Journal of Data Analysis and Information Processing, 2022, 10, 78-89
https://www.scirp.org/journal/jdaip
ISSN Online: 2327-7203
ISSN Print: 2327-7211

COVID-19 Virus Prediction Using CNN and Logistic Regression Classification Strategies

Asadi Srinivasulu, Tarkeshwar Barua, Srinivas Nowduri, Madhusudhana Subramanyam, Sivaram Rajeyyagari

Data Science Research Laboratory, BlueCrest University, Monrovia, Liberia
Email: srinu.asadi@gmail.com, tbarua1@gmail.com, drnsrini@gmail.com mmsnaidu@yahoo.com, dr.sivaram@su.edu.sa

How to cite this paper: Srinivasulu, A., Barua, T., Nowduri, S., Subramanyam, M. and Rajeyyagari, S. (2022) COVID-19 Virus Prediction Using CNN and Logistic Regression Classification Strategies. *Journal of Data Analysis and Information Processing*, **10**, 78-89.
https://doi.org/10.4236/jdaip.2022.101005

Received: December 10, 2021
Accepted: February 25, 2022
Published: February 28, 2022

Abstract

COVID-19 virus is certainly considered as one of the harmful viruses amongst all the illnesses in biological science. COVID-19 symptoms are fever, cough, sore throat, and headache. The paper gave a singular function for the prediction of most of the COVID-19 virus diseases and presented with the Convolutional Neural Networks and Logistic Regression which might be the supervised learning and gaining knowledge of strategies for most of COVID-19 virus diseases detection. The proposed system makes use of an 8-fold pass determination to get a correct result. The COVID-19 virus analysis dataset is taken from Microsoft Database, Kaggle, and UCI websites gaining knowledge of the repository. The proposed studies investigate Convolutional Neural Networks (CNN) and Logistic Regression (LR) about the usage of the UCI database, Kaggle, and Google Database Datasets. This paper proposed a hybrid method for COVID-19 virus, most disease analyses through reducing the dimensionality of capabilities the usage of Logistic Regression (LR), after which making use of the brand new decreased function dataset to Convolutional Neural Networks and Logistic regression. The proposed method received the accuracy of 78.82%, sensitiveness of 97.41%, and specialness of 98.73%. The overall performance of the proposed system is appraised thinking about performance, accuracy, error rate, sensitiveness, particularity, correlation and coefficient. The proposed strategies achieved the accuracy of 78.82% and 97.41% respectively through Convolutional Neural Networks and Logistic Regression.

Keywords

Machine Learning, COVID-19 Virus, Deep Learning, ANN, CNN and LR

1. Introduction

This developing hobby is extended with the aid of using less expensive compu-

ting electricity and low-fee memory. Thus, a large number of statistics may be stored, processed and analyzed efficiently. Machine gaining knowledge performs an important function in a huge variety of crucial applications, which include statistics mining, natural language processing, photograph recognition, professional structures and predictions [1].

Research Background

This paper specializes in most of the COVID-19 virus disease diagnoses [1]. In any country, most of the COVID-19 virus diseases [2] have been determined to be the maximum number. The latest records from the World Health Organization (WHO) [3] suggest that most of the COVID-19 virus diseases are the maximum extensively identified to be in humans and animals across the world. There are nearly 2.7 million new instances that are detected in 2020-2021. Most of the COVID-19 virus diseases outperformed the location of 5th for the motive of loss of life in humans. In several international locations with a superior generation in clinical science, the 3 - 9 weeks survival rate in the literature about most of the COVID-19 virus diseases is 80% - 90% [4]. Doctors may also wrongly diagnose lung infection when most of the COVID-19 virus diseases are detected early enough, the survival rate will increase due to the fact that better remedies may be provided for chronic diseases. The aim is to get the benefits of dimensionality discount and device gaining knowledge of most of the COVID-19 virus diseases. The attention of this paper is to combine those approaches gaining knowledge of strategies with characteristic selection or characteristic extraction strategies and examine their performances to become aware of the maximum appropriate approach.

