Comparing Novel Multiple Logistic Regression to Bayesian Linear Regression for Accuracy in Death Ratio Analysis of Covid Patients

B. Bharath Kumar Raju[1], N. Deepa[2*]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode:602105.

ABSTRACT

Aim: Examining global data from different countries to estimate the covid infection fatality ratio. **Materials and Methods:** Accuracy of people who are suffering in different countries due to covid19 is analyzed. The sample size is 70 for both the algorithms and both these algorithms belong to supervised learning techniques. Each algorithm consists of 35 sample sizes from which calculate G-power which is 80%. All these algorithms come under supervised learning. **Results:** Calculated the total number of deaths and gained accuracy from the datasets and came to a conclusion that Novel Multiple Logistic Regression has an accuracy of 96% and Bayesian Linear Regression has an accuracy of 74%. The significance for accuracy is determined as p=0.042 (p<0.05). **Conclusion:** Novel Multiple Logistic Regression has a better accuracy value than Bayesian LinearRegression.

Keywords: Novel Coronavirus, Supervised Learning, COVID-19, Death Ratio, Infection Fatality Rate, Novel Multiple Logistic Regression, Bayesian LinearRegression, Machine Learning.

INTRODUCTION

The pandemic which they are facing was first recorded in China wuhan city and from there it spread through many regions called as COVID-19. Evaluating the death ratio of individuals suffering in the crisis with the disease.(Preuss, Kalava, and King 2021).It is used to To know each individual's health condition during the pandemic situation.(Galanis 2021) and is also used to To know the widespread of the disease spread through each region [Citation error]. In machine learning it is used to predict future death rates[Citation error]. The analyses allowed us to see relevant patterns in admission rates in big data(Azevedo et al. 2021). In big data it is used to predict the underlying cause of death. The various applications of both the algorithms are they are used to predict the death rate and to find the similarities in patients symptoms which is used in analyzation of records in hospital.

In the area of Big Data Analytics ,several research papers are available in IEEE and Science Direct. From IEEE Xplore digital library 41 journals are identified 23,863 articles from ScienceDirect 18,700 articles from GoogleScholar 14,603 articles from Springer. The most cited article (Kwekha-Rashid, Abduljabbar, and Alhayani 2021) has a citation of 63.(Panigrahi et al. 2021) has a citation of 91. (Panigrahi et al. 2021; Tan et al. 2021) has a citation of 14 his work is based on pedagogical components accuracy.(Khakharia et al. 2021) has a citation of 20 and discusses the development and external evaluation of patients. Mortality rate of India using statistical neural network models (Dhamodharavadhani, Rathipriya, and Chatterjee 2020)) has a citation of 240.

The proposed work discusses predicting performances of risk of patients suffering with covid19 disease (Her et al. 2021) . Analysis of health records of covid patients who all are suffering with the pandemic disease (Riley et al. 2019)Previously our team has a rich experience in working on various research projects across multiple disciplines(Sathish et al. 2020; Arivazhagan et al. 2020; Pandurangan, Veeraiyan, and Nesappan 2020; Saravanan et al. 2021)

The existing system has a lacunae that it only considers people in a few states but there are more people suffering with that in different places. The existing system also has less accuracy compared to my proposed work. The proposed work has more accuracy than the existing system and has taken data of people from various states. Hence the proposed method aims at comparing algorithms to know which algorithm was giving more accuracy than the Bayesian Linear Regression. The research work aims to provide better accuracy than the existing system and calculate the death ratio of people who are suffering in the pandemic.

MATERIALS AND METHODS

The research work was carried out in a data analytics and image processing lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences where the laboratory facilitates higher configurations to perform better experimental results. Two groups are used for this study with a sample size of (N=35) (Smithson and Shou 2019). The computation is done with G-power with 80% with alpha value 0.05 and beta value is 0.95 with a confidence interval at 95%.

In sample preparation group 1, Novel Multiple Logistic Regression is used and this algorithm is used to train the statistical model at the back end. It gives us good accuracy for large and small data sets. Multiple Logistic Regression descends from logistic regression. Multi-class classification is supported by Multiple Logistic Regression and it is a supervised learning technique. Probability of multinomial is calculated by Multiple Logistic Regression. It has a resistance to overfitting. This model will help us to get accuracy. Death ratio of individuals is provided with this model. Accuracy for Novel Multiple Logistic Regression is 94%. Table 1 represents Novel Multiple Logistic Regression.

