PREDICTION OF BREAST CANCER USING NOVEL MULTI LAYER PERCEPTRON IN COMPARISON WITH NAIVE BAYES TO IMPROVE ACCURACY

C S. Iyswarya Lakshmi[1], N. Bharatha Devi[2*]

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

Abstract

Aim: The objective of the study is to detect Breast Cancer with the help of Chest Diagnostic image dataset by using Machine Learning based on Naive Bayes Algorithm. To achieve accuracy, a novel augmented dataset classification is used.

Materials and Methods: Accuracy and Loss of Breast cancer detection are performed using the kaggel library. Dataset of 220 chest image dataset with total sample size 20 is used and the two groups are Naive Bayes (N=10) and Novel Multilayer Perceptron (N=10).

Results: This study proved that Naive Bayes achieved better accuracy of 99.66% which is higher, compared to Novel Multilayer Perceptron accuracy of 99.13%. Finally, the Naive Bayes appears significantly better than the Novel Multilayer Perceptron. The statistical analysis shows insignificant differences between the sample groups with p=0.184 ($p<0.05$) and confidence level of 95%.

Conclusion: This study shows that Naive Bayes achieves better accuracy than the Novel Multilayer Perceptron.

Keywords: Naive Bayes, Novel Multilayer Perceptron, Augmented Dataset, Classification, Breast Cancer, Detection, Accuracy.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and technical Science, Saveetha University, Chennai, Tamil Nadu, 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, Tamil Nadu, India. Pincode: 602105.

1. Introduction

Breast cancer disease is an infectious disease and its virus is spread by the cough or sneezes of an infected person. Virus serious symptoms are difficulty in breathing or shortness of breath, chest pain, loss of speech or movement.(Lee et al. 2022) First virus affects the lung, so that the patient isn't able to breathe properly. It is very important in today's world as it is a very effective and low budget method with low risk of spreading virus due to less contact with humans. It can be used in the Medical Field and also in education sites(Yu 2017). Deep learning based methods of detecting Breast cancer positive patients is a promising way to test with low risk of spreading and high accuracy of detection.(Liu and Zhou 2020) Novel Multilayer Perceptrons identify the features of images to analyze its shape, elements in images (Gilanie et al. 2021).

Total Number of articles published in Detecting Breast cancer Virus over the past 5 years is 19 in IEEE Xplore.(Peretti and Amenta 2016) The spread of Breast cancer throughout the world is increasing rapidly. Research has been done to understand the classes of virus to prevent the spreading of virus (Saad et al. 2021). The chest Diagnostic image is the imaging technique for diagnosing Breast cancer, and the Novel Multilayer Perceptron of is showing remarkable results and accuracy for detecting Breast cancer positive and negative patients.(Mojarad and University of Newcastle upon Tyne. School of Electrical and Electronic Engineering 2012) The accurate detection of Breast cancer virus in patients may decrease the death rate of people all over the world and the machine vision is approaching to detect the Breast cancer virus with high accuracy, the budget and maintenance of the method is minimal so users can use it multiple times.(Basmadjian et al. 2022) There have been two classes of image augmented dataset as Breast cancer positive and negative chest images, in case of classification the Novel Multilayer Perceptron has maintained several parameters as similar image size, distributing image dataset in a number of epochs to get high accuracy in results. Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

Based on the survey, Novel Multilayer Perceptron has low performance on accuracy while classifying Breast cancer virus.(Salem Abdull 2011) The aim of the study is to improve the accuracy of Detecting Breast cancer virus and reduce the loss of training, testing Dataset(Baum 1993).

2. Materials and Methods

The study setting of the proposed work was carried out in the Machine Learning laboratory lab at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.No ethical or human sample required. The two groups were identified for the study. Group 1 is the Naive Bayes algorithm and group 2 is the Novel Multilayer Perceptron algorithm. Using G power, 15 sample sizes and a total of 30 sample sizes have been carried out for our study (Pathan, Siddalingaswamy, and Ali 2021). The testing setup proposed system to implement with the following system configuration of hardware and software are: Operating system-Desktop with 64-bit Operating system, RAM-8 GB minimum, CPU-AMD RYZEN %, Memory- 6GB, Platform- Python, IDE- Google Colab, Technology-Novel Multilayer Perceptron,(Bychkov et al. 2022) CPU Core- Quad core. Testing the develop model, the model has been tested using the test augmented dataset. Naive Bayes algorithm applied for classifying the images in two different groups with multiple slices. Resizing the images into the same width, depth, height and augmenting the data to rotate the image randomly in different dimensions.