2. Literature Survey

Machine Learning (ML) is a kind of synthetic intelligence that makes a specialty of the improvement of laptop applications [5] which can extrude whilst uncovered to brand new facts. It makes use of laptop fashions and records received from beyond and former facts to useful resource type, prediction [6] and sensing procedure. This research article turned into planned to assess a number of extensively used type algorithmic rules and software in COVID-19 virus most disease diagnosing. Features [7] attribute may be decreased the usage of an appropriate function choice or functional method [8]. There are numerous techniques used to lessen the scale of capabilities in the dataset. Features choice strategies contains deciding on a set of capabilities from the authentic set of capabilities [9]. Feature extraction is done in order to obtain objectives at producing new capabilities with the aid of using merging the authentic capabilities. A large range of capabilities could affect the overall performance of a system studying version. This painting used 4 special techniques to remedy the excessive dimensionality problem. Several current strategies were advanced with the development of the generation for forecasting of COVID-19 virus most disease. The paintings as-

sociated with the area printed rapidly as follows. Azar *et al.* [10] projected a singular approach for the sensing [11] of the COVID-19 virus most disease. This technique used three types of algorithmic rules called radial foundation function, probability-based neural network and multiple layer classification. The approach skilled the property of COVID-19 virus most disease dataset which is trying out method additionally carried out. The overall performance of the machine is calculated in phrases of a few systems studying overall performance degree indices quality, particularity, sensitiveness, etc. MLP has done accuracy of 79.80% and 97.41% for education and trying out individuals. The authors in [12] proven a machine for the identity of COVID-19 virus most diseases that are carried out for two special Wisconsin COVID-19 virus [13] datasets the usage of GA function choice gets rid of the useless attributes of the facts and affords appropriate facts that could accelerate the machine. Various system studying strategies were carried out for type purposes. The better accuracy of 99.48% is done with the aid of using the Rotation Forest version with GA-primarily based attributes choice.

3. System Methodology

3.1. Existing System

There are various existing methods that are accessible in profound learning [14] strategies, for example, CNN [15] and Artificial Neural Networks LR [16] calculation drawbacks: The disadvantages are as per the following:

- Less precision;
- High time complexity;
- High execution time;
- High error rate;
- Less data size.

CNN calculation downsides: The inconveniences are as per the following:

- Less precision;
- High time complexity;
- High execution time;
- High error rate;
- Less data size.

3.2. Proposed System

There are two proposed techniques are available in machine learning such as Logistic Regression (LR) and Extended Convolutional Neural Networks (CNN).

LR algorithm Advantages: The advantages are as follows:

- ✓ High accuracy;
- ✓ Less time complexity;
- ✓ Less execution time;
- ✓ Less error rate;
- ✓ Large data size.

CNN algorithm advantages: The advantages are as follows:

- ✓ High accuracy;
- ✓ Less time complexity;
- ✓ Less execution time;
- ✓ Less error rate;
- ✓ Large data size.

4. Experimental Results

The fundamental thought of our framework plan and execution is to guarantee that, the COVID-19 virus infection patient's data worked in research that can oblige, the arrangement, sections for their initial expectation. This framework configuration is hence a technique or strong point of depicting the arrangement, parts, modules, interfaces, and information for an appropriate construction to fulfill fundamentals. There are some spread and joint exertion with the informational collections as far as their constructions evaluation, frameworks strategy and frameworks structure. Implementation or capability is assessed depending on their yield predictable by the application. Essential specifics have been found to consume a huge impact on the examination of their system. Given the fitting patients' essential Characteristics, brings about a possible construction to a prevalent structure; that in the end fits into our necessary condition. It additionally hopes to lay on an extraordinary degree with the current customers of the current system, through the need specifics.

4.1. Linear Regression Algorithm

In light of the COVID-19 virus disease informational index for example a complete COVID-19 virus disease dataset, our experimentation contained the accompanying thirteen stages:

Stage 1: Loading the modules and packages required;

Stage 2: Input the dataset and transform;

Stage 3: Design a linear regression model and fit;

Stage 4: Find the results of model fitting to validate the model for better accuracy;

Stage 5: Improve the model for better prediction;

Stage 6: Visualize data with function plot ().

In this research the perforce of our calculation is discovered to be of more precision, devouring little executing instance of time; specifying the COVID-19 virus disease cases in the initial expectation.

4.2. CNN Algorithm

Considering accomplishing more exactness, execution and time intricacy, we are compelled to expand CNN, to an all-inclusive CNN (S).

Stage 1: Import libraries of pandas, numpy, seaborne, matplotlib;

Stage 2: Load the dataset of the COVID-19 virus;

Stage 3: Construct visualization;

Stage 4: To perform the target values into a data frame;

Stage 5: Converting categorical data to numerical data;

Step 6: Initialize use only one column for the target value;

Stage 7: Call corer () on data frame X;

Stage 8: Generating the heat map that is function (sns. heatmap ();)

Stage 9: Reducing the attributes in X data frame;

Stage 10: Apply CNN for the COVID-19 virus data;

Stage 11: Compare both model fitting and predicted values;

Stage 12: Apply confusion matrix;

Stage 13: Finally print classification [4] report for COVID-19 virus.

