IMPROVED ACCURACY FOR PREDICTION OF LEAF WETNESS USING LOGISTIC REGRESSION ALGORITHM COMPARED WITH K-NEAREST NEIGHBOUR ALGORITHM

T. Manvitha[1], K. Sashi Rekha[2*]

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

Abstract

Aim: The main aim of this research work is to compare the accuracy percentage of leaf wetness predicted by the Novel Logistic Regression algorithm to that predicted by the K-Nearest Neighbour algorithm using meteorological data. **Materials and methods:**The accuracy of leaf wetness prediction was evaluated using Novel Logistic Regression and K-Nearest Neighbour algorithms with a sample size of 20 at different times. **Results:** Novel Logistic Regression has a significantly better accuracy percentage (91.89%) compared to K-Nearest Neighbour accuracy (79%). Between Novel Logistic Regression and K-Nearest Neighbour, The statistical significance difference $p=0.07$ ($p<0.05$) independent sample T-test value state that the results in the study are insignificant. **Conclusion:** The K-Nearest Neighbour method fared much worse than Novel Logistic Regression.

KEYWORDS: Novel Logistic Regression, K-Nearest Neighbour Algorithm, Machine Learning Algorithms, Relative Humidity, Classification, Prediction, Leaf Wetness.

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

The focus of this research is to compare the Novel Logistic Regression method and the K-Nearest Neighbour algorithm in order to enhance the accuracy % of forecasting leaf wetness. Leaf moisture, or the presence of water droplets on the surface of the leaves, is necessary for the development of plant-fungal diseases (Kerstiens 1996). The leaf wetness mainly occurred because of temperature in the atmosphere and relative humidity (Pais 1982). With the help of these two parameters, the accuracy of the prediction of leaf wetness will be improved. This paper compares the Novel Logistic Regression algorithm against the K-Nearest Neighbour technique. Visual inspection, threshold or model structure, and visualisations are some of the current approaches for establishing the existence of leaf wetness (Uy et al. 2021). The direct measuring strategy is not a practical option for implementation since it is momentary, requires manual workers, and is subject to spectator perception. (Sutton, Reis, and Baker 2008). The application of the prediction of Leaf Wetness is predicting the developed fungal pathogens in the leaf.

Researchers have debated the origins, impacts, and trends of leaf wetness since. On leaf wetness prediction, there were 228 publications published in IEEE Explore and 198 published articles in Google Scholar (Uy et al. 2021). The leaf wetness prediction using various ML algorithms is presented in this part. It is difficult to extract relevant information from noisy data. To do this objective, ML approaches are required(Matti et al. 2021). (Karadag 2017) used a number of Machine Learning methods in conjunction with spectral reflectance values to detect fusarium disease in peppers(Knapik et al. 2020). The author recommends the article with the greatest citations and was most visited for this research work (R. M. Gillespie and Gary 2009; M. B. Gillespie 2016) and also focused on Novel Logistic Regression with an accuracy of 90%. The learning algorithm of a machine that executes Logistic Regression in a Novel Way focuses mostly on classification (Sutton, Reis, and Baker 2008). When a sensor fails, reporting is stopped, and real-time monitoring is insufficient to prevent infection when leaf moisture is predicted, according to previous studies. Finally, (Nishiura et al. 2020)Machine that does Logistic Regression in a Novel Way The learning algorithm focuses primarily on classification.

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). The previous research work shows a research gap in that sensor failure stops reports, and real-time monitoring is unable to prevent infection when leaf moisture is predicted. Finally, (Nishiura et al. 2020) the implications of the research and the research's future trajectory are depicted in the concluding part. Machine learning can be used to educate people how to forecast how moist their leaves are. The Novel Logistic Regression technique is used in this proposed system to improve accuracy in predicting leaf wetness.

2. Materials And Methods

The suggested work was accomplished in Saveetha School of Engineering's Open Source Lab. For the investigation, a total of two groups have been identified**.** In Group 1, K-Nearest Neighbour algorithm is employed, while in Group 2, Novel Logistic Regression is used. The novel Logistic Regression technique and the K-Nearest Neighbour algorithm were repeated at varied intervals with a sample size of 20, a confidence interval of 95 percent, and a pretest power of 80 percent(Peng et al. 2021).

The Meteorological dataset was utilised as the real-time dataset. Meteorological dataset was used as the input dataset for the proposed research. csv file downloaded from kaggle.com(Giannakopoulou et al. 2021)."Relative Humidity (rh)", "Temp (temperature)", "atmos pres (atmospheric pressure)", "ceil hgt (ceil height)", "ws (wind speed)", "wd (wind density)", "dew point", and "visibility" were the key attributes used to enhance accuracy (percent). Above all, it pertains to the meteorological dataset's attribute description.

