incorporating variables such as population density, weather, healthcare systems, and patient history. The proposed models can also be extended to incorporate multimodal inputs (text, image, audio, sensor data) using advanced deep learning methods and evolutionary algorithms for analyzing the spread of other epidemics and pandemics.

BIBLIOGRAPHY

[1] Hossein Abbasimehr and Reza Paki. Prediction of covid-19 confirmed cases combining deep learning methods and bayesian optimization. *Chaos, Solitons & Fractals*, 142:110511, 2021.

[2] Hossein Abbasimehr and Reza Paki. Improving time series forecasting using lstm and attention models. *Journal of Ambient Intelligence and Humanized Computing*, 13(1):673–691, 2022.

[3] Alaa Abd-Alrazaq, Asmaa Hassan, Israa Abuelezz, Arfan Ahmed, Mahmood Saleh Alzubaidi, Uzair Shah, Dari Alhuwail, Anna Giannicchi, Mowafa Househ, et al. Overview of technologies implemented during the first wave of the covid-19 pandemic: Scoping review. *Journal of Medical Internet Research*, 23(9):e29136, 2021.

[4] Rasheed Omobolaji Alabi, Akpojoto Siemuri, and Mohammed Elmusrati. Covid-19: Easing the coronavirus lockdowns with caution. *medRxiv*, 2020.

[5] Talha Burak Alakus and Ibrahim Turkoglu. Comparison of deep learning approaches to predict covid-19 infection. *Chaos, Solitons & Fractals*, 140:110120, 2020.

[6] Madini O Alassafi, Mutasem Jarrah, and Reem Alotaibi. Time series predicting of covid-19 based on deep learning. *Neurocomputing*, 468:335–344, 2022.

[7] Huda M Alshanbari, Hasnain Iftikhar, Faridoon Khan, Moeeba Rind, Zubair Ahmad, and Abd Al-Aziz Hosni El-Bagoury. On the implementation of the artificial neural network approach for forecasting different healthcare events. *Diagnostics*, 13(7):1310, 2023.

[8] Yousif Alyousifi, Mahmod Othman, Ibrahima Faye, Rajalingam Sokkalingam, and Petronio CL Silva. Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution. *International Journal of Fuzzy Systems*, 22:1468–1486, 2020.

[9] AJIDM Anirudh. Mathematical modeling and the transmission dynamics in predicting the covid-19-what next in combating the pandemic. *Infectious Disease Modelling*, 5:366–374, 2020.

[10] Melissa M Arons, Kelly M Hatfield, Sujan C Reddy, Anne Kimball, Allison James, Jesica R Jacobs, Joanne Taylor, Kevin Spicer, Ana C Bardossy, Lisa P Oakley, et al. Presymptomatic sars-cov-2 infections and transmission in a skilled nursing facility. *New England journal of medicine*, 382(22):2081–2090, 2020.

Bibliography

[11] Parul Arora, Himanshu Kumar, and Bijaya Ketan Panigrahi. Prediction and analysis of covid-19 positive cases using deep learning models: A descriptive case study of india. *Chaos, Solitons & Fractals*, 139:110017, 2020.

[12] Sankalap Arora and Satvir Singh. Butterfly optimization algorithm: a novel approach for global optimization. *Soft computing*, 23:715–734, 2019.

[13] Edilson F Arruda, Marcelo D Fragoso, João BR do Val, Sinnu S Thomas, et al. A novel queue-based stochastic epidemic model with adaptive stabilising control. *ISA transactions*, 2023.

[14] KE ArunKumar, Dinesh V Kalaga, Ch Mohan Sai Kumar, Govinda Chilkoor, Masahiro Kawaji, and Timothy M Brenza. 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, 2021.

[15] KE ArunKumar, Dinesh V Kalaga, Ch Mohan Sai Kumar, Masahiro Kawaji, and Timothy M Brenza. Comparative analysis of gated recurrent units (gru), long short-term memory (lstm) cells, autoregressive integrated moving average (arima), seasonal autoregressive integrated moving average (sarima) for forecasting covid-19 trends. *Alexandria Engineering Journal*, 61(10):7585–7603, 2022.

[16] Peter Ashcroft, Sonja Lehtinen, Daniel C Angst, Nicola Low, and Sebastian Bonhoeffer. Quantifying the impact of quarantine duration on covid-19 transmission. *Elife*, 10:e63704, 2021.

[17] Md Abdul Awal, Mehedi Masud, Md Shahadat Hossain, Abdullah Al-Mamun Bulbul, SM Hasan Mahmud, and Anupam Kumar Bairagi. A novel bayesian optimization-based machine learning framework for covid-19 detection from inpatient facility data. *Ieee Access*, 9:10263–10281, 2021.

[18] Dzmitry Bahdanau. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[19] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[20] Samit Bhanja, Banani Ghose, and Abhishek Das. Multi-step-ahead time series forecasting using deep learning and fuzzy time series-based error correction method. *Journal of Universal Computer Science*, 30(11):1569, 2024.

[21] Tarun Bhatnagar, Sirshendu Chaudhuri, Manickam Ponnaiah, Pragya D Yadav, R Sabarinathan, Rima R Sahay, Faheem Ahmed, S Aswathy, Pankaj Bhardwaj, Anil Bilimale, et al. Effectiveness of bbv152/covaxin and

azd1222/covishield vaccines against severe covid-19 and b. 1.617. 2/delta variant in india, 2021: a multi-centric hospital-based case-control study. *International Journal of Infectious Diseases*, 122:693–702, 2022.

[22] Mohamed Aziz Bhouri, Francisco Sahli Costabal, Hanwen Wang, Kevin Linka, Mathias Peirlinck, Ellen Kuhl, and Paris Perdikaris. Covid-19 dynamics across the us: A deep learning study of human mobility and social behavior. *Computer Methods in Applied Mechanics and Engineering*, 382:113891, 2021.

[23] Alessandro Borri, Pasquale Palumbo, Federico Papa, and Corrado Possieri. Optimizing restrictions in epidemics via piecewise time-varying sird models: Application to the covid-19 italian emergency. *European Journal of Control*, 75:100902, 2024.

[24] Mahua Bose and Kalyani Mali. Designing fuzzy time series forecasting models: A survey. *International Journal of Approximate Reasoning*, 111:78–99, 2019.

[25] Arthur Bousquet, William H Conrad, Said Omer Sadat, Nelli Vardanyan, and Youngjoon Hong. Deep learning forecasting using time-varying parameters of the sird model for covid-19. *Scientific Reports*, 12(1):3030, 2022.

[26] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[27] Álvaro Briz-Redón and Ángel Serrano-Aroca. A spatio-temporal analysis for exploring the effect of temperature on covid-19 early evolution in spain. *Science of the total environment*, 728:138811, 2020.

[28] Ellen Brooks-Pollock, Jonathan M Read, Angela R McLean, Matt J Keeling, and Leon Danon. Mapping social distancing measures to the reproduction number for covid-19. *Philosophical Transactions of the Royal Society B*, 376(1829):20200276, 2021.