Hence the outcome of calculations, are discovered to be of more precision, devouring little executing time; specifying the COVID-19 virus infections in the initial expectation.

5. Results

The following are the results for COVID-19 disease detection by integrating CNN and Linear Regression.

Figure 1 illustrates the execution flow through Epoches COVID-19 disease dataset from Google database, UCI and Kaggle dataset.

Figure 2 illustrates the execution environment with resources on COVID-19 disease detection dataset from Google database, UCI and Kaggle dataset.

Figure 3 illustrates the CNN model for the COVID-19 disease dataset from Google database, Microsoft DB, Amazon and UCI.

Figure 4 illustrates the graph between accuracy and loss according to the number of iterations on the COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Figure 5 illustrates the graph between loss and time according to the number of iterations on the COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Figure 6 illustrates the graph between accuracy and time according to the number of iterations on the COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Figure 7 illustrates the COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Linear Regression Results: The following are the results for COVID-19 virus disease detection by integrating CNN and Linear Regression.

Figure 8 illustrates the execution flow of the COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Figure 9 illustrates the dataset vs the number of epochs of COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Figure 10 illustrates the number of classifications of the COVID-19 dataset from Google database, Microsoft DB, Amazon and UCI.

Evaluation Methods

We used the following methodologies to demonstrate and assess the effects of our suggested technique on LR and CNN. Actual positive (AP), Untrue Positive (UP), Untrue negative (UN), and Actual Negative (AN) are initially defined onan individual basis to investigate the confusion matrix. Due to OP, the number of cases was effectively predicted as required. At the same time, the number of examples required was incorrectly estimated due to B measures.

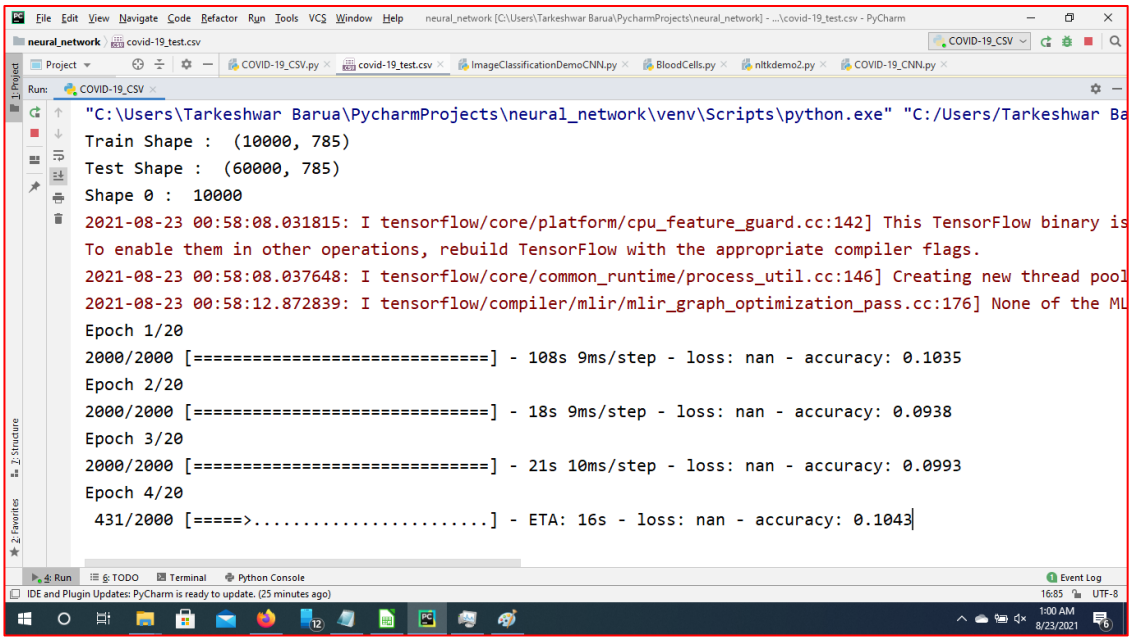

Figure 1. Execution flow of COVID-19 virus disease detection CNN.