In sample preparation group 2, Bayesian Linear regression is used and also trained this model with statistical analysis at the back end. Bayesian Linear Regression is approached from Linear Regression in statistical analysis which lies under Bayesian interface and it is a supervised learning technique. Bayesian Linear regression deals with insufficient data. In the proposed system it is used as a comparison algorithm. The accuracy for Bayesian Linear Bayesian Linear regression allows mixing prior information to guide the statistical inference process.Bayesian Linear Regression accuracy is 64%. Table 2 represents Bayesian Linear regression.

For this proposed model the Jupyter notebook is used as an implementation tool. I have implemented codes in that itself. Hardware configurations of the system which I worked on consists of 8GB RAM and ROM of 1TB HDD+256 SSD

with a processor of 11th gen intel(R) core i5-1135G7 @2. 40 GHZ. There are two groups: group 1 consists of Novel Multiple Logistic Regression and group 2 consists of Bayesian Lasso Regression.

For the proposed system dataset was taken from kaggle. The dataset contains 239 places. It consists of different regions and has a data of total number of cases registered and total number of deaths occured and total number of vaccines available due to covid.

Statistical Analysis

The Analysis was done in SPSS of version 26.0. Independent sample test is carried out for analysis. Independent variables are State/UTs and Unnamed. Dependent variables are Discharge ratio, Active ratio and Death ratio from the dataset. Analysis of health records of covid patients who all are suffering with the pandemic disease (Riley et al. 2019) are collected from the data.

RESULTS

Table 1 represents the Novel Multiple Logistic Regression and it is a supervised learning technique. Every program needs certain libraries so at first libraries are declared which helps for the program to execute. Then the data is splitted into training and testing sets which will be helpful for obtaining our results. At last accuracy is calculated for the respective model and shows Novel Multiple Logistic Regression has better accuracy than Bayesian Linear regression.

Table 2 represents the Bayesian Linear regression. Every program needs certain libraries so at first libraries are declared which helps for the program to execute. Then the data is splitted into training and testing sets which will be helpful for obtaining our results. At last accuracy is calculated for the respective model and shows Novel Multiple Logistic Regression has better accuracy than Bayesian Linear regression.

Table 3 represents various columns each column has its own specifications first column represents total number of sample size and second column represents about the significance of both the algorithm and third column represents about the standard deviation and fourth column represents about the standard error mean. Novel Multiple Logistic Regression has a better accuracy than Bayesian Linear regression.

Table 4 represents the inputs of the spss tool which is used to calculate accuracy for both Multiple Logistic Regression and Bayesian Linear Regression.

Table 5 also represents different columns; each column has its own specifications. Significance and Standard deviation are the two columns which are mentioned in the above table. The significance values for both Novel Multiple Logistic Regression and Bayesian Linear Regression are holding the same value of 0.042.

Figure 1 represents both algorithms of the research work. The algorithms which are used for this research work are Multiple Logistic Regression and Bayesian Linear Regression. The accuracy for both algorithms are obtained from the figure. The accuracy for Multiple Logistic Regression is 96.10 whereas accuracy for Bayesian Linear Regression is 74.86 . From the above mentioned accuracy can

conclude that Multiple logistic regression has a better and accurate accuracy in Bayesian Linear Regression.

DISCUSSION

Based on the results obtained in the Independent sample T-Test analysis, the significant difference between the two groups obtained is 0.042 ($p<0.05$) for the selected dataset. Bayesian Linear Regression algorithm gives mean accuracy is 74%. Novel Multiple Logistic Regression has means accuracy as 96%.

In this paper they discussed the various machine learning algorithms which are used to analyze the cases of individuals who are suffering with the covid19 disease in the pandemic (Kwekha-Rashid, Abduljabbar, and Alhayani 2021) . In this paper the author discussed the usage of data driven approaches for automatic risk prediction and how it is used in calculating the patients death accuracy in covid19 (Shamout et al. 2021). In this paper they discussed how the prediction model helps in management of available resources[Citation error]. In this paper the author discussed the multiple ways of how multivariable modeling is used in health care[Citation error] . The below paper mentions how various machine learning applications and various algorithms are used to deal with covid19(Kwekha-Rashid, Abduljabbar, and Alhayani 2021). When it comes to analyzing death ratio, the accuracy of Multiple Logistic Regression is better when compared with other machine learning and supervised learning algorithms. Multiple Logistic Regression accuracy depends upon training and testing data sets. In this study, accuracy appears to be better than Bayesian Linear Regression and other algorithms. However, the average error appears to be higher in our proposed work which should be decreased.