The image dataset for classifying Breast cancer patients have total 2400 images as 1200 of covid positive chest Diagnostic images and 1200 of Breast cancer negative chest Diagnostic images, from which 80% of dataset will train to detect Breast cancer virus and 20% will test the image dataset accuracy.

Novel Multilayer Perceptron

A Novel Multilayer Perceptron is a class sometimes strictly referring to networks composed of multiple layers of perceptrons. Novel Multilayer Perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer. A Multi Layer Perceptron consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. Multi Layer Perceptron utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguishes Multi Layer Perceptron from a linear perceptron. It can distinguish data that is not linearly separable.

Naive Bayes

Breast cancer is known as the most common invasive cancer type among women and automatic breast cancer detection systems are in demand. Thus, various machine learning and pattern recognition techniques have been proposed to detect breast cancer. One of these techniques is the Bayes classifier. Naive Bayesian is known to be a simple classifier, which is based on the Bayes theorem. There have been so many applications used in literature. In this paper, Naive Bayes (weighted NB) classifier was proposed and its application on breast cancer detection was presented. Several experiments were conducted to evaluate the performance of the weighted NB on the breast cancer database. The experiments were realized with a 5-fold cross validation test. Moreover, various performance evaluation techniques namely sensitivity, specificity and accuracy are considered.

Statistical Analysis

SPSS software is used for statistical analysis of MLP and Multilayer Perceptron based methods. The independent variable is MPLaccuracy and the dependent variable is log loss. The independent T test analyses are carried out to calculate the accuracy of the MLP for both methods

3. Results

Naive Bayes and Novel Multilayer Perceptron algorithm accuracy and loss values are collected. The overall Mean, Standard Deviation and Standard Error Mean are all collected by 10 epochs from both algorithms as shown in Table 1.

Table 2 shows the architecture diagram of Naive Bayes, Image, Hyperspectral imaging of image dataset for pooling images, Reshaping, Merging all flatten dataset and connected into Fully Connected dataset.

Table 3 shows Novel Multilayer Perceptron, Convolution image by pooling the image augmented dataset multiple times (Feature Learning) classifying images by connecting dataset in fully connected dataset. The accuracy of the algorithm can be checked while training and loss of algorithm can be detected by subtraction on accuracy value with 100. The output of an independent sample test got a insignificance of 0.184 which is greater than the level of significance 0.05 shown in Table 3.

The train augmented dataset loss and valuation loss of the Naive Bayes dataset, Independent Sample Effect size applies for Confidence interval of 95% shown in Table 3. The accuracy of Naive Bayes is higher compared to the Novel Multilayer Perceptron.

In Figure 1, the train loss is decreasing as the number of epochs are increasing. Training accuracy and valuation accuracy of the Naive Bayes. The two graphs are used to compare the loss and accuracy of a Naive Bayes. Accuracy of Naive Bayes is 99.66% and Accuracy of Novel Multilayer Perceptron is 99.13% So, the performance of Naive Bayes algorithm is better than Novel Multilayer Perceptron as shown in Figure 1.

4. Discussion

In this research study, the Naive Bayes algorithm has better significance in detecting prediction accuracy than the Novel Multilayer Perceptron algorithm. The statistical insignificant difference obtained between the two groups is $p=0.184$ ($p<0.05$) (Lee et al. 2022). The accuracy result of Naive Bayes is 99.66% and the accuracy of Novel Multilayer Perceptron is 99.13%. Research paper which is related to classification of Breast cancer positive or negative patients (“Risk Prediction in Breast Cancer” 2007).