[29] Qisen Cai, Defu Zhang, Wei Zheng, and Stephen CH Leung. A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowledge-Based Systems*, 74:61–68, 2015.

[30] Jamie M Caldwell, Xuan Le, Lorin McIntosh, Michael T Meehan, Samson Ogunlade, Romain Ragonnet, Genevieve K O'Neill, James M Trauer, and Emma S McBryde. Vaccines and variants: Modelling insights into emerging issues in covid-19 epidemiology. *Paediatric Respiratory Reviews*, 39:32–39, 2021.

[31] Nicolò Cangiotti, Marco Capolli, Mattia Sensi, and Sara Sottile. A survey on lyapunov functions for epidemic compartmental models. *Bollettino dell'Unione Matematica Italiana*, pages 1–17, 2023.

57

Bibliography

[32] Alessandro M Carabelli, Thomas P Peacock, Lucy G Thorne, William T Harvey, Joseph Hughes, Sharon J Peacock, Wendy S Barclay, Thushan I De Silva, Greg J Towers, and David L Robertson. Sars-cov-2 variant biology: immune escape, transmission and fitness. *Nature Reviews Microbiology*, 21(3):162–177, 2023.

[33] Oscar Castillo, Juan R Castro, Martha Pulido, and Patricia Melin. Interval type-3 fuzzy aggregators for ensembles of neural networks in covid-19 time series prediction. *Engineering Applications of Artificial Intelligence*, 114:105110, 2022.

[34] Vitor Cerqueira, Luis Torgo, and Igor Mozetič. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 109(11):1997–2028, 2020.

[35] Meric Cetin and Selami Beyhan. Vaccination, lock-down, measures and time-varying reproduction number based estimation and control of covid-19 outbreak in turkey. In *Artificial Intelligence and Machine Learning Methods in COVID-19 and Related Health Diseases*, pages 139–159. Springer, 2022.

[36] Suman Chakraborti, Arabinda Maiti, Suvamoy Pramanik, Srikanta Sannigrahi, Francesco Pilla, Anushna Banerjee, and Dipendra Nath Das. Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific covid-19 analysis. *Science of the Total Environment*, 765:142723, 2021.

[37] Robert Challen, Ellen Brooks-Pollock, Jonathan M Read, Louise Dyson, Krasimira Tsaneva-Atanasova, and Leon Danon. Risk of mortality in patients infected with sars-cov-2 variant of concern 202012/1: matched cohort study. *bmj*, 372, 2021.

[38] Thing-Yuan Chang, Cheng-Kui Huang, Cheng-Hsiung Weng, and Jing-Yuan Chen. Feature-based deep neural network approach for predicting mortality risk in patients with covid-19. *Engineering Applications of Artificial Intelligence*, 124:106644, 2023.

[39] Ray-Ming Chen. Track the dynamical features for mutant variants of covid-19 in the uk. *Mathematical Biosciences and Engineering*, 18(4):4572–4585, 2021.

[40] Shyi-Ming Chen. Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems*, 81(3):311–319, 1996.

[41] Shyi-Ming Chen. Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems*, 33(1):1–16, 2002.

[42] Shyi-Ming Chen, Xin-Yao Zou, and Gracius Cagar Gunawan. Fuzzy time series forecasting based on proportions of intervals and particle swarm optimization techniques. *Information Sciences*, 500:127–139, 2019.

[43] Yixiang Chen, Min Chen, Bo Huang, Chao Wu, and Wenjia Shi. Modeling the spatiotemporal association between covid-19 transmission and population mobility using geographically and temporally weighted regression. *GeoHealth*, 5(5):e2021GH000402, 2021.

[44] Zhongxiang Chen, Jun Yang, and Binxiang Dai. Forecast possible risk for covid-19 epidemic dissemination under current control strategies in japan. *International Journal of Environmental Research and Public Health*, 17(11):3872, 2020.

[45] Vinay Kumar Reddy Chimmula and Lei Zhang. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, 135:109864, 2020.

[46] Nalini Chintalapudi, Gopi Battineni, and Francesco Amenta. Covid-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*, 2020.

[47] Rajiv Chowdhury, Kevin Heng, Md Shajedur Rahman Shawon, Gabriel Goh, Daisy Okonofua, Carolina Ochoa-Rosales, Valentina Gonzalez-Jaramillo, Abbas Bhuiya, Daniel Reidpath, Shamini Prathapan, et al. Dynamic interventions to control covid-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. *European journal of epidemiology*, 35:389–399, 2020.

[48] Rangan Das, Sagnik Sen, and Ujjwal Maulik. A survey on fuzzy deep neural networks. *ACM Computing Surveys (CSUR)*, 53(3):1–25, 2020.

[49] Ankan Ghosh Dastider, Farhan Sadik, and Shaikh Anowarul Fattah. An integrated autoencoder-based hybrid cnn-lstm model for covid-19 severity prediction from lung ultrasound. *Computers in Biology and Medicine*, 132:104296, 2021.

[50] Orhun O Davarci, Emily Y Yang, Alexander Viguerie, Thomas E Yankeelov, and Guillermo Lorenzo. Dynamic parameterization of a modified seird model to analyze and forecast the dynamics of covid-19 outbreaks in the united states. *Engineering with Computers*, 40(2):813–837, 2024.

[51] Ugo Avila-Ponce de León, Ángel GC Pérez, and Eric Avila-Vales. An seiard epidemic model for covid-19 in mexico: mathematical analysis and state-level forecast. *Chaos, Solitons & Fractals*, 140:110165, 2020.

[52] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7:1–30, 2006.

[53] Sarah Schaffer DeRoo, Natalie J Pudalov, and Linda Y Fu. Planning for a covid-19 vaccination program. *Jama*, 323(24):2458–2459, 2020.

Bibliography

[54] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[55] Caichang Ding, Yiqin Chen, Zhiyuan Liu, and Tianyin Liu. Prediction on transmission trajectory of covid-19 based on particle swarm algorithm. *Pattern Recognition Letters*, 152:70–78, 2021.

[56] Weiping Ding, Mohamed Abdel-Basset, Hossam Hawash, and Ahmed M Ali. Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey. *Information Sciences*, 615:238–292, 2022.

[57] Stefania Dispinseri, Massimiliano Secchi, Maria Franca Pirillo, Monica Tolazzi, Martina Borghi, Cristina Brigatti, Maria Laura De Angelis, Marco Baratella, Elena Bazzigaluppi, Giulietta Venturi, et al. Neutralizing antibody responses to sars-cov-2 in symptomatic covid-19 is persistent and critical for survival. *Nature communications*, 12(1):2670, 2021.

[58] Ankit Dixit and Shikha Jain. Contemporary approaches to analyze non-stationary time-series: Some solutions and challenges. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 16(2):61–80, 2023.

[59] Jacqueline Duhon, Nicola Bragazzi, and Jude Dzevela Kong. The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of covid-19: A cross-country study. *Science of The Total Environment*, 760:144325, 2021.