Following the collection of the dataset, the procedure for preprocessing was implemented.. Both random effect model and incomplete information in the dataset were removed during the cleaning process. Following cleaning, feature extraction was utilised to vectorize the data, which included strings, words, and characters that were converted to 0 and 1 values. The data set obtained was excellent for evaluating the machine learning method because it was free of blank and empty values. The dataset was split into two halves during preprocessing and explored as a 25 percent of overall testing set and a 75percent of overall training set.

Logistic Regression

One of the most prominent supervised machine learning algorithms is logistic regression. Logistic Regression is primarily concerned with the classification of instances 0 and 1 as well as pass or fail. It helps predict dependent variables that are stratified using many parameters. It converts projected values to probabilities using the Novel sigmoid function and decides which variables to adopt as a result and which to refuse .

```
Import numPy as np
Import pandas as pd
Import seaborn as sb
Import matplotlib.pyplot as plt
Import sklearn
From pandas import Series, DataFrame
From pylab import rcParams
From sklearn import preprocessing
From sklearn.linear_model import LogisticRegression
From sklearn.model_selection import train_test_split
From sklearn imports metrics
From sklearn.metrics import classification_report
df=predocs'('/final dataset 1.csv')
df.head()
df.isnull().sum()
fcc_location=pd.get_dummies(df['atmos_pres'],drop_first=True)
fcc_location.head()
df.shape
df.info()
df.head()
sb.heatmap(df.corr())
x=df.iloc[:,1].values
y=df.iloc[:,1].values
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LgisticRegression
LogReg = LogisticRegression()
LogReg.fit(x_train,y_train)
y_predicted=LogReg.predict(x_test)
percentage = score*100
print(f'Accuracy: {round(percentage,2)}%')
```

Fig. 1. Pseudocode of Logistic Regression

K-Nearest Neighbour algorithm

K Nearest Neighbour (KNN) is a simple, uncomplicated, and versatile machine learning technique.Handwriting recognition, picture recognition, and video recognition are just a few of the applications that use it. When obtaining labelled data is prohibitively expensive or unattainable, KNN can tackle a wide spectrum of high-accuracy prediction challenges. KNN is a simple algorithm that learns an unknown function with the requisite precision and accuracy by using the target stored process local minimum. In addition, the method identifies the position of an unknown input's neighbourhood, as well as its range or distance from it and other criteria. It is predicated on the notion of "learning," in which the algorithm selects the most effective method.

```
Import numpy as num
Import matplotlib.pyplot as mtp
Import pandas as pd
Data_set=pd.read_csv('/final dataset 1.csv')
x=data_set.iloc[:,[2,3]].values
y=data_set.iloc[:,4].values
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25, random_state=0)
from sklearn.preprocessing import standardScalar
st_x=standardscaler()
x_train=st_x.fit_tranform(x_train)
x_test=st_x.transform(x_test)
from sklearn.neighbors import KNeighborsClassifier
classifier=kneighborsClassifier(n_neighbors=5,metric='minkowski',p=2)
classifier.fit(x_train,y_train)
y_pred=classifier.predict(x_test)
from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_pred)
percentage=score*100
print('f Accuracy: { round(percentage,2)}%)
```

Fig. 2. Pseudocode for KNN algorithm

Google collab was utilised to assess the Novel Logistic Regression and RF techniques. An Intel Core i3 processor and 8GB of RAM were included in the system configuration. A 64-bit operating system was included in the system. The operating system Windows 10 is included in the software configuration.

Statistical analysis

The statistical analysis was carried out using the Statistical Package for the Social Sciences (SPSS)(Windrim et al. 2018) and Google colab. The significant factors include relative humidity, temperature, air pressure, ceiling height, ws, wd, dew point, and visibility. Accuracy was the constant variable. To evaluate the efficiency of the proposed algorithm, an Independent Sample t-test was used.

3. RESULTS

Table 1 demonstrates the accuracy of the Logistic Regression and K-Nearest Neighbour Algorithms, which were run in Google collab with a sample size of 20 at different times. The K-Nearest Neighbour Technique algorithm appears to be less accurate than Logistic Regression. In Table 2, we utilised an independent sample T-test to compare the accuracy of Logistic Regression and K-Nearest Neighbour Algorithms, and we observed a statistically significant difference (p=