According to recent estimates, 3 billion people throughout the world do not have access to soap or water at home, 900 million children do not have access to soap and water at school, and up to 40% of health care institutions lack hand hygiene equipment. On average, according to our analysis, the proportion of the population in Africa with access to water and soap hand-washing facilities is 35% (95% CI: 26–44).

The factors that affect the death ratio are that there is a lack of instruments in the medical field and there is no proper diet to be maintained and lack of cleanliness are some of the factors that affect the death ratio. The factors that affect the death ratio are that there are so many states and so much population so when collecting the data from different states there may lead to some miscalculations which lead to a change in the death ratio. These are the limitations which are faced during calculation of death ratio.In future death ratio is used to calculate the death and mortality ratio. With the death ratio one can know each and every detail about the cases, how many had passed away, how many are cured etc., These are the future scope of death ratio and in this way death ratio purpose is served.

CONCLUSION

Novel Multiple Logistic Regression and Linear Regression are both machine learning techniques which use averaging to improve the accuracy. The work shows the death ratio accuracy of people suffering in

the pandemic through covid disease using Multiple Logistic Regression, Linear Regression, Lasso Regression,Logistic Regression and Bayesian Linear Regression. It is found that Multiple Logistic Regression gained more accurate results than Linear Regression, Logistic Regression, Lasso Regression, Bayesian Linear Regression. Hence, it is concluded that Multiple Logistic Regression provides more accuracy when compared with other algorithms.

REFERENCES

1. Arivazhagan, R., N. B. Geetha, P. Sivasamy, P. Kumaran, M. Kumara Gnanamithra, S. Sankar, Ganesh Babu Loganathan, and A. Arivarasan. 2020. "Review on Performance Assessment of Phase Change Materials in Buildings for Thermal Management through Passive Approach." *Materials Today: Proceedings* 22 (January): 419–31.
2. Azevedo, Ana, José Manuel Azevedo, James Onohuome Uhomoibhi, and Ebba Ossiannilsson. 2021. *Advancing the Power of Learning Analytics and Big Data in Education*. IGI Global.
3. Dhamodharavadhani, S., R. Rathipriya, and Jyotir Moy Chatterjee. 2020. "COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models." *Frontiers in Public Health* 8 (August): 441.
4. Galanis, Petros. 2021. "Healthcare Workers during the Covid-19 Pandemic: An Emergency Situation." *Health & Research Journal*. https://doi.org/10.12681/healthresj.26787.
5. Riley, Richard D., Danielle van der Windt, Peter Croft, and Karel G. M. Moons. 2019. *Prognosis Research in Healthcare: Concepts, Methods, and Impact*. Oxford University Press.
6. Saravanan, A., P. Senthil Kumar, Dai-Viet N. Vo, S. Jeevanantham, S. Karishma, and P. R. Yaashikaa. 2021. "A Review on Catalytic-Enzyme Degradation of Toxic Environmental Pollutants: Microbial Enzymes." *Journal of Hazardous Materials* 419 (October): 126451.
7. Sathish, T., R. V. Sabariraj, K. Muthukumar, and S. Karthick. 2020. "Experimental Investigation of Convective Heat Transfer Coefficient on Nano Particles Mixture Used in Automobile Radiator Based on Mass Flow Rate." *Materials Today: Proceedings* 33 (January): 2524–29.
8. Shamout, Farah E., Yiqiu Shen, Nan Wu, Aakash Kaku, Jungkyu Park, Taro Makino, Stanisław Jastrzębski, et al. 2021. "An Artificial Intelligence System for Predicting the Deterioration of COVID-19 Patients in the Emergency Department." *NPJ Digital Medicine* 4 (1): 80.
9. Smithson, Michael, and Yiyun Shou. 2019. *Generalized Linear Models for Bounded and Limited Quantitative Variables*. SAGE Publications.
10. Tan, Liang, Keping Yu, Ali Kashif Bashir, Xiaofan Cheng, Fangpeng Ming, Liang Zhao, and Xiaokang Zhou. 2021. "Toward Real-Time and Efficient Cardiovascular Monitoring for COVID-19 Patients by 5G-Enabled Wearable Medical Devices: A Deep Learning Approach." *Neural Computing & Applications*, July, 1–14.

TABLES AND FIGURES

Table 1. Pseudo code for Novel Multiple Logistic Regression algorithm. Multiple Logistic Regression descends from logistic regression. Multi-class classification is supported by Multiple Logistic Regression. Probability of multinomial is calculated by Multiple Logistic Regression. It has a resistance to overfitting.

