AN EFFICIENT ANALYSIS OF SPREAD OF COVID-19 AND PREDICTION IN INDIA USING CONVOLUTIONAL NEURAL NETWORK IN COMPARISON TO NAIVE BAYES WITH IMPROVED ACCURACY

M. Sasi Kumar[1], E. K.Subramanian[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The research is to show that An Efficient analysis of spread of covid-19 and prediction in India using Novel Convolutional Neural Network in comparison to Naive Bayes Algorithm.
Materials and Methods: Novel Convolutional Neural Network algorithm and Naive Bayes algorithm have been used in this research. By using the G-power software sample size is calculated and 10 sample values are taken from per group pretest value is 87%.
Results: Novel Convolutional Neural Network machine learning algorithm has provided 89% compared to Naive Bayes algorithm with 88% in the prediction of covid-19 in India with improved accuracy. There is a significant difference between the study groups with p=0.0001(p<0.001, 2 tailed).
Conclusion: Novel Convolutional neural Network algorithm provides more accuracy than Naive Bayes algorithm.

Keywords: Novel Convolutional Neural Network, Prediction, Machine Learning, Covid-19, Accuracy, Naive Bayes Algorithm.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University,Chennai,Tamilnadu,India,Pincode: 602105.

1. Introduction

COVID-19, a contagious coronavirus disease, was firstly reported in Dec 2019 in Wuhan, China, and has since spread to 212 countries or regions, affecting millions of people. India, a country with 1.3 billion population, a youngster travelling from Wuhan was diagnosed with the sickness in January 2020. As of May 3, 2020, India had over 37,000 confirmed illnesses, and the number is steadily increasing.(Ghosh, Ghosh, and Chakraborty 2020). All governments are attempting to save the lives of their citizens by enacting policies such as travel restrictions, quarantines, event postponements and cancellations, social distance, testing, and harsh and soft lockdowns (Arora, Kumar, and Panigrahi 2020). (Sy et al. 2022)Applications of efficient analysis of covid-19 include methods for forecasting future cases based on existing data, detecting infectious diseases, banking industry, insurance industry(Awassa et al. 2022) . Machine Learning approaches are used and two solutions, one for predicting the chances of being infected and other for forecasting the number of positive cases, are discussed. CoronaVirus is a disease that comes suddenly and gets worse quickly. Many times, the cause can't be found. The immediate treatment will help to recover fast. This research will help to identify the spread of coronavirus in hospitals (Shah et al. 2021).

The spread of Coronavirus has been carried out by researchers and google scholar articles of 5210, one ieee xplore paper and science direct articles of 569. The disease's contagious nature and lack of vaccines hampered social bonds and resulted in economic collapse, putting a strain on global economic and healthcare institutions. (Bhatti and Bhatti 2019) This prompted a rush of study into the virus to try to halt its spread, find a cure, and assist local governments throughout the world in determining how to stop it from spreading.(Rajaraman et al. 2020) The demand for assistance in assisting countries in identifying how to best reopen their economies and handle healthcare logistics is growing. The dissemination of COVID-19 must be predicted with precision and clarity. There is a need to anticipate patterns in the virus's transmission using available data. Many governments depend significantly on such forecasts to determine their next steps, whether it's allocating medical resources or easing or increasing lockdown levels.Machine learning is a data-driven analytic technique that focuses on the integration of a wide range of information into patterns that can be used for prediction (Martha 2020). In the existing systems, the accuracy is very low. If the accuracy is increased it will be easy to predict the Coronavirus (Ghosh, Ghosh, and Chakraborty 2020).Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)

The research gap in the existing paper is less accuracy. Here a small amount of data is used to determine accuracy because a huge amount of data will influence the accuracy. The machine learning techniques are used to conduct a comparative study on the spread of coronavirus and the accuracy values of the algorithms are found. The aim of this project is to find better accuracy to find the efficient spread of coronavirus.

2. Materials and Methods

The study is performed at Saveetha school of engineering. There are two groups in the study, which are group 1 is Novel Convolutional Neural Network and group 2 is Naive Bayes. G-power calculation is used to calculate the required samples for this analysis. (Chisholm-Burns, Vaillancourt, and Shepherd 2014) The minimum power of the analysis is fixed at 0.8, while the maximum accepted error is fixed at 0.2. The data set has the patient details of the coronavirus symptoms. (Randhawa et al. 2020) Attributes are patient name, age, gender, positive, negative. The dataset should be in a comma-separated value file format. The required samples for this analysis are done using G power calculation (p=0.001,2-tailed). The data had been collected from Kaggle. The Weka tool is a software that is used to test the dataset (Xu, Chen, and Tang 2020). The system has an 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz processor, 8GB of RAM, and 256GB of SSD storage, and it runs on Windows 10. After the collection by preprocessing, all the null values and missing values present in the dataset were removed by cleaning the data of the dataset. The algorithms are compared to find which one is better. The testing setup for the proposed system to implement with the Weka 3.8.5. SPSS tool used for statistical analysis.

Convolutional Neural Network (Cnn)
A Novel convolutional neural network is a class of artificial neural networks, most commonly applied to analyse visual imagery.CNN's are regularised versions of multilayer Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer.CNN uses relatively little pre-processing compared to other image classification algorithms and it lacks in the development of fast data mining models capable of making quick predictions.

Pseudo code for CNN:
Step 1: Import the training dataset.
Step 2: Preprocess the imported data.
Step 3: Select the classification
Step 4: Select the Novel Convolutional Neural Network algorithm.
Step 5: Start the process.

Naive Bayes:
Naive Bayes classifiers are a collection of classification algorithms based on the Naive Bayes Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, every pair of features being classified is independent of each other.
Pseudo code for Naive Bayes:
Step 1: Import the training dataset.
Step 2: Calculate the coefficient.
Step 3: Preprocess the imported data.
Step 4: Select the classification.
Step 5: Select the Naive Bayes algorithm.
Step 6: Start the process.

Statistical Analysis
For statistical analysis, the SPSS tool is used. In SPSS the data set is prepared using a sample size of 10 the data is analysed using Novel Convolutional Neural Network algorithm and Naive Bayes. To find equality of means, an independent sample test was performed using the SPSS tool.

3. Results

CNN appears to be performing significantly better as compared to Naive Bayes with the value of p=0.075. Accuracy was calculated using equation (1).

Accuracy=a+d/a+b+c+d—(1)

Class 'a' denotes true positive, 'b' denotes false positive, and 'd' denotes false negative are the expected values, whereas True and False are the actual values. The SPSS application is used for statistical analysis. In SPSS, the data is analyzed using Convolution Neural Network and Naive Bayes, and statistical analysis is performed on the two groups using the train and test sets. Table 2 covers the independent samples test categories for equality of variance, and the t-test for equality of mean variations and error differences. Table 3 presents group statistics for mean, standard deviation, and standard error mean for a sample size of ten. Both the significance level and the confidence interval are set at p=0.075. The experimental setup's system architecture is depicted in Figure 1. The Novel Convolutional Neural Network seems to perform much better than Naive Bayes, with a p=0.075 value.

4. Discussions

The data evaluation was performed using IBM SPSS version 21. Data analysis is performed by independent sample T-test and group statistics carried out. The Novel Convolutional Neural Network algorithm which has a mean accuracy is 89% and error rate is11% and Naive Bayes algorithm has mean accuracy of 88% and the error rate is 12%. The values of the CNN algorithm and NB algorithm are analysed statistically and the significant difference found between the two algorithms and better accuracy for CNN machine learning algorithms in predicting the spread of coronavirus was greater than 88%. It was claimed that Naive Bayes is minimal and CNN is the most effective algorithm.(Padmanaban, n.d.; Chen and Jahanshahi 2018). When compared to Naive Bayes, CNN proves to be the most significant and successful. The Naive Bayes algorithm is significantly degraded when irrelevant characteristics exist and drawbacks of the project is that it takes the minimum size of data to predict the accuracy. Many applications can be developed to predict accurately for sensitivity from verified platforms. (Putra et al. 2019) The limitation of the spread of coronavirus is that it is a time-consuming procedure with increased errors in relational analysis, and it requires a higher level of interpretation in order to reduce errors when dealing with complicated cases. One of the drawbacks of the project is that it takes the minimum size of data to predict the accuracy. Research can further try for developing applications to predict accurately for sensitivity from verified platforms.

5. Conclusion

The accuracy value for the prediction of CoronaVirus symptoms using machine learning techniques, Novel Convolutional Neural Network over Naive Bayes is estimated in this proposed

work. Novel Convolutional Neural Network appears to be a more accurate percentage than the Naive Bayes, according to the summary. The Novel Convolutional Neural Network is fast and significant since Naive Bayes is slow and shows improved accuracy for huge data.

Declaration

Conflict of interest

No conflicts of interest in this manuscript

Author contributions

Author M.S K was involved in data collection, data analysis, and manuscript writing. Author E.K.S was involved in conceptualization, data validation, and critical review of a manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this research work successfully.

Funding:

We thank the following organisations for providing financial support that enabled us to complete the study.

1. Kaashiv infotech, Anna Nagar, Chennai,
2. Saveetha University,
3. Saveetha Institute of Medical And Technical Sciences,
4. Saveetha School of Engineering.

6. References

Arora, Parul, Himanshu Kumar, and Bijaya Ketan Panigrahi. 2020. "Prediction and Analysis of COVID-19 Positive Cases Using Deep Learning Models: A Descriptive Case Study of India." Chaos, Solitons, and Fractals 139 (October): 110017.

Awassa, Lamia, Imen Jdey, Habib Dhahri, Ghazala Hcini, Awais Mahmood, Esam Othman, and Muhammad Haneef. 2022. "Study of Different Deep Learning Methods for Coronavirus (COVID-19) Pandemic: Taxonomy, Survey and Insights." Sensors 22 (5). https://doi.org/10.3390/s22051890.

Bhatti, Sidra Ghayour, and Aamer Iqbal Bhatti. 2019. "Adaptive Measurement Noise Covariance Matrix R for JPDAF Based Multitarget Tracking." 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). https://doi.org/10.1109/icecce47252.2019.8940659.

Chen, Fu-Chen, and Mohammad R. Jahanshahi. 2018. "NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion." IEEE Transactions on Industrial Electronics. https://doi.org/10.1109/tie.2017.2764844.

Chisholm-Burns, Marie A., Allison M. Vaillancourt, and Marv Shepherd. 2014. Pharmacy Management, Leadership, Marketing, and Finance. Jones & Bartlett Publishers.

Ghosh, Palash, Rik Ghosh, and Bibhas Chakraborty. 2020. "COVID-19 in India: Statewise Analysis and Prediction." JMIR Public Health and Surveillance 6 (3): e20341.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." Computers and Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.

Martha, London. 2020. The Spread of COVID-19. Core Library.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh,

Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." Environmental Research. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." Carbohydrate Polymers 287 (July): 119349.

Padmanaban, Harini. n.d. "Comparative Analysis of Naive Bayes and Tree Augmented Naive Bayes Models." https://doi.org/10.31979/etd.n7jg-e3uh.

Putra, Yupi Kuspandi, Universitas Hamzanwadi, Muhamad Sadali, and Universitas Hamzanwadi. 2019. "Perbandingan Algoritma Naive Bayes Dan Naive Bayes Berbasis PSO Untuk Analisis Kredit Pada PT. BPR Syariah Paokmotong." Infotek : Jurnal Informatika Dan Teknologi. https://doi.org/10.29408/jit.v2i2.1460.

Rajaraman, Sivaramakrishnan, Jen Siegelman, Philip O. Alderson, Lucas S. Folio, Les R. Folio, and Sameer K. Antani. 2020. "Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays." IEEE Access : Practical Innovations, Open Solutions 8 (June): 115041–50.

Randhawa, Gurjit S., Maximillian P. M. Soltysiak, Hadi El Roz, Camila P. E. de Souza, Kathleen A. Hill, and Lila Kari. 2020. "Machine Learning Using Intrinsic Genomic Signatures for Rapid Classification of Novel Pathogens: COVID-19 Case Study." PloS One 15 (4): e0232391.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." Chemosphere 298 (July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." Chemosphere 298 (July): 134265.

Shah, Saloni, Aos Mulahuwaish, Kayhan Zrar Ghafoor, and Halgurd S. Maghdid. 2021. "Prediction of Global Spread of COVID-19 Pandemic: A Review and Research Challenges." Artificial Intelligence Review, July, 1–22.

Sy, Cheng Len, Pao-Yu Chen, Chun-Wen Cheng, Ling-Ju Huang, Ching-Hsun Wang, Tu-Hsuan Chang, Yi-Chin Chang, et al. 2022. "Recommendations and Guidelines for the Treatment of Infections due to Multidrug Resistant Organisms." Journal of Microbiology, Immunology, and Infection = Wei Mian Yu Gan Ran Za Zhi, March. https://doi.org/10.1016/j.jmii.2022.02.001.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-022-06636-5.

Xu, Youyao, Yizhen Chen, and Xiaoyan Tang. 2020. "Guidelines for the Diagnosis and Treatment of Coronavirus Disease 2019 (COVID-19) in China." Global Health & Medicine. https://doi.org/10.35772/ghm.2020.01015.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." Environmental Research. https://doi.org/10.1016/j.envres.2022.113114.

Tables and Figures

Table 1:It shows the comparison data of the accuracies of both the Convolutional Neural Network and Naive Bayes.

Sno	CNN	NB
1	95.00	93.00

2	94.00	92.00
3	91.23	90.00
4	90.75	89.00
5	90.00	88.45
6	89.15	88.00
7	88.30	87.00
8	87.00	86.57
9	86.00	85.00
10	85.00	83.00

Table 2. Group statistics Result-Convolutional Neural Network(89.64%) and Naive Bayes (88.20%) which has a sample size of 10 for each group. This table also shows standard deviation and standard error rate.

	Groups	**N**	**Mean**	**Std. Deviation**	**Std. Error Rate**
Accuracy	Convolutional Neural Network	10	89.64	3.25	1.029
	Naive Bayes	10	88.20	3.03	0.95

Table 3. Independent sample T-test - Convolutional Neural Network seems to be significantly better than Naive Bayes(p=0.75).

	Levene's test for equality of variances.		**T- test for equality of means**						
ACCURACY	**F**	**Sig.**	t	**df**	**Sig.(2-tailed)**	**Mean difference**	**Std. error difference**	**95% confidence interval of the difference**	
								Lower	**Upper**
Equal variance assumed	0.10	0.075	1.024	18	0.0001	1.44100	1.40733	-1.51570	4.39770
Equal variances			1.024	17.91	0.0001	1.44100	1.40733	-1.51675	4.39875

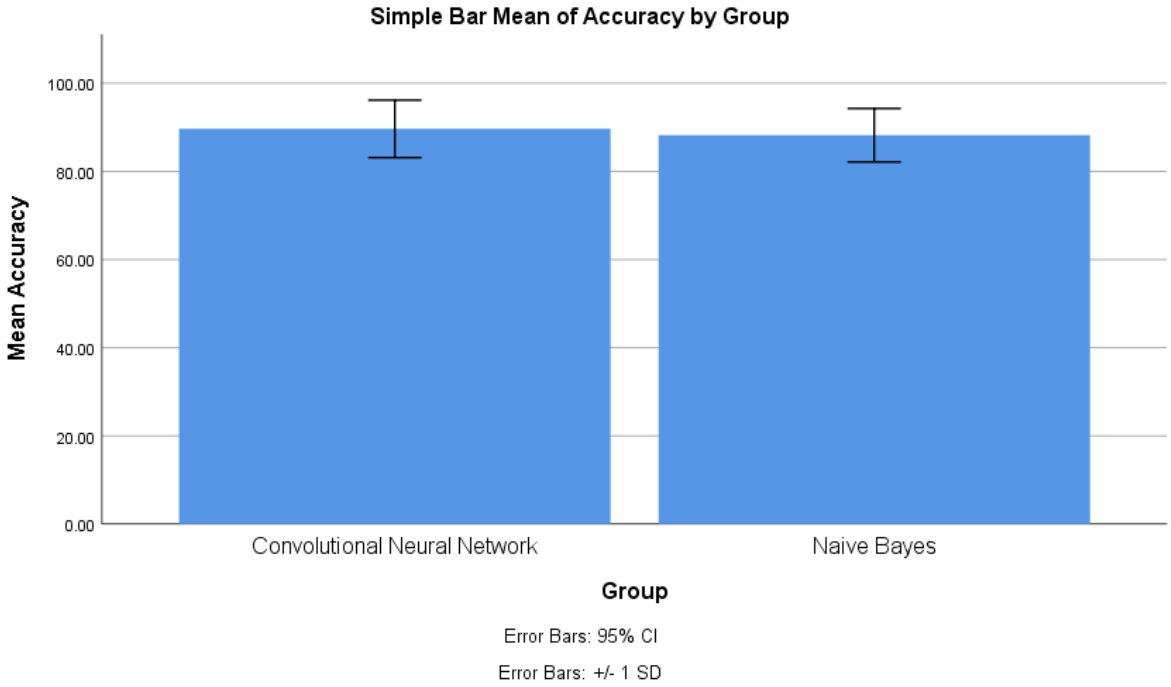

Fig 1.Bar graph comparison on mean accuracy of Convolutional Neural Network (89.64%) and Naive Bayes(88.20%).X axis:Convolutional Neural Network,Naive Bayes,Y axis with mean Accuracy with +/-1 SD