Prediction of Breast cancer cases using Novel Multilayer Perceptron with Diagnostic, the result obtained in Breast cancer prediction using Novel Multilayer Perceptron with training accuracy of 99% and 98.5% by emphasizing the transfer learning algorithm in disease prediction. Hybrid-covid: a novel hybrid 3D Novel Multilayer Perceptron based on cress-domain adaptation approach for Breast cancer screening from chest Diagnostic images. Experimental results show that the proposed framework can achieve reasonable performances when evaluated on a collected augmented dataset. Notably, it achieved a sensitivity of 98.33%, a specificity of 98.68% and an overall accuracy of 96.91%.Classification of CT images in Breast cancer and non-Breast cancer using Novel Multilayer Perceptron to extract features and multiple classifiers,(Walker and Eeles 2007) The methodology showed an accuracy of 99.79%(“Risk Prediction in Breast Cancer” 2007; McCann 2000), recall of 99.79%, accuracy of 99.80%. The Breast cancer images(Gayathri and Nandhini 2011)with the area under the receiver operating characteristic curve (AUC) value of 0.944, accuracy of 87.5%, sensitivity of 86.9%(Hartmann 2009), specificity of 90.1%, and F1-score of 82.0%. Deep learning method is proposed, which enables the rapid diagnosis of Breast cancer disease by 2019, the accuracy, sensitivity, and specificity of our method reach 98.5%, 99.9%, and 97% . Our team has extensive knowledge and research experience that has translate into high quality publication(Baum 1993) (Devi, Deepa, and Jaisharma 2020)

The limitation of Naive Bayes is that the training and testing dataset is small because the small dataset does not cover all breast cancer symptoms. Due to the small number of datasets, the accuracy of results were also affected. The image resolution of the dataset is low quality. In the future, image datasets will grow in size and large datasets will be collected to know all the symptoms of the breast cancer virus. As the image dataset grows in size, all elements of breast cancer are analyzed with the new Multilayer Perceptron. In the future, high resolution image datasets for breast cancer analysis and classification will improve the resolution of image datasets. High resolution image datasets produce more accurate results.

5. Conclusion

The research study found that the accuracy of Naive Bayes with novel augmented dataset Classification has higher accuracy over the Novel Multilayer Perceptron algorithm. The Naive Bayes algorithm has accuracy of 99.66%, and Novel Multilayer Perceptron algorithm has accuracy of 99.13% . The purpose of this comparative analysis was to find the most accurate machine learning algorithm that could serve as a tool for diagnosing breast tissue according to the predicted results. From this we can conclude that the shopping cart algorithm has the highest accuracy for a particular dataset.

Declarations

Conflicts of Interest

The author declares no conflict of Interest.

Authors Contribution

Author SI was involved in data collection, data analysis, manuscript writing. Author NBD was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

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 work successfully.

Funding

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

1.CK Technology Pvt. Ltd.
2.Saveetha University.
3.Saveetha Institute of Medical and Technical Sciences.
4.Saveetha School of Engineering.

6. REFERENCES

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

Basmadjian, Robert B., Shiying Kong, Devon J. Boyne, Tamer N. Jarada, Yuan Xu, Winson Y. Cheung, Sasha Lupichuk, May Lynn Quan, and Darren R. Brenner. 2022. "Developing a Prediction Model for Pathologic Complete Response Following Neoadjuvant Chemotherapy in Breast Cancer: A Comparison of Model Building Approaches." *JCO Clinical Cancer Informatics* 6 (February): e2100055.

Baum, M. 1993. "Prognosis and Prediction for Early Breast Cancer." *Adjuvant Therapy of Breast Cancer IV*. https://doi.org/10.1007/978-3-642-84745-5_12.

Bychkov, Dmitrii, Heikki Joensuu, Stig Nordling, Aleksei Tiulpin, Hakan Kücükel, Mikael Lundin, Harri Sihto, et al. 2022. "Outcome and Biomarker Supervised Deep Learning for Survival Prediction in Two Multicenter Breast Cancer Series." *Journal of Pathology Informatics* 13 (January): 9.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

Devi, T., N. Deepa, and K. Jaisharma. 2020. "Client-Controlled HECC-as-a-Service (HaaS)." In *Lecture Notes on Data Engineering and Communications Technologies*, 312–18. Cham: Springer International Publishing.