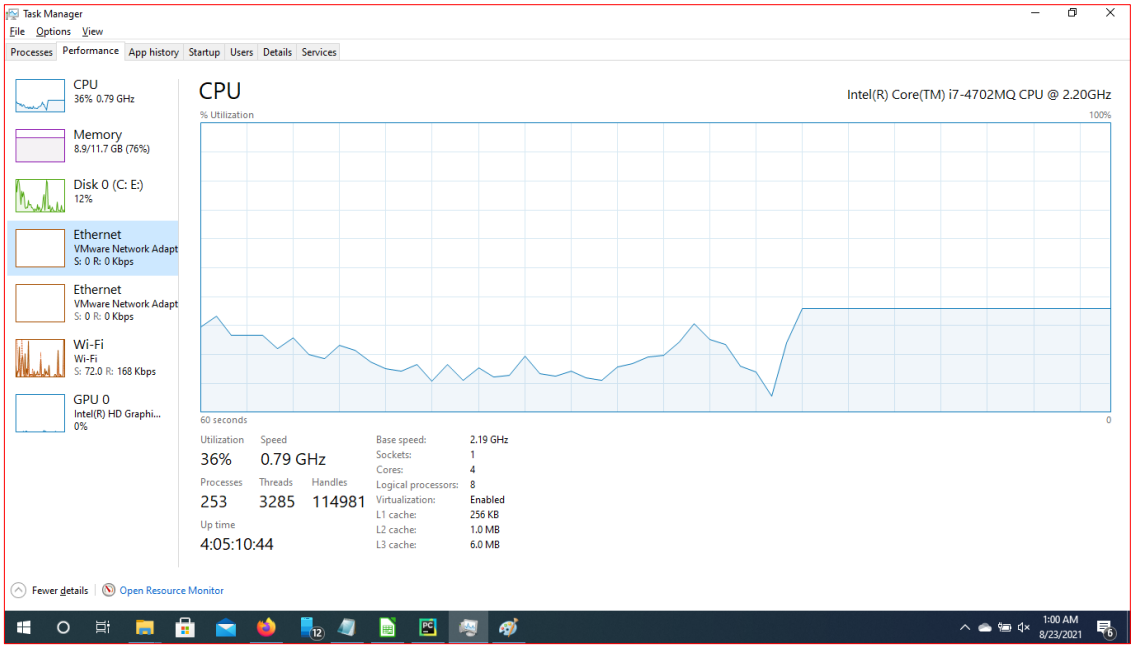

Figure 2. Resource utilization of COVID-19 disease detection CNN.

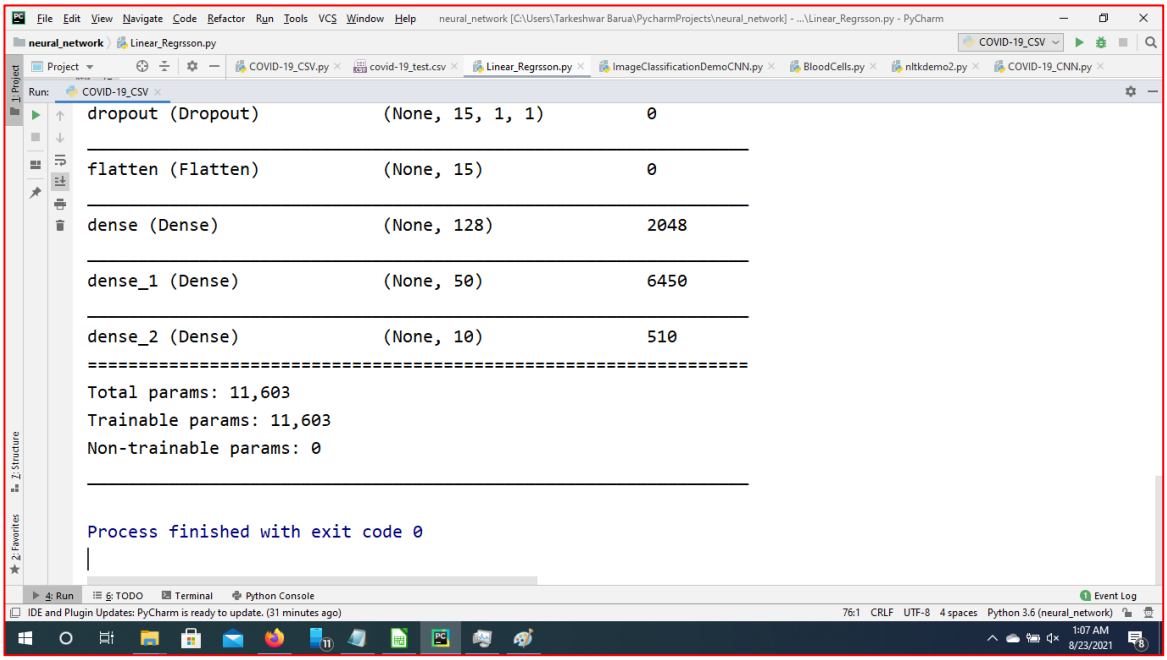

Figure 3. CNN model for COVID-19 disease dataset.

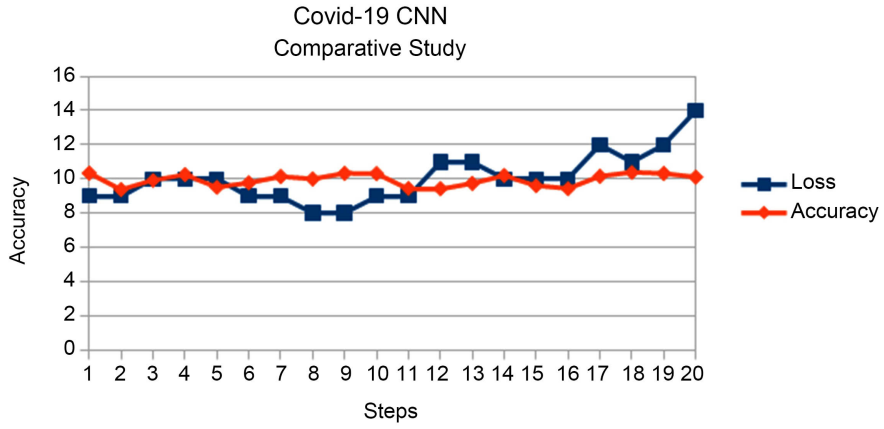

Figure 4. The graph between accuracy vs loss COVID-19 disease using CNN.

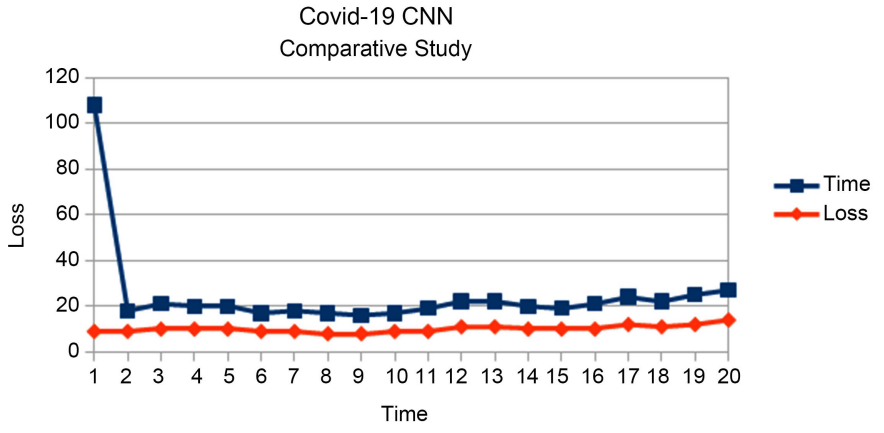

Figure 5. The graph between time vs loss COVID-19 disease using CNN.

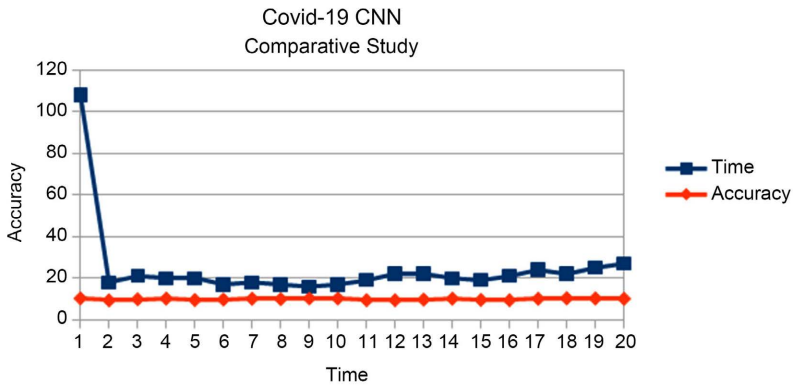

Figure 6. The graph between time vs accuracy COVID-19 disease using CNN.

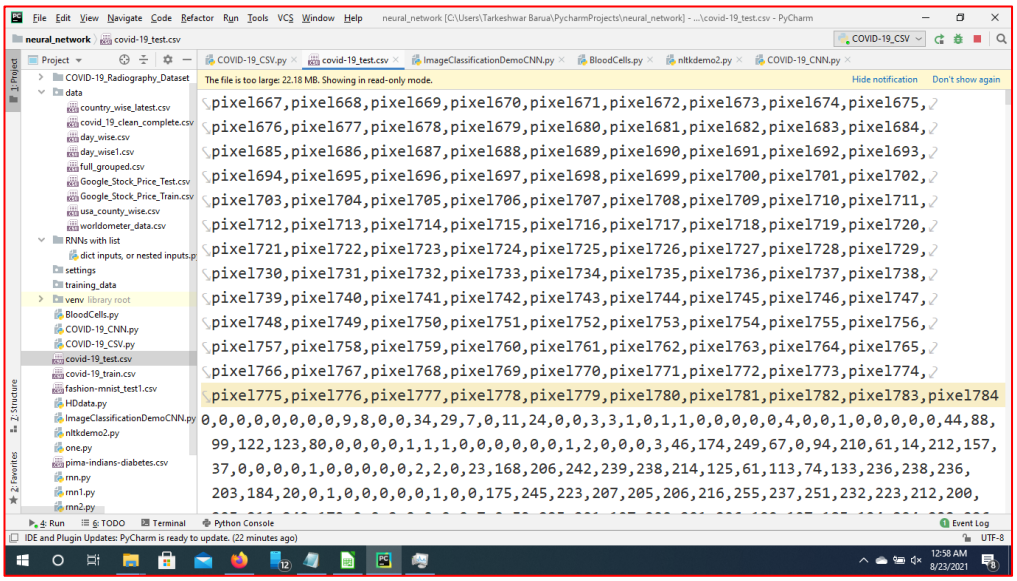

Figure 7. The graph between time vs accuracy COVID-19 disease using CNN.

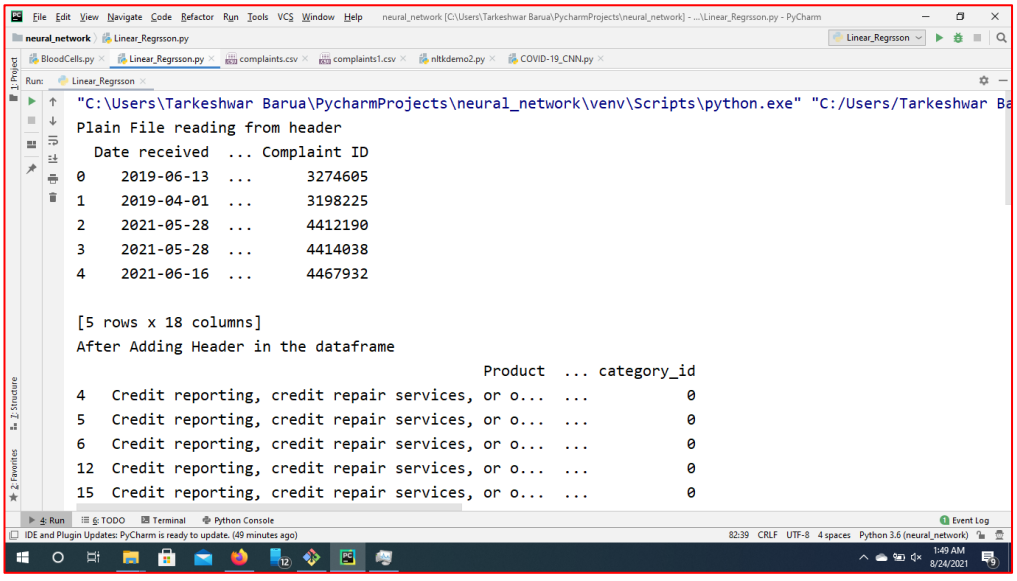

Figure 8. The graph between time vs accuracy COVID-19 disease using Linear Regression.

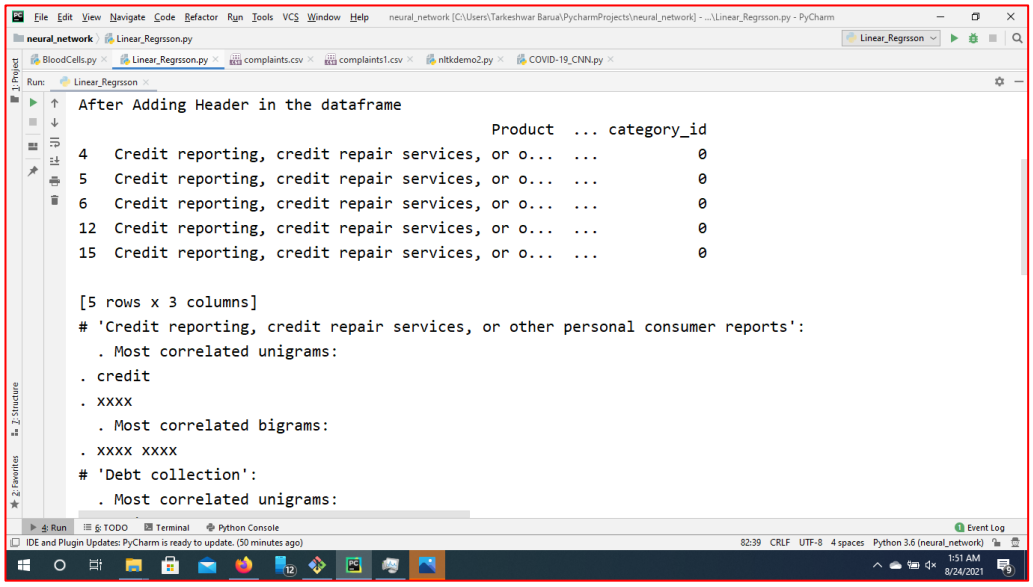

Figure 9. The graph between dataset vs number of epochs of COVID-19 disease using Linear Regression.

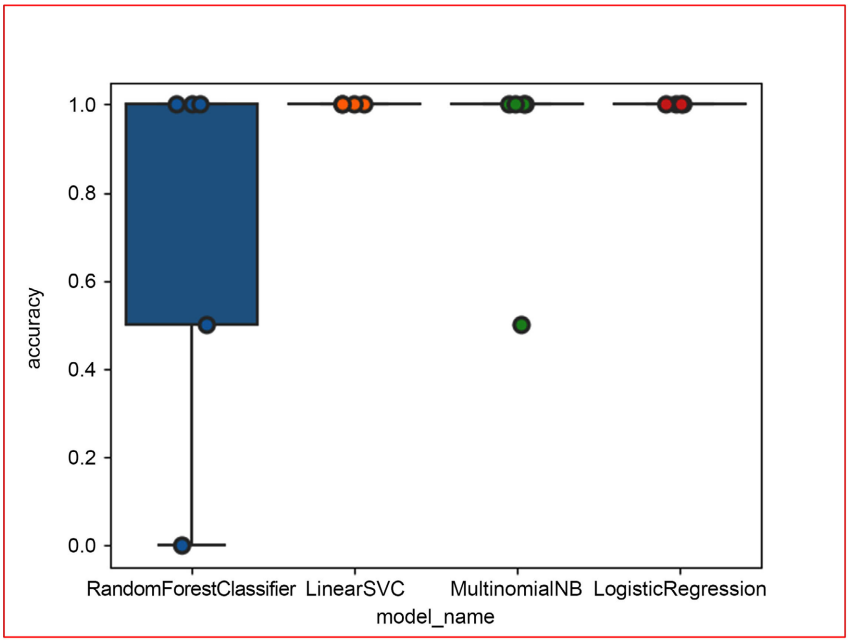

Figure 10. Box plot of COVID-19 dataset using Linear Regression.

$$\text{Quality} = \frac{AP + UN}{AP + UP + AN + UN}$$

$$\text{Preciseness} = \frac{AP}{AP + UP}$$

$$\text{Callback} = \frac{AP}{AP + UN}$$

$$\text{F-measure} = \frac{2 \times \text{Preciseness} \times \text{Callback}}{\text{Preciseness} + \text{Callback}}$$

Table 1. COVID-19 virus disease dataset with LR vs. CNN.

Parameter	LR	CNN
Accuracy	78.82	97.41
Loss	68.76	49.07
Iterations	105	105
Time Complexity	O(n3)	O(n2)
Data Size	133 MB	133 MB
Number of Parameters	03	03

Table 1 illustrates the execution flow through iterations between data loss vs accuracy one COVID-19 virus disease dataset from Microsoft, Amazon and UCI dataset. For exhibiting the likelihood of illness, every data set and all parameters are taken into account, and the preciseness of jeopardy prediction is supposed to rely on diverse assortment highlights of clinical information. Which is, higher in the exactness, the better the element presentation of the disease becomes. The precision rate in our study was 81.22 percent and 85.93 percent.

Execution Time/Time Complexity: It has been discovered that our methodology takes 50% less time than other existing techniques. The use of a Graphic Processing Unit (GPU) and a Numpy, seaborn can reduce this time even more. The time it takes to complete this task is also dependent on the system's performance. Finally, system performance is determined by system software and system hardware.

Table 1 illustrates the comparison [12] of COVID-19 Virus disease detection using LR vs. CNN with parameters Loss, Accuracy, Time complexity, data size and iterations of the given dataset that is COVID-19 Virus disease dataset from Microsoft and Amazon dataset.

6. Conclusion

The proposed research work of study contributes to the development of the state-of-the-art method for detecting COVID-19 virus disease early. In this research paper, integration of CNN and Linear Regression was applied as a highlighted choice to track down the ideal critical elements that influence COVID-19 virus disease classification. The proposed techniques achieved very good results on the 133 MB COVID-19 dataset from Kaggle. In this research, LR can assist with creating better learning and speculation capacity in the CNN classifier. In this research, the LR-CNN model was moderately steady and can merge quicker universally. To approve the presentation of the LR-CNN model, correlations with LR-CNN and CNN were carried out. The final results showed that the LR-CNN model yielded the best grouping execution for precision, explicitness, less error rate, better performance, better accuracy, affect-ability and AUC contrasted with two different models. In this research, the proposed model can be utilized to help clinical professionals in medical services exercises for the advanced discovery of COVID-19 virus disease's prediction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Islam, M.M., *et al.* (2017) Prediction of COVID-19 Virus Using Support Vector Machine and K-Nearest Neighbors. 2017 *IEEE Region* 10 *Humanitarian Technology Conference* (*R*10-*HTC*), Dhaka, 21-23 December 2017, 226-229. https://doi.org/10.1109/R10-HTC.2017.8288944

[2] Omondiagbe, D.A., *et al.* (2019) Machine Learning Classification Techniques for COVID-19 Virus Diagnosis. *IOP Conference Series*: *Materials Science and Engineering*, **495**, Article ID: 012033. https://doi.org/10.1088/1757-899X/495/1/012033

[3] Srinivasulu, A., *et al.* (2021) Advanced Lung COVID-19 Virus Prediction Based on Blockchain Material Using Extended CNN. *Applied Nanoscience*, **1**, 1-13.

[4] Polat, K., *et al.* (2007) COVID-19 Virus and Liver Disorders Classification Using Artificial Immune Recognition System (AIRS) with Performance Evaluation by Fuzzy Resource Allocation Mechanism. *Expert Systems with Applications*, **32**, 172-183. https://doi.org/10.1016/j.eswa.2005.11.024

[5] Chen, H.-L., *et al.* (2011) A Support Vector Machine Classifier with Rough Set-Based Feature Selection for COVID-19 Virus Diagnosis. *Expert Systems with Applications*, **38**, 9014-9022. https://doi.org/10.1016/j.eswa.2011.01.120

[6] Sun, Y. and Wu, D. (2008) A RELIEF Based Feature Extraction Algorithm. *Proceedings of the* 8*th SIAM International Conference on Data Mining*, Atlanta, 24-26 April 2008, 188-195. https://doi.org/10.1137/1.9781611972788.17

[7] Tanatavikorn, H. and Yamashita, Y. (2016) Fuzzy Treatment Method for Outlier Detection in Process Data. *Journal of Chemical Engineering of Japan*, **49**, 864-873. https://doi.org/10.1252/jcej.16we042

[8] Xu, L. and Yuille, A.L. (1995) Robust Artificial Neural Networks by Self-Organizing Rules Based on Statistical Physics Approach. *IEEE Transactions on Neural Networks*, **6**, 131-143. https://doi.org/10.1109/72.363442

[9] Karabatak, M. and Cevdet, M. (2009) An Expert System for Detection of COVID-19 Virus Based on Association Rules and Neural Network. *Expert Systems with Applications*, **36**, 3465-3469. https://doi.org/10.1016/j.eswa.2008.02.064

[10] Kovalerchuc, B., Triantaphyllou, E., Ruiz, J.F. and Clayton, J. (1997) Fuzzy Logic in Computer-Aided COVID-19 Virus-COVID-19 Virus Diagnosis: Analysis of Lobulation. *Artificial Intelligence in Medicine*, **11**, 75-85. https://doi.org/10.1016/S0933-3657(97)00021-3

[11] Zhou, Z.H. and Jiang, Y. (2003) Medical Diagnosis with C4.5 Rule Preceded by Artificial Neural Network Ensemble. *IEEE Transactions on Information Technology in Biomedicine*, **7**, 37-42. https://doi.org/10.1109/TITB.2003.808498

[12] Delen, D., Walker, G. and Kadam, A. (2005) Predicting COVID-19 Virus Survivability: A Comparison of Three Data Mining Methods. *Artificial Intelligence in Medicine*, **34**, 113-127. https://doi.org/10.1016/j.artmed.2004.07.002

[13] Lundin, M., Lundin, J., Burke, H.B., Toikkanen, S., Pylkkanen, L., *et al.* (1999) Artificial Neural Networks Applied to Survival Prediction in COVID-19 Virus. *Oncology*, **57**, 281-286. https://doi.org/10.1159/000012061

[14] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*: *Series*

B, **39**, 1-38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[15] Rubin, D.B. and Schenker, N. (1991) Multiple Imputation in Health-Care Databases—An Overview and Some Applications. *Statistics in Medicine*, **10**, 585-598. https://doi.org/10.1002/sim.4780100410

[16] Cristianini, N. and Shawe-Taylor, J. (2000) An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, London. https://doi.org/10.1017/CBO9780511801389