[60] Raches Ella, Siddarth Reddy, William Blackwelder, Varsha Potdar, Pragya Yadav, Vamshi Sarangi, Vinay K Aileni, Suman Kanungo, Sanjay Rai, Prabhakar Reddy, et al. Efficacy, safety, and lot-to-lot immunogenicity of an inactivated sars-cov-2 vaccine (bbv152): interim results of a randomised, double-blind, controlled, phase 3 trial. *The Lancet*, 398(10317):2173–2184, 2021.

[61] Duccio Fanelli and Francesco Piazza. Analysis and forecast of covid-19 spreading in china, italy and france. *Chaos, Solitons & Fractals*, 134:109761, 2020.

[62] Chellai Fatih, Ahmed Hamimes, and Pradeep Mishra. A note on covid-19 statistics, strange trend and forecasting of total cases in the most infected african countries: An arima and fuzzy time series approaches. *African Journal of Applied Statistics*, 2(2):961–975, 2020.

[63] Filipe Fernandes, Stéfano Frizzo Stefenon, Laio Oriel Seman, Ademir Nied, Fernanda Cristina Silva Ferreira, Maria Cristina Mazzetti Subtil, Anne Carolina Rodrigues Klaar, and Valderi Reis Quietinho Leithardt. Long short-term memory stacking model to predict the number of cases and deaths

caused by covid-19. *Journal of Intelligent & Fuzzy Systems*, 42(6):6221–6234, 2022.

[64] Santiago Fernández, Alex Graves, and Jürgen Schmidhuber. Sequence labelling in structured domains with hierarchical recurrent neural networks. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI 2007*, 2007.

[65] Jesús Fernández-Villaverde and Charles I Jones. Estimating and simulating a sird model of covid-19 for many countries, states, and cities. *Journal of Economic Dynamics and Control*, 140:104318, 2022.

[66] Luisa Ferrari, Giuseppe Gerardi, Giancarlo Manzi, Alessandra Micheletti, Federica Nicolussi, Elia Biganzoli, and Silvia Salini. Modeling provincial covid-19 epidemic data using an adjusted time-dependent sird model. *International Journal of Environmental Research and Public Health*, 18(12):6563, 2021.

[67] Gentile Francesco Ficetola and Diego Rubolini. Containment measures limit environmental effects on covid-19 early outbreak dynamics. *Science of the Total Environment*, 761:144432, 2021.

[68] Milton Friedman. A comparison of alternative tests of significance for the problem of m rankings. *The annals of mathematical statistics*, 11(1):86–92, 1940.

[69] Yuting Fu, Hanqing Jin, Haitao Xiang, and Ning Wang. Optimal lockdown policy for vaccination during covid-19 pandemic. *Finance research letters*, 45:102123, 2022.

[70] Mudasir A Ganaie, Minghui Hu, Ashwani Kumar Malik, Muhammad Tanveer, and Ponnuthurai N Suganthan. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, 2022.

[71] Seyyed Mostafa Ghadami, Seyyed Masoud Seyyedi, and Rituraj Rituraj. Nonlinear control of covid-19 pandemic based on the sird model. In *2022 IEEE 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, pages 000449–000456. IEEE, 2022.

[72] Ramy Mohamed Ghazy, Rasha Ashmawy, Noha Alaa Hamdy, Yasir Ahmed Mohammed Elhadi, Omar Ahmed Reyad, Dina Elmalawany, Abdallah Almaghraby, Ramy Shaaban, and Sarah Hamed N Taha. Efficacy and effectiveness of sars-cov-2 vaccines: a systematic review and meta-analysis. *Vaccines*, 10(3):350, 2022.

[73] Antonio Gómez-Corral, Maria Jesus Lopez-Herrero, and D Taipe. A markovian epidemic model in a resource-limited environment. *Applied Mathematics and Computation*, 458:128252, 2023.

Bibliography

[74] Wenping Gong, Seppo Parkkila, Xueqiong Wu, and Ashok Aspatwar. Sars-cov-2 variants and covid-19 vaccines: Current challenges and future strategies. *International reviews of immunology*, 42(6):393–414, 2023.

[75] Leon Gordis. *Epidemiology e-book*. Elsevier Health Sciences, 2013.

[76] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.

[77] Xiaojing Guo, Hui Zhang, Luyao Kou, and Yufan Hou. Modeling the external, internal, and multi-center transmission of infectious diseases: the covid-19 case. *Journal of Social Computing*, 3(2):171–181, 2022.

[78] Michael E Habicht, F Donald Pate, Elena Varotto, and Francesco M Galassi. Epidemics and pandemics in the history of humankind and how governments dealt with them a review from the bronze age to the early modern age. *Rivista Trimestrale Di Scienza Dell'Amministrazione*, 2, 2020.

[79] Ikbel Hadj Hassine. Covid-19 vaccines and variants of concern: a review. *Reviews in medical virology*, 32(4):e2313, 2022.

[80] Milad Haghani, Michiel CJ Bliemer, Floris Goerlandt, and Jie Li. The scientific literature on coronaviruses, covid-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. *Safety Science*, page 104806, 2020.

[81] Yan Hao, Ting Xu, Hongping Hu, Peng Wang, and Yanping Bai. Prediction and analysis of corona virus disease 2019. *PloS one*, 15(10):e0239960, 2020.

[82] Mohammad Mehedi Hassan, Mabrook S AlRakhami, Ahmed Zohier Elhendi, and Salman A AlQahtani. A distributed ensemble of diverse deep learning models for predicting covid-19 cases. In *IEEE EUROCON 2023-20th International Conference on Smart Technologies*, pages 585–589. IEEE, 2023.

[83] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.

[84] Arash Heidari, Nima Jafari Navimipour, Mehmet Unal, and Shiva Toumaj. Machine learning applications for covid-19 outbreak management. *Neural Computing and Applications*, 34(18):15313–15348, 2022.

[85] Andres Hernandez-Matamoros, Hamido Fujita, Toshitaka Hayashi, and Hector Perez-Meana. Forecasting of covid19 per regions using arima models and polynomial functions. *Applied soft computing*, 96:106610, 2020.

[86] The Hindu. Worldś largest vaccination programme begins in india [online]. available: https://www.thehindu.com/news/national/coronavirus-worlds-largest-vaccination-programme-begins-in-india-on-january-16/article33582069.ece, 15 January 2021.

[87] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[88] Jun Hu and Wendong Zheng. A deep learning model to effectively capture mutation information in multivariate time series prediction. *Knowledge-Based Systems*, 203:106139, 2020.

[89] Li Huang, Shuai Ding, Shouhao Yu, Juan Wang, and Ke Lu. Chaos-enhanced cuckoo search optimization algorithms for global optimization. *Applied Mathematical Modelling*, 40(5-6):3860–3875, 2016.

[90] Kunhuang Huarng. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy sets and systems*, 123(3):387–394, 2001.

[91] Katherine Huerne, Kristian B Filion, Roland Grad, Pierre Ernst, Andrea S Gershon, and Mark J Eisenberg. Epidemiological and clinical perspectives of long covid syndrome. *American Journal of Medicine Open*, 9:100033, 2023.

[92] Our World in Data. Covid-19 dataset. [online]. available: https://ourworldindata.org/covid-vaccinations.

[93] Our World in Data. Covid-19 dataset. [online]. https://github.com/owid/covid-19-data.

[94] Github Inc. Covid-19 cases. https://github.com/cssegisanddata/covid-19.

[95] Vito Janko, Nina Reščič, Aljoša Vodopija, David Susič, Carlo De Masi, Tea Tušar, Anton Gradišek, Sophie Vandepitte, Delphine De Smedt, Jana Javornik, et al. Optimizing non-pharmaceutical intervention strategies against covid-19 using artificial intelligence. *Frontiers in public health*, 11:1073581, 2023.

[96] Nadia Jebril. World health organization declared a pandemic public health menace: A systematic review of the coronavirus disease 2019 “covid-19”, up to 26th march 2020. *Available at SSRN 3566298*, 2020.

[97] Chia-Feng Juang, Ren-Bo Huang, and Wei-Yuan Cheng. An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems. *IEEE Transactions on fuzzy systems*, 18(4):686–699, 2010.

[98] Leonid Kalachev, Erin L Landguth, and Jon Graham. Revisiting classical sir modelling in light of the covid-19 pandemic. *Infectious Disease Modelling*, 8(1):72–83, 2023.

Bibliography

[99] Md Kamrujjaman, Pritam Saha, Md Shahidul Islam, and Uttam Ghosh. Dynamics of seir model: A case study of covid-19 in italy. *Results in Control and Optimization*, 7:100119, 2022.

[100] Perumandla Karunakar, K Shiva Reddy, and Snehashish Chakraverty. Stability analysis and approximate solution of interval mathematical model for the covid-19 pandemic. *Mathematical Methods in the Applied Sciences*, 2023.

[101] Nataliia Kashpruk, Cezary Piskor-Ignatowicz, and Jerzy Baranowski. Time series prediction in industry 4.0: a comprehensive review and prospects for future advancements. *Applied Sciences*, 13(22):12374, 2023.

[102] James Kennedy. Swarm intelligence. In *Handbook of nature-inspired and innovative computing*, pages 187–219. Springer, 2006.

[103] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995.

[104] William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics–i. 1927. *Bulletin of mathematical biology*, 53(1-2):33–55, 1991.

[105] Muhamad Khairulbahri. The seir model incorporating asymptomatic cases, behavioral measures, and lockdowns: Lesson learned from the covid-19 flow in sweden. *Biomedical Signal Processing and Control*, 81:104416, 2023.

[106] Sanaa L Khalaf and Hadeer S Flayyih. Analysis, predicting, and controlling the covid-19 pandemic in iraq through sir model. *Results in Control and Optimization*, 10:100214, 2023.

[107] Sultan Daud Khan, Louai Alarabi, and Saleh Basalamah. Toward smart lockdown: a novel approach for covid-19 hotspots prediction using a deep hybrid neural network. *Computers*, 9(4):99, 2020.

[108] Wasiq Khan, Abir Hussain, Sohail Ahmed Khan, Mohammed Al-Jumailey, Raheel Nawaz, and Panos Liatsis. Analysing the impact of global demographic characteristics over the covid-19 spread using class rule mining and pattern matching. *Royal Society open science*, 8(1):201823, 2021.

[109] Cem Kocak, Erol Egrioglu, and Eren Bas. A new deep intuitionistic fuzzy time series forecasting method based on long short-term memory. *The Journal of Supercomputing*, 77:6178–6196, 2021.

[110] Archana Koirala, Ye Jin Joo, Ameneh Khatami, Clayton Chiu, and Philip N Britton. Vaccines for covid-19: The current state of play. *Paediatric respiratory reviews*, 35:43–49, 2020.

[111] Liang Kong, Yanhui Guo, and Chung-wei Lee. Enhancing covid-19 prevalence forecasting: A hybrid approach integrating epidemic differential equations and recurrent neural networks. *AppliedMath*, 4(2):427–441, 2024.

[112] Lingcai Kong, Mengwei Duan, Jin Shi, Jie Hong, Zhaorui Chang, and Zhijie Zhang. Compartmental structures used in modeling covid-19: a scoping review. *Infectious diseases of poverty*, 11(1):72, 2022.

[113] Gourav Kumar, Sanjeev Jain, and Uday Pratap Singh. Stock market forecasting using computational intelligence: A survey. *Archives of computational methods in engineering*, 28(3):1069–1101, 2021.

[114] Naresh Kumar and Seba Susan. Covid-19 pandemic prediction using time series forecasting models. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–7. IEEE, 2020.

[115] Naresh Kumar and Seba Susan. Particle swarm optimization of partitions and fuzzy order for fuzzy time series forecasting of covid-19. *Applied Soft Computing*, 110:107611, 2021.

[116] Naresh Kumar and Seba Susan. Epidemic modeling using hybrid of time-varying sird, particle swarm optimization, and deep learning. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7. IEEE, 2023.

[117] Rajani Kumari, Sandeep Kumar, Ramesh Chandra Poonia, Vijander Singh, Linesh Raja, Vaibhav Bhatnagar, and Pankaj Agarwal. Analysis and predictions of spread, recovery, and death caused by covid-19 in india. *Big Data Mining and Analytics*, 4(2):65–75, 2021.

[118] Cheng-Pin Kuo and Joshua S Fu. Evaluating the impact of mobility on covid-19 pandemic with machine learning hybrid predictions. *Science of The Total Environment*, 758:144151, 2021.

[119] I-Hong Kuo, Shi-Jinn Horng, Yuan-Hsin Chen, Ray-Shine Run, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, and Tsung-Lieh Lin. Forecasting taifex based on fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 37(2):1494–1502, 2010.

[120] I-Hong Kuo, Shi-Jinn Horng, Tzong-Wann Kao, Tsung-Lieh Lin, Cheng-Ling Lee, and Yi Pan. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert Systems with applications*, 36(3):6108–6117, 2009.

[121] Samuel Lalmuanawma, Jamal Hussain, and Lalrinfela Chhakchhuak. Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review. *Chaos, Solitons & Fractals*, page 110059, 2020.

Bibliography

[122] Jeffrey V Lazarus, Scott C Ratzan, Adam Palayew, Lawrence O Gostin, Heidi J Larson, Kenneth Rabin, Spencer Kimball, and Ayman El-Mohandes. A global survey of potential acceptance of a covid-19 vaccine. *Nature medicine*, 27(2):225–228, 2021.

[123] Hyojung Lee, Sol Kim, Minyoung Jeong, Eunseo Choi, Hyeonjeong Ahn, and Jeehyun Lee. Mathematical modeling of covid-19 transmission and intervention in south korea: A review of literature. *Yonsei medical journal*, 64(1):1, 2023.

[124] Li-Wei Lee, Li-Hui Wang, and Shyi-Ming Chen. Temperature prediction and taifex forecasting based on fuzzy logical relationships and genetic algorithms. *Expert Systems with Applications*, 33(3):539–550, 2007.

[125] Li-Wei Lee, Li-Hui Wang, and Shyi-Ming Chen. Temperature prediction and taifex forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques. *Expert Systems with Applications*, 34(1):328–336, 2008.

[126] Mariana Lenharo. Who declares end to covid-19's emergency phase. *Nature*, 882(10.1038), 2023.

[127] SP Levashkin, OI Zakharova, SV Kuleshov, and AA Zaytseva. Adaptive-compartmental model of coronavirus epidemic and its optimization by the methods of artificial intelligence. In *Journal of Physics: Conference Series*, volume 1864, page 012108. IOP Publishing, 2021.

[128] Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):20200209, 2021.

[129] Felipe Tomazelli Lima and Vinicius MA Souza. A large comparison of normalization methods on time series. *Big Data Research*, 34:100407, 2023.

[130] Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, and Michael Weyrich. A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, 99:650–655, 2021.

[131] Chenghao Liu, Steven CH Hoi, Peilin Zhao, and Jianling Sun. Online arima algorithms for time series prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[132] Leonardo López and Xavier Rodo. A modified seir model to predict the covid-19 outbreak in spain and italy: simulating control scenarios and multi-scale epidemics. *Results in Physics*, 21:103746, 2021.

[133] Chao Luo, Chenhao Tan, Xingyuan Wang, and Yuanjie Zheng. An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction. *Applied Soft Computing*, 78:150–163, 2019.

[134] Shiyang Lyu, Oyelola Adegboye, Kiki Adhinugraha, Theophilus I Emeto, and David Taniar. Unlocking insights: Analysing covid-19 lockdown policies and mobility data in victoria, australia, through a data-driven machine learning approach. *Data*, 9(1):3, 2023.

[135] Xin Ma, Xue Zhao, and Pengfei Guo. Cope with the covid-19 pandemic: Dynamic bed allocation and patient subsidization in a public healthcare system. *International Journal of Production Economics*, 243:108320, 2022.

[136] Ahmed Maged, Abdullah Ahmed, Salah Haridy, Arthur W Baker, and Min Xie. Seir model to address the impact of face masks amid covid-19 pandemic. *Risk Analysis*, 43(1):129–143, 2023.

[137] Ganapathy Mahalakshmi, S Sridevi, and Shyamsundar Rajaram. A survey on forecasting of time series data. In *2016 international conference on computing technologies and intelligent data engineering (ICCTIDE'16)*, pages 1–8. IEEE, 2016.

[138] Parikshit N Mahalle, Nilesh P Sable, Namita P Mahalle, and Gitanjali R Shinde. Data analytics: Covid-19 prediction using multimodal data. *Intelligent systems and methods to combat Covid-19*, pages 1–10, 2020.

[139] Sharif Naser Makhadmeh, Mohammed Azmi Al-Betar, Ammar Kamal Abasi, Mohammed A Awadallah, Iyad Abu Doush, Zaid Abdi Alkareem Alyasseri, and Osama Ahmad Alomari. Recent advances in butterfly optimization algorithm, its versions and applications. *Archives of Computational Methods in Engineering*, 30(2):1399–1420, 2023.

[140] Anvita Gupta Malhotra, Pranjali Borkar, Rashmi Chowdhary, and Sarman Singh. Combating covid-19 by employing machine learning predictions and projections. In *Advanced Methods in Biomedical Signal Processing and Analysis*, pages 175–203. Elsevier, 2023.

[141] Manotosh Mandal, Soovoojeet Jana, Swapan Kumar Nandi, Anupam Khatua, Sayani Adak, and TK Kar. A model based study on the dynamics of covid-19: Prediction and control. *Chaos, Solitons & Fractals*, page 109889, 2020.

[142] Pavodi Maniamfu and Keisuke Kameyama. Lstm-based forecasting using policy stringency and time-varying parameters of the sir model for covid-19. In *2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, pages 111–116. IEEE, 2023.

[143] Douglas Martins, Amit Bhaya, and Fernando Pazos. N-step-ahead optimal control of a compartmental model of covid-19. *Journal of Control, Automation and Electrical Systems*, 34(3):455–469, 2023.

Bibliography

[144] Ricardo P Masini, Marcelo C Medeiros, and Eduardo F Mendes. Machine learning advances for time series forecasting. *Journal of economic surveys*, 37(1):76–111, 2023.

[145] Gemma Massonis, Julio R Banga, and Alejandro F Villaverde. Structural identifiability and observability of compartmental models of the covid-19 pandemic. *Annual reviews in control*, 51:441–459, 2021.

[146] Patricia Melin, Daniela Sánchez, Juan R Castro, and Oscar Castillo. Design of type-3 fuzzy systems and ensemble neural networks for covid-19 time series prediction using a firefly algorithm. *Axioms*, 11(8):410, 2022.

[147] Risto Miikkulainen, Olivier Francon, Elliot Meyerson, Xin Qiu, Darren Sargent, Elisa Canzani, and Babak Hodjat. From prediction to prescription: evolutionary optimization of nonpharmaceutical interventions in the covid-19 pandemic. *IEEE Transactions on Evolutionary Computation*, 25(2):386–401, 2021.

[148] Luis Miralles-Pechuán, Ankit Kumar, and Andrés L Suárez-Cetrulo. Forecasting covid-19 cases using dynamic time warping and incremental machine learning methods. *Expert Systems*, page e13237, 2023.

[149] Seyedali Mirjalili. The ant lion optimizer. *Advances in engineering software*, 83:80–98, 2015.

[150] Seyedali Mirjalili. Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural computing and applications*, 27:1053–1073, 2016.

[151] Seyedali Mirjalili and Andrew Lewis. The whale optimization algorithm. *Advances in engineering software*, 95:51–67, 2016.

[152] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in engineering software*, 69:46–61, 2014.

[153] Abolfazl Mollalo, Behzad Vahedi, and Kiara M Rivera. Gis-based spatial modeling of covid-19 incidence rate in the continental united states. *Science of the total environment*, 728:138884, 2020.

[154] Kensaku Murano, Youjia Guo, and Haruhiko Siomi. The emergence of sars-cov-2 variants threatens to decrease the efficacy of neutralizing antibodies and vaccines. *Biochemical Society Transactions*, 49(6):2879–2890, 2021.

[155] Khondoker Nazmoon Nabi, Md Toki Tahmid, Abdur Rafi, Muhammad Ehsanul Kader, and Md Asif Haider. Forecasting covid-19 cases: A comparative analysis between recurrent and convolutional neural networks. *Results in Physics*, 24:104137, 2021.

[156] Zhang Nannan and Luo Chao. Adaptive online time series prediction based on a novel dynamic fuzzy cognitive map. *Journal of Intelligent & Fuzzy Systems*, 36(6):5291–5303, 2019.

[157] Hafiz Muhammad Naveed, Yao HongXing, Bilal Ahmed Memon, Shoaib Ali, Mohammed Ismail Alhussam, and Jan Muhammad Sohu. Artificial neural network (ann)-based estimation of the influence of covid-19 pandemic on dynamic and emerging financial markets. *Technological Forecasting and Social Change*, 190:122470, 2023.

[158] Peter Bjorn Nemenyi. *Distribution-free multiple comparisons.* Princeton University, 1963.

[159] ABC News. Johnson & johnson single-shot vaccine 85% effective against severe covid-19 disease, June 29, 2021.

[160] Xiao Ning, Linlin Jia, Yongyue Wei, Xi-An Li, and Feng Chen. Epi-dnns: Epidemiological priors informed deep neural networks for modeling covid-19 dynamics. *Computers in biology and medicine*, 158:106693, 2023.

[161] Ruiwu Niu, Yin-Chi Chan, Eric WM Wong, Michaël Antonie van Wyk, and Guanrong Chen. A stochastic seihr model for covid-19 data fluctuations. *Nonlinear dynamics*, 106:1311–1323, 2021.

[162] Alberto Olivares and Ernesto Staffetti. Uncertainty quantification of a mathematical model of covid-19 transmission dynamics with mass vaccination strategy. *Chaos, Solitons & Fractals*, 146:110895, 2021.

[163] World Health Organization et al. Origin of sars-cov-2, 26 march 2020. Technical report, World Health Organization, 2020.

[164] Sarah P Otto, Troy Day, Julien Arino, Caroline Colijn, Jonathan Dushoff, Michael Li, Samir Mechai, Gary Van Domselaar, Jianhong Wu, David JD Earn, et al. The origins and potential future of sars-cov-2 variants of concern in the evolving covid-19 pandemic. *Current Biology*, 31(14):R918–R929, 2021.

[165] OWID. Covid-19 data maintained by our world in data. [online]. available: https://github.com/owid/covid-19-data/tree/master/public/data, 2024.

[166] Chintamani Pai, Ankush Bhaskar, and Vaibhav Rawoot. Investigating the dynamics of covid-19 pandemic in india under lockdown. *Chaos, Solitons & Fractals*, 138:109988, 2020.

[167] Sibarama Panigrahi and Himansu Sekhar Behera. A study on leading machine learning techniques for high order fuzzy time series forecasting. *Engineering Applications of Artificial Intelligence*, 87:103245, 2020.

[168] Jieun Park, Dokkyun Yi, and Sangmin Ji. Analysis of recurrent neural network and predictions. *Symmetry*, 12(4):615, 2020.

Bibliography

[169] Showmick Guha Paul, Arpa Saha, Al Amin Biswas, Md Sabab Zulfiker, Mohammad Shamsul Arefin, Md Mahfujur Rahman, and Ahmed Wasif Reza. Combating covid-19 using machine learning and deep learning: applications, challenges, and future perspectives. *Array*, 17:100271, 2023.

[170] Nicola Perra. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Physics Reports*, 2021.

[171] pharmeasy.in. Covaxin vs covishield - a detailed comparison - efficacy, side effects, August 10, 2023.

[172] Marvin G Pizon, Ronald R Baldo, Ruthlyn N Villarante, and Jessica D Balatero. Path analysis of covid-19 with the influence of air pressure, air temperature, and relative humidity. *arXiv preprint arXiv:2105.05451*, 2021.

[173] Fernando P Polack, Stephen J Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L Perez, Gonzalo Pérez Marc, Edson D Moreira, Cristiano Zerbini, et al. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. *New England journal of medicine*, 383(27):2603–2615, 2020.

[174] Dimiter Prodanov. Computational aspects of the approximate analytic solutions of the sir model: applications to modelling of covid-19 outbreaks. *Nonlinear Dynamics*, pages 1–19, 2023.

[175] Zongxi Qu, Yutong Li, Xia Jiang, and Chunhua Niu. An innovative ensemble model based on multiple neural networks and a novel heuristic optimization algorithm for covid-19 forecasting. *Expert Systems with Applications*, 212:118746, 2023.

[176] Ali A Rabaan, Shamsah H Al-Ahmed, Ranjit Sah, Ruchi Tiwari, Mohd Yatoo, Shailesh Kumar Patel, Mamta Pathak, Yashpal Singh Malik, Kuldeep Dhama, Karam Pal Singh, et al. Sars-cov-2/covid-19 and advances in developing potential therapeutics and vaccines to counter this emerging pandemic. *Annals of Clinical Microbiology and Antimicrobials*, 19(1):1–37, 2020.

[177] Firas A Rabi, Mazhar S Al Zoubi, Ghena A Kasasbeh, Dunia M Salameh, and Amjad D Al-Nasser. Sars-cov-2 and coronavirus disease 2019: what we know so far. *Pathogens*, 9(3):231, 2020.

[178] Seyed Ali Rakhshan, Mahdi Soltani Nejad, Marzie Zaj, and Fatemeh Helen Ghane. Global analysis and prediction scenario of infectious outbreaks by recurrent dynamic model and machine learning models: A case study on covid-19. *Computers in Biology and Medicine*, 158:106817, 2023.

[179] Muhammad Rendana and Wan Mohd Razi Idris. New covid-19 variant (b. 1.1. 7): Forecasting the occasion of virus and the related meteorological factors. *Journal of Infection and Public Health*, 2021.

[180] Chiara Reno, Francesco Sanmarchi, Michael A Stoto, Maria Pia Fantini, Jacopo Lenzi, and Davide Golinelli. The impact of health policies and vaccine rollout on the covid-19 pandemic waves in italy. *Health policy and technology*, 11(2):100604, 2022.

[181] Reuters. Russiaś sputnik v shot around 90developers say, January 29, 2021.

[182] Matheus Henrique Dal Molin Ribeiro, Ramon Gomes da Silva, Viviana Cocco Mariani, and Leandro dos Santos Coelho. Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil. *Chaos, Solitons & Fractals*, page 109853, 2020.

[183] Ali Roziqin, Syasya YF Mas'udi, and Iradhad T Sihidi. An analysis of indonesian government policies against covid-19. *Public Administration and Policy*, 2021.

[184] Ingrid Różyło-Kalinowska and Kaan Orhan. Applications of machine learning and artificial intelligence in the covid-19 pandemic. In *Artificial Intelligence in Dentistry*, pages 247–257. Springer, 2024.

[185] Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byung-Won On, Waqar Aslam, and Gyu Sang Choi. Covid-19 future forecasting using supervised machine learning models. *IEEE access*, 8:101489–101499, 2020.

[186] Sara Saadatmand, Khodakaram Salimifard, Reza Mohammadi, Alex Kuiper, Maryam Marzban, and Akram Farhadi. Using machine learning in prediction of icu admission, mortality, and length of stay in the early stage of admission of covid-19 patients. *Annals of Operations Research*, 328(1):1043–1071, 2023.

[187] Tanzila Saba, Ibrahim Abunadi, Mirza Naveed Shahzad, and Amjad Rehman Khan. Machine learning techniques to detect and forecast the daily total covid-19 infected and deaths cases under different lockdown types. *Microscopy Research and Technique*, 2021.

[188] Jyoti Prakash Sahoo, Ambika Prasad Mishra, Laxmipriya Behera, Sudhanya Nath, and Kailash Chandra Samal. New mutant covid-19 strain (vui–202012/01)–more contagious than current status. *Biotica Research Today*, 2(12):1331–1333, 2020.

[189] Fernando Saldaña and Jorge X Velasco-Hernández. Modeling the covid-19 pandemic: a primer and overview of mathematical epidemiology. *SeMA Journal*, pages 1–27, 2022.

[190] Rebecca Salles, Kele Belloze, Fabio Porto, Pedro H Gonzalez, and Eduardo Ogasawara. Nonstationary time series transformation methods: An experimental review. *Knowledge-Based Systems*, 164:274–291, 2019.

Bibliography

[191] Daniela Sánchez, Patricia Melin, and Oscar Castillo. A grey wolf optimizer for modular granular neural networks for human recognition. *Computational intelligence and neuroscience*, 2017, 2017.

[192] Kankan Sarkar, Subhas Khajanchi, and Juan J Nieto. Modeling and forecasting the covid-19 pandemic in india. *Chaos, Solitons & Fractals*, 139:110049, 2020.

[193] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

[194] Ahmad Sedaghat, Shahab Band, Amir Mosavi, and Laszlo Nadai. Predicting trends of coronavirus disease (covid-19) using sird and gaussian-sird models. In *2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE)*, pages 000267–000274. IEEE, 2020.

[195] Carlos A Severiano, Petrônio CL Silva, Hossein Javedani Sadaei, and Frederico Gadelha Guimarães. Very short-term solar forecasting using fuzzy time series. In *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2017.

[196] Warda M Shaban, Asmaa H Rabie, Ahmed I Saleh, and MA Abo-Elsoud. Detecting covid-19 patients based on fuzzy inference engine and deep neural network. *Applied soft computing*, 99:106906, 2021.

[197] Sourabh Shastri, Kuljeet Singh, Sachin Kumar, Paramjit Kour, and Vibhakar Mansotra. Time series forecasting of covid-19 using deep learning models: India-usa comparative case study. *Chaos, Solitons & Fractals*, 140:110227, 2020.

[198] Honghao Shi, Jingyuan Wang, Jiawei Cheng, Xiaopeng Qi, Hanran Ji, Claudio J Struchiner, Daniel AM Villela, Eduard V Karamov, and Ali S Turgiev. Big data technology in infectious diseases modeling, simulation, and prediction after the covid-19 outbreak. *Intelligent Medicine*, 2023.

[199] Leyuan Shi and Sigurdur Olafsson. Nested partitions method for global optimization. *Operations research*, 48(3):390–407, 2000.

[200] Xuan-Li Shi, Feng-Feng Wei, and Wei-Neng Chen. A swarm-optimizer-assisted simulation and prediction model for emerging infectious diseases based on seir. *Complex & Intelligent Systems*, 9(2):2189–2204, 2023.

[201] Gitanjali R Shinde, Asmita B Kalamkar, Parikshit N Mahalle, Nilanjan Dey, Jyotismita Chaki, and Aboul Ella Hassanien. Forecasting models for coronavirus disease (covid-19): a survey of the state-of-the-art. *SN Computer Science*, 1:1–15, 2020.

[202] Andrey S Shkoda, Vladimir A Gushchin, Darya A Ogarkova, Svetlana V Stavitskaya, Olga E Orlova, Nadezhda A Kuznetsova, Elena N Keruntu, Andrei A Pochtovyi, Alexander V Pukhov, Denis A Kleymenov, et al. Sputnik v effectiveness against hospitalization with covid-19 during omicron dominance. *Vaccines*, 10(6):938, 2022.

[203] Muhammad Shoaib, Muhammad Asif Zahoor Raja, Muhammad Touseef Sabir, Ayaz Hussain Bukhari, Hussam Alrabaiah, Zahir Shah, Poom Kumam, and Saeed Islam. A stochastic numerical analysis based on hybrid nar-rbfs networks nonlinear sitr model for novel covid-19 dynamics. *Computer Methods and Programs in Biomedicine*, 202:105973, 2021.

[204] Sakshi Shringi, Harish Sharma, Pushpa Narayan Rathie, Jagdish Chand Bansal, Atulya Nagar, and Daya Lal Suthar. Predicting covid-19 outbreak in india using modified sird model. *Applied Mathematics in Science and Engineering*, 32(1):2305191, 2024.

[205] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pages 1394–1401. Ieee, 2018.

[206] Sarbjit Singh, Kulwinder Singh Parmar, Sidhu Jitendra Singh Makkhan, Jatinder Kaur, Shruti Peshoria, and Jatinder Kumar. Study of arima and least square support vector machine (ls-svm) models for the prediction of sars-cov-2 confirmed cases in the most affected countries. *Chaos, Solitons & Fractals*, 139:110086, 2020.

[207] Catrin Sohrabi, Zaid Alsafi, Niamh ONeill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, and Riaz Agha. World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19). *International Journal of Surgery*, 2020.

[208] Qiang Song and Brad S Chissom. Forecasting enrollments with fuzzy time series—part i. *Fuzzy sets and systems*, 54(1):1–9, 1993.

[209] Qiang Song and Brad S Chissom. Fuzzy time series and its models. *Fuzzy sets and systems*, 54(3):269–277, 1993.

[210] Qiang Song and Brad S Chissom. Forecasting enrollments with fuzzy time series-part ii. *Fuzzy sets and systems*, 62(1):1–8, 1994.

[211] Derya Soydaner. Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*, 34(16):13371–13385, 2022.

[212] Xiaoying Su, Yanfeng Sun, Hongxi Liu, Qiuling Lang, Yichen Zhang, Jiquan Zhang, Chaoyong Wang, and Yanan Chen. An innovative ensemble model

based on deep learning for predicting covid-19 infection. *Scientific Reports*, 13(1):12322, 2023.

[213] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1):43–62, 1997.

[214] Abdusy Syarif, Novi Azman, Viktor Vekky Ronal Repi, Ernawati Sinaga, and Muhamad Asvial. Unas-net: A deep convolutional neural network for predicting covid-19 severity. *Informatics in Medicine Unlocked*, 28:100842, 2022.

[215] Yoshiki Takeguchi and Kazuyuki Yagasaki. Optimal control of the seir epidemic model using a dynamical systems approach. *Japan Journal of Industrial and Applied Mathematics*, 41(1):297–316, 2024.

[216] Hai Tao, Sani I Abba, Ahmed M Al-Areeq, Fredolin Tangang, Sandeep Samantaray, Abinash Sahoo, Hugo Valadares Siqueira, Saman Maroufpoor, Vahdettin Demir, Neeraj Dhanraj Bokde, et al. Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: A comprehensive review, assessment, and possible future research directions. *Engineering Applications of Artificial Intelligence*, 129:107559, 2024.

[217] Sean J Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

[218] LaFraniere S Thomas K, Weiland N. F.d.a. advisory panel gives green light to pfizer vaccine. [online]. available: https://www.nytimes.com/2020/12/10/health/covid-vaccine-pfizer-fda.html, (December 10, 2020.

[219] Vincent S Tseng, Josh Jia-Ching Ying, Stephen TC Wong, Diane J Cook, and Jiming Liu. Computational intelligence techniques for combating covid-19: A survey. *IEEE Computational Intelligence Magazine*, 15(4):10–22, 2020.

[220] Vedide Rezan Uslu, Eren Bas, Ufuk Yolcu, and Erol Egrioglu. A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations. *Swarm and Evolutionary Computation*, 15:19–26, 2014.

[221] Anil Utku. Deep learning based hybrid prediction model for predicting the spread of covid-19 in the world's most populous countries. *Expert Systems with Applications*, 231:120769, 2023.

[222] Nghiem Van Tinh. Forecasting of covid-19 confirmed cases in vietnam using fuzzy time series model combined with particle swarm optimization. *Comput. Res. Progr. Appl. Sci. Eng.*, 6(2):114–120, 2020.

[223] Nghiem Van Tinh and Nguyen Cong Dieu. A new hybrid fuzzy time series forecasting model based on combining fuzzy c-means clustering and particle swam optimization. *Journal of Computer Science and Cybernetics*, 35(3):267–292, 2019.

[224] HP Vinutha, B Poornima, and BM Sagar. Detection of outliers using interquartile range technique from intrusion dataset. In *Information and Decision Sciences: Proceedings of the 6th International Conference on FICTA*, pages 511–518. Springer, 2018.

[225] Philip V'kovski, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, and Volker Thiel. Coronavirus biology and replication: implications for sars-cov-2. *Nature Reviews Microbiology*, 19(3):155–170, 2021.

[226] Weina Wang, Jiapeng Shao, and Huxidan Jumahong. Fuzzy inference-based lstm for long-term time series prediction. *Scientific Reports*, 13(1):20359, 2023.

[227] Wikipedia. Timeline of the covid-19 pandemic in india [online]. available: https://en.wikipedia.org/wiki/covid-19_pandemic_in_india, 2023.

[228] Wikipedia. Timeline of the covid-19 pandemic in the united states [online]. available: https://en.wikipedia.org/wiki/timeline_of_the_covid-19_pandemic_in_the_united_states, 2023.

[229] Feuer Will. Texas, mississippi lift covid restrictions and mask mandates, despite cdc warnings, March 2, 2021.

[230] Shaun S Wulff. Time series analysis: Forecasting and control. *Journal of Quality Technology*, 49(4):418, 2017.

[231] Sidong Xian, Kaiyuan Chen, and Yue Cheng. Improved seagull optimization algorithm of partition and xgboost of prediction for fuzzy time series forecasting of covid-19 daily confirmed. *Advances in Engineering Software*, 173:103212, 2022.

[232] Xiaobing Xian, Liang Wang, Xiaohua Wu, Xiaoqing Tang, Xingpeng Zhai, Rong Yu, Linhan Qu, and Mengliang Ye. Comparison of sarima model, holt-winters model and ets model in predicting the incidence of foodborne disease. *BMC Infectious Diseases*, 23(1):803, 2023.

[233] Yue Xiang, Yonghong Jia, Linlin Chen, Lei Guo, Bizhen Shu, and Enshen Long. Covid-19 epidemic prediction and the impact of public health interventions: A review of covid-19 epidemic models. *Infectious Disease Modelling*, 6:324–342, 2021.

[234] Lu Xu, Rishikesh Magar, and Amir Barati Farimani. Forecasting covid-19 new cases using deep learning methods. *Computers in biology and medicine*, 144:105342, 2022.

Bibliography

[235] Zhenxing Xu, Chang Su, Yunyu Xiao, and Fei Wang. Artificial intelligence for covid-19: Battling the pandemic with computational intelligence. *Intelligent medicine*, 2(1):13–29, 2022.

[236] Xin-She Yang. Firefly algorithms for multimodal optimization. In *International symposium on stochastic algorithms*, pages 169–178. Springer, 2009.

[237] Chong You, Yuhao Deng, Wenjie Hu, Jiarui Sun, Qiushi Lin, Feng Zhou, Cheng Heng Pang, Yuan Zhang, Zhengchao Chen, and Xiao-Hua Zhou. Estimation of the time-varying reproduction number of covid-19 outbreak in china. *International journal of hygiene and environmental health*, 228:113555, 2020.

[238] LA Zadeh. Fuzzy sets information and control, vol. 8 (1965). *MathSciNet zbMATH*, pages 338–353.

[239] Ghaida Zaid. Zoonotic infections. In *Coronavirus Disease*, pages 21–28. Elsevier, 2022.

[240] Yuri Zelenkov and Ivan Reshettsov. Analysis of the covid-19 pandemic using a compartmental model with time-varying parameters fitted by a genetic algorithm. *Expert Systems with Applications*, 224:120034, 2023.

[241] Abdelhafid Zeroual, Fouzi Harrou, Abdelkader Dairi, and Ying Sun. Deep learning methods for forecasting covid-19 time-series data: A comparative study. *Chaos, Solitons & Fractals*, 140:110121, 2020.

[242] Xiaolei Zhang, Renjun Ma, and Lin Wang. Predicting turning point, duration and attack rate of covid-19 outbreaks in major western countries. *Chaos, Solitons & Fractals*, page 109829, 2020.

[243] Yangyi Zhang, Sui Tang, and Guo Yu. An interpretable hybrid predictive model of covid-19 cases using autoregressive model and lstm. *Scientific reports*, 13(1):6708, 2023.

[244] Nanning Zheng, Shaoyi Du, Jianji Wang, He Zhang, Wenting Cui, Zijian Kang, Tao Yang, Bin Lou, Yuting Chi, Hong Long, et al. Predicting covid-19 in china using hybrid ai model. *IEEE transactions on cybernetics*, 50(7):2891–2904, 2020.

[245] Binggui Zhou, Guanghua Yang, Zheng Shi, and Shaodan Ma. Interpretable temporal attention network for covid-19 forecasting. *Applied Soft Computing*, 120:108691, 2022.

[246] Luyu Zhou, Chun Zhao, Ning Liu, Xingduo Yao, and Zewei Cheng. Improved lstm-based deep learning model for covid-19 prediction using optimized approach. *Engineering applications of artificial intelligence*, 122:106157, 2023.

[247] Xiao Zhou, Xiaohu Zhang, Paolo Santi, and Carlo Ratti. Phase-wise evaluation and optimization of non-pharmaceutical interventions to contain the covid-19 pandemic in the us. *Frontiers in Public Health*, 11:1198973, 2023.

[248] Yimin Zhou, Zuguo Chen, Xiangdong Wu, Zengwu Tian, Lingjian Ye, and Leyi Zheng. Retrospect: the outbreak evaluation of covid-19 in wuhan district of china. In *Healthcare*, volume 9, page 61. MDPI, 2021.