Gayathri, A., and V. Nandhini. 2011. "HVS Based Enhanced Medical Image Fusion." *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-642-25734-6_156.

Gilanie, Ghulam, Usama Ijaz Bajwa, Mustansar Mahmood Waraich, Mutyyba Asghar, Rehana Kousar, Adnan Kashif, Rabab Shereen Aslam,

Muhammad Mohsin Qasim, and Hamza Rafique. 2021. "Coronavirus (COVID-19) Detection from Chest Radiology Images Using Convolutional Neural Networks." *Biomedical Signal Processing and Control* 66 (April): 102490.

Hartmann, Lynn C. 2009. "Benign Breast Disease: Toward Molecular Prediction of Breast Cancer Risk." https://doi.org/10.21236/ada552165.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Lee, Inki, Ilhan Lim, Byung Hyun Byun, Byung Il Kim, Chang Woon Choi, Kyo Chul Lee, Choong Mo Kang, et al. 2022. "The Prediction of HER2-Targeted Treatment Response Using Cu-Tetra-Azacyclododecanetetra-Acetic Acid (DOTA)-Trastuzumab PET/CT in Metastatic Breast Cancer: A Case Report." *Journal of Breast Cancer*, January. https://doi.org/10.4048/jbc.2022.25.e5.

Liu, Dingxie, and Kehua Zhou. 2020. "BRAF/MEK Pathway Is Associated With Breast Cancer in ER-Dependent Mode and Improves ER Status-Based Cancer Recurrence Prediction." *Clinical Breast Cancer*. https://doi.org/10.1016/j.clbc.2019.08.005.

McCann, Jenny. 2000. "Mammographic Prediction of Early Breast Cancer Survival." *Breast Cancer Research*. https://doi.org/10.1186/bcr-2000-66658.

Mojarad, Shirin Ameiryan, and University of Newcastle upon Tyne. School of Electrical and Electronic Engineering. 2012. *A Reliable Neural Network-Based Decision Support System for Breast Cancer Prediction*.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Pathan, Sameena, P. C. Siddalingaswamy, and Tanweer Ali. 2021. "Automated Detection of Covid-19 from Chest X-Ray Scans Using an Optimized CNN Architecture." *Applied Soft Computing* 104 (June): 107238.

Peretti, Alessandro, and Francesco Amenta. 2016. "Breast Cancer Prediction by Logistic Regression with CUDA Parallel Programming Support." *Breast Cancer: Current Research*. https://doi.org/10.4172/2572-4118.1000111.

"Risk Prediction in Breast Cancer." 2007. *Hereditary Breast Cancer*. https://doi.org/10.3109/9781420020366-6.

Saad, Waleed, Wafaa A. Shalaby, Mona Shokair, Fathi Abd El-Samie, Moawad Dessouky, and Essam Abdellatef. 2021. "COVID-19 Classification Using Deep Feature Concatenation Technique." *Journal of Ambient Intelligence and Humanized Computing*, March, 1–19.

Salem Abdull, Mohamed A. 2011. *Data Mining Techniques and Breast Cancer Prediction: A Case Study of Libya*.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806.

Walker, Lisa, and Rosalind Eeles. 2007. "Risk Prediction in Breast Cancer." *Hereditary Breast Cancer*. https://doi.org/10.3109/9781420020366-3.

Whangchai, Niwooti, Daovieng Yaibouathong,

Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Yu, Mengjie. 2017. *Breast Cancer Prediction Using Machine Learning Algorithm*.

Tables and Figures

Table 1. Predicted Accuracy of Breast Cancer Prediction for 48 different sample sizes.

GROUP	Algorithms	Average accuracy
1	Novel Naive bayes	99.66
2	Multi Layer Perceptron	99.13

Table 2. Statistical analysis of novel naive bayes and Multi Layer Perceptron. Mean accuracy, Standard deviation, and standard error values are obtained for 48 sample datasets.

	Algorithms	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	Novel naive bayes	10	99.66	.26282	.08311
	Multi Layer Perceptron	10	99.13	.16732	.05291

Table 3. Independent sample T-test with a confidence interval at 95% and level of significance as 0.05.