Input: Covid Dataset
Output: Accuracy
Step 1: Import and read the dataset.
Step 2: Select some features from the dataset.
Step 3: Generate the parameter.
Step 4: Analyze the dataset by changing the dependent and independent variables.
Step 5: Predict the output in numerical variables.
Step 6: Predict the output by using functions.

Table 2. Pseudo code for BLR approaches linear regression in which statistical analysis within the bayesian interface. Bayesian Linear Regression deals with insufficient data. In my proposed system it is used as a comparison algorithm. The accuracy for Bayesian Linear Regression allows mixing prior information to guide the statistical inference process.

Input: Covid Dataset.
Output: Accuracy.
Step 1: Training set identified
Step 2: Calculate the values of predictor variables of the classes.
Step 4: process the probability for each class.
Step 5: Calculate likelihood of each class.
Step 6: Get the greatest likelihood.

Table 3. For both Novel Multiple Logistic Regression and Bayesian Linear Regression independent sample tests are calculated which has mean accuracy of 96.10 and 74.86. And standard deviation for both the algorithms are calculated and they have a value of 1.834 and 1.556.

	Comparison Techniques	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accurate value	MLR	35	96.10	1.834	.310

	BLR	35	74.86	1.556	.263

Table 4. Rawdata table accuracy for both Bayesian Linear Regression and Multiple Logistic Regression using SPSS statistics.

Group_Id	**MLR**	**BLR**
1	92	75
2	98	74
3	94	76
4	95	77
5	93	74
6	90	75
7	98	76
8	94	76
9	97	75
10	96	78
11	97	79

12	95	74
13	98	75
14	96	79
15	97	75
16	96	74
17	97	73
18	98	75
19	95	77
20	97	78
21	95	76
22	95	75
23	96	77
24	95	76
25	97	74
26	94	75
27	95	76
28	94	77
29	98	74
30	98	79
31	96	77
32	95	76
33	94	75
34	95	77
35	96	76

Table 5. For both Novel Multiple Logistic Regression and Bayesian Linear Regression confidence intervals have been calculated and it has a value of 95%. The value of significance for accuracy is determined as 0.042 ($p<0.05$).

		F	**Sig.**	**T**	**df**	**Sig((2-trailed)**	**Mean diff**	**Std. Error diff**	**Lower**	**Upper**
In % of accurate results	Variance defined	.672	.042	48.562	68	<. 001	19.743	.407	18.932	20.554
	variances not defined			48.562	68.23	<. 001	19.743	.407	18.931	220.556

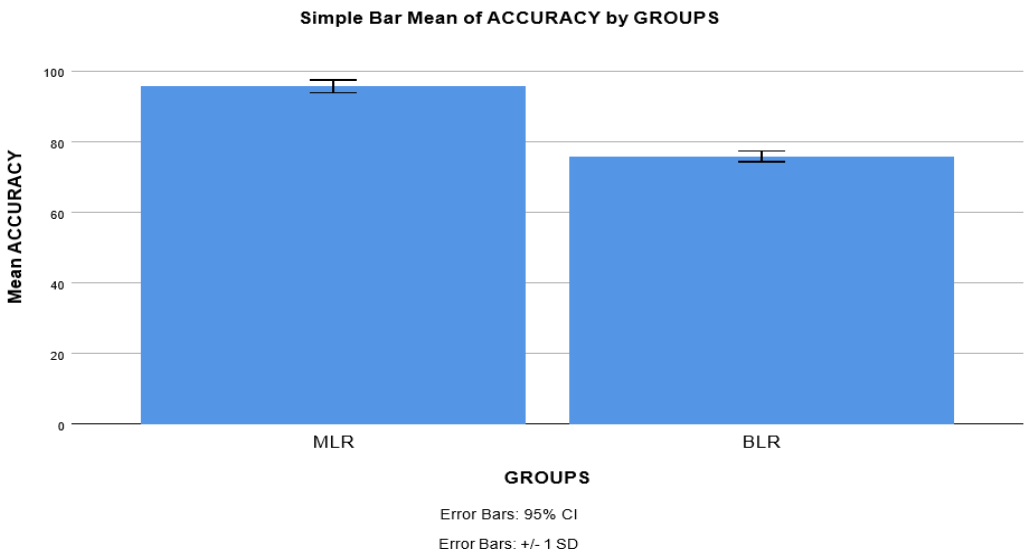

Fig 1. The **s**imple graph is calculated for both Novel Multiple Logistic Regression and Bayesian Linear Regression and MLR has an better accuracy than BLR algorithm.X label : MLR in comparison with BLR Y label: Mean accuracy of detection ± 1 SD.