

COVID-19 IDENTIFIER USING SMLT AND DATA SCIENCE

Mr. A. SATHEESH, AJITH S, PRANESH S, RICHARD JOSEPH P P, SURYA K
DEPT OF COMPUTER SCIENE AND ENGINEERING
JEPPIAAR INSTITUTE OF TECHNOLOGY

Abstract- The worldwide population has been devastated by the Coronavirus Disease 2019 (COVID-19) pandemic, which began in Wuhan, China, and has overloaded advanced healthcare systems around the world. WHO is actively monitoring and responding to this epidemic. The current rapid and exponential increase in the number of patients has prompted the use of AI approaches to forecast the likely result of an infected patient in order to provide suitable therapy. The goal is to determine whether machine learning-based algorithms can accurately anticipate whether or not covid-19 recovery is achievable. The supervised machine learning approach (SMLT) is used to evaluate the dataset and capture various information such as variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments, and data validation, data cleaning/preparing, and data analysis.

Keywords: Covid-19 and Deep Learning, Machine Learning Algorithms, Supervised Machine Learning Techniques (SMLT)

INTRODUCTION

Data science is an interdisciplinary field that use scientific methods, procedures, algorithms, and systems to extract knowledge and insights from structured and unstructured data, as well as to apply that knowledge and actionable insights to a variety of application areas.

The term "data science" dates back to 1974, when Peter Naur proposed it as a replacement for the term "computer science." The International Federation of

Classification Societies was the first conference to address data science expressly in 1996. The definition, however, was still in motion.

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may additionally be applied to any machine that exhibits traits related to a person's mind like learning and problem-solving.

Artificial intelligence (AI) is intelligence demonstrated by machines, as opposition the natural intelligence displayed by humans or animals. Leading AI textbooks define the sector because the study of "intelligent agents" any system that perceives its environment and takes actions that maximize its chance of achieving its goals. Some popular accounts use the term "artificial intelligence" to explain machines that mimic "cognitive" functions that humans go with the human mind, like "learning" and "problem solving", however this definition is rejected by major AI researchers.

I.PROBLEM STATEMENT

In December 2019, a virulent disease named COVID-19 broke enter Wuhan, China, and during a few weeks, it spread to quite 200 countries worldwide. Every country infected with the disease started taking necessary measures to prevent the spread and supply the simplest possible medical facilities to infected patients and take precautionary measures to regulate the spread. because the infection spread was exponential, there arose a necessity to model infection spread patterns to estimate the patient volume computationally. Such patients' estimation is that

the key to the required actions that local governments may desire counter the spread, control hospital load, and resource allocations. The existing method is only patient count and it does not classify whether covid or not the patient is recovered or not.

II. PROPOSED SYSTEM

Multiple datasets from different sources would be combined to create a generalized dataset, then different machine learning algorithms would be applied to extract patterns and to get results with maximum accuracy. In this section of the report will load within the data, check for cleanliness, and so trim and clean given dataset for analysis. confirm that the document steps carefully and justify for cleaning decisions.

III. MODULES

In this paper, there are four stages:

Data Pre-processing

Machine learning validation approaches are used to calculate the error rate of the Machine Learning (ML) model, which is as close to the genuine error rate of the dataset as possible. Validation approaches may not be required if the data volume is large enough to be representative of the population. However, in real-world circumstances, it is necessary to work with data samples that are not always representative of the population of a dataset. Duplicate the value and the data type description to identify the missing value, whether it is a float variable or an integer variable. While tuning model hyper parameters, a sample of data is employed to offer an unbiased evaluation of a model fit on the training dataset.

- Import libraries for access and functional purpose and read the given dataset
- General Properties of Analyzing the given dataset

- Display the given dataset in the form of data frame
- show columns and shape of the data frame
- To describe the data frame
- Checking data type and information about dataset
- Checking for duplicate data
- Checking Missing values and unique values of data frame

Data Analysis of Visualization

In December 2019, a virus named COVID-19 broke go in Wuhan, China, and during a few weeks, it spread to quite 200 countries worldwide. Every country infected with the disease started taking necessary measures to prevent the spread and supply the simplest possible medical facilities to infected patients and take precautionary measures to manage the spread. because the infection spread was exponential, there arose a desire to model infection spread patterns to estimate the patient volume computationally. Such patients' estimation is that the key to the required actions that local governments may fancy counter the spread, control hospital load, and resource allocations. In applied statistics and machine learning, data Visualization may be a crucial ability. Statistics is anxious with quantitative data descriptions and estimations. Data visualization is a crucial set of tools for analysis. This may well be useful for spotting patterns, faulty data, outliers, and other things when exploring and progressing to know a dataset.

Comparing Algorithm with prediction in the form of best accuracy result

1. It is critical to compare the performance of various different machine learning algorithms consistently, and this tutorial will show you

how to develop a test harness in Python using scikit-learn to do so.

2. This test harness can be used as a framework for your own machine learning tasks, with additional and different algorithms to compare.
3. The performance characteristics of each model will vary. You may gain an idea of how reliable each model is on unseen data using resampling approaches like cross validation.

Flask

Flask is a Python-based microweb framework. Because it does not require any specific tools or libraries, it is characterized as a micro-framework. It lacks a database abstraction layer, form validation, or any other components for which third-party libraries already exist.

Extensions, on the other hand, can be used to add application functionalities as if they were built into Flask itself. Object-relational mappers, form validation, upload handling, different open authentication protocols, and other framework-related tools all have extensions.

In this module is used to help to the Consumer to View the Payment status report to check the payment, Paid or Not with the land longitude and the will update the report along with their opinion and the will be stored the database.

IV.CLASSIFICATION

ALGORITHM

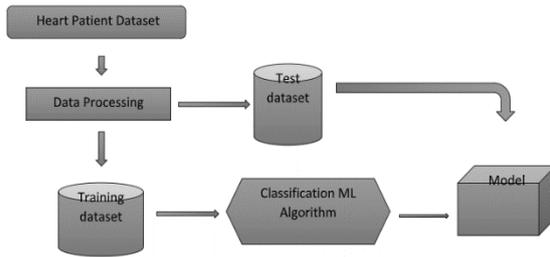
Classification is a supervised learning strategy in machine learning and statistics in which a computer

program learns from the data input supplied to it and then applies that learning to classify fresh observations. This data collection could be bi-class (for example, determining whether the person is male or female or whether the email is spam or not) or multi-class. Speech recognition, handwriting recognition, biometric identification, document classification, and other classification challenges are examples.

V.METHODOLOGY

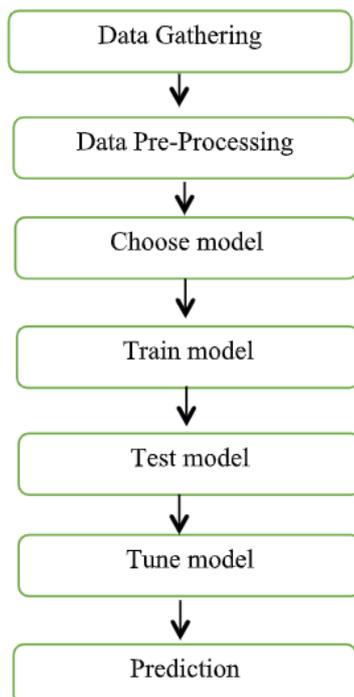
A. Data Extraction and Transformation

- For predictive analysis, data are extracted from Covid-19 information web portal [14]. The daily updates of new cases, recovered cases, deaths, and critical cases are updated on this web portal and are maintained by the Ministry of Health, Pakistan. This Covid-19 data source is the official resource that contains the province wise stats of all the above-listed parameters. We have used the data set available till the submission date of this article for model training and testing.
- Once the data are extracted and required preprocessing (on data columns containing dates using date and time functions of Pandas) has been performed, it is divided into two partitions: training data and testing data. It is essential to convert time-series data into a simple structure with input and output components before it fits into the supervised learning model.



B. Long Short-Term Memory

- LSTM is a special type of artificial neural network that performs well on classification and regression tasks. LSTM is a special type of recurrent neural network (RNN) that uses memory units to serve as short-term memory. LSTM has given outstanding performance on speech recognition, handwriting recognition, and pattern recognition.
- After configuring all parameters, training data are stored in batches and are passed to the compiled model for the training purpose. In this process, the training data are repeated according to a defined number of epochs.



VII.FUTURE ENCHANCEMENT

- Covid Disease prediction to connect with Cloud.
- To optimize the work to implement in Artificial Intelligence environment. Covid-19 pandemic is a growing manifold daily. With the ever-increasing number of cases, bulk testing of cases swiftly may be required.
- In this work, we experimented with multiple CNN models in an attempt to classify the Covid-19 affected patients using their chest X-ray scans.

CONCLUSION

In this article, we have used the LSTM model to predict the COVID-19 PC in Pakistan. Due to the limited availability of PCR COVID-19 test kits, we cannot directly predict PC. Instead, we have calculated PC by predicting both PPP and total tests per day separately. Results revealed that our predicted PC is much closer to the actual PC between June 1,2020, and June 24, 2020. The proposed model can also predict Covid-19 cases, area wise, that help smart lockdown decisions and area wise sampling. In the future, we have plans to train the model on death cases and will try to forecast the mortality ratio and its correlations with the critical cases.

REFERENCES

- [1] F. Wu et al., "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020.
- [2] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, pp. 497–506, May 2020.
- [3] Coronavirus Disease 2019 (COVID-19) Situation Report, World Health Organization, Geneva, Switzerland, 2020.
- [4] E. Dong, H. Du, and L. Gardner, "An interactive Web-based dashboard to track COVID-19 in real time," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, May 2020.
- [5] Coronavirus Disease 2019 (COVID-19) Situation Report, World Health Organization, Geneva, Switzerland, 2020.
- [6] World Health Organization. Accessed: May 31, 2020. [Online]. 2019
- [7] K. J. Clerkin et al., "COVID-19 and cardiovascular disease," *Circulation*, vol. 141, no. 20, pp. 1648–1655, 2020.
- [8] B. Resnik. World Health Organization. Accessed: May 31, 2020. [2/21197617/coronavirus-pandemic-covid-19-death-rate-transmission-risk-factors-lockdowns-social-distancing](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/20200221197617/coronavirus-pandemic-covid-19-death-rate-transmission-risk-factors-lockdowns-social-distancing)
- [9] Report of the Who-China Joint Mission on Coronavirus Disease 2019 (COVID-19), World Health Organization, Geneva, Switzerland, 2020.
- [10] P. K. Ozili and T. Arun, "Spillover of COVID-19: Impact on the global economy," *Tech. Rep.*, Mar. 2020. [Online]. Available:
- [11] M. Gupta et al., "COVID-19 and economy," *Dermatol. Therapy* vol. 33, no. 4, 2020, Art. no. e13329, doi: 10.1111/dth.13329.
- [12] F. Al-Obeidat, Á. Rocha, M. S. Khan, F. Maqbool, and S. Razzaq, "Parallel tensor factorization for relational learning," *Neural Comput. Appl.*, vol. 4, pp. 1–10, Jan. 2021.
- [13] S. Razzaq, F. Maqbool, and A. Hussain, "Modified cat swarm optimization for clustering," in *Proc. Int. Conf. Brain Inspired Cognit. Syst. Cham, Switzerland: Springer*, 2016, pp. 161–170.
- [14] Health Advisory Platform. COVID-19 Pakistan. Accessed: Jun. 24, 2020.
- [15] S. Tiwari, S. Kumar, and K. Guleria, "Outbreak trends of coronavirus disease–2019 in India: A prediction," *Disaster Med. Public Health Preparedness*, vol. 14, pp. 1–6, Oct. 2020.
- [16] M. Gatto et al., "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 19, pp. 10484–10491, May 2020.
- [17] M. R. Rahman, A. H. M. H. Islam, and M. N. Islam, "Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches," *Model. Earth Syst. Environ.*, vol. 7, pp. 1–29, Sep. 2020.

[18] G. D. Sharma, G. Talan, M. Srivastava, A. Yadav, and R. Chopra, "A qualitative enquiry into strategic and operational responses to COVID-19 challenges in South Asia," *J. Public Affairs*, Jun. 2020.

[19] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial intelligence (AI) applications for COVID-19 pandemic," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 14, no. 4, pp. 337–339, Jul. 2020.

[20] A. Tariq et al., "Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020," *BMC Med.*, vol. 18, no. 1, pp. 1–14, Dec. 2020.

[21] S. Deb and M. Majumdar, "A time series method to analyze incidence pattern and estimate reproduction number of COVID-19," 2020, arXiv:2003.10655. [Online].

[22] A. J. Kucharski et al., "Early dynamics of transmission and control of COVID-19: A mathematical modelling study," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553–558, 2020.

[23] S. K. Dey, M. M. Rahman, U. R. Siddiqi, and A. Howlader, "Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach," *J. Med. Virology*, vol. 92, no. 6, pp. 632–638, Jun. 2020.

AUTHORS PROFILE



Mr. A Satheesh is currently working as an Assistant Professor in the Department of Computer Science and Engineering at Jeppiaar Institute of Technology, Chennai. He has 5.5 years of Teaching experience. He is handling various Computer Science subjects like Programming in c, object oriented and programming, operating systems, computer graphics, object-oriented analysis and design, cryptography and network security, software testing, professional ethics in engineering. He also presented 3 papers in various National and International conferences.



AJITH S is currently pursuing his bachelor's degree in the field of Computer Science and Engineering at Jeppiaar Institute of Technology, Kanchipuram. He did his schooling at Bharathidhasan Matriculation and Higher Secondary School, Kanchipuram.



PRANESH S is currently pursuing his bachelor's degree in the field of Computer Science and Engineering at Jeppiaar Institute of Technology, Kanchipuram. He did his schooling at Cheran Matriculation Higher Secondary School, Karur.



RICHARD JOSEPH P P is currently pursuing his bachelor's degree in the field of Computer Science and Engineering at Jeppiaar Institute of Technology, Kanchipuram. He did his schooling at Don

Bosco Matriculation and Higher Secondary School, Pannur.



SURYA K is currently pursuing his bachelor's degree in the field of Computer Science and Engineering at Jeppiaar Institute of Technology, Kanchipuram. He did his schooling at Sri Jayendra Golden Jubilee Matriculation and

Higher Secondary School, Thirunelveli.