	Levene's Test for Equality of Variances		T-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Differences	Std.Error Differences	95% Confidence Interval of the Differences Lower | Upper	
Equal Variances assumed	1.906	.184	29.911	18	.000	2.9470	.09853	2.7400	3.115399
Equal Variances not assumed	-	-	29.911	15.266	.000	2.9470	.09853	2.7373	3.115668

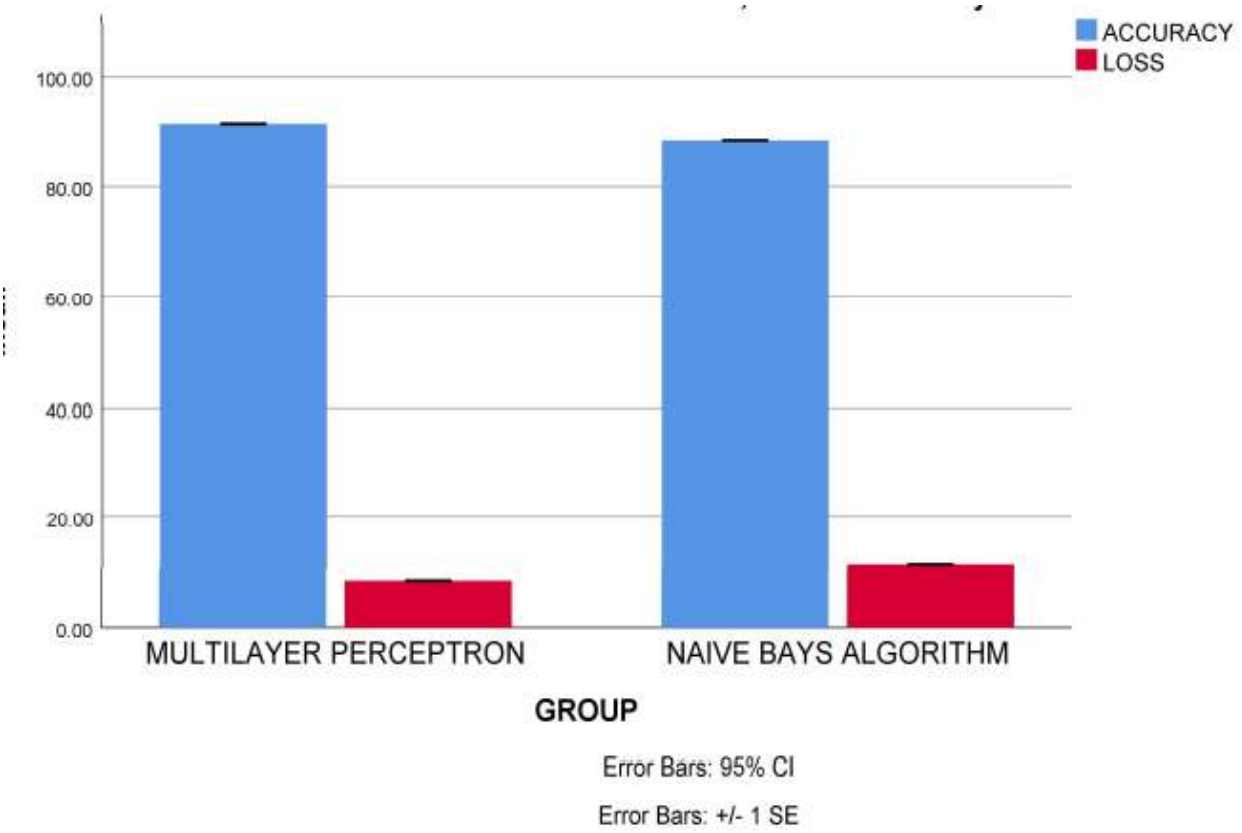

Fig. 1. Comparison of Novel MultiLayer Perceptron in terms of means and accuracy. The mean accuracy of the Novel MultiLayer Perceptron is better than Naive Bayes. X-axis: Novel MultiLayer Perceptron vs Naive Bayes, Y-axis: Mean accuracy. Error Bar +/-1 SD with 95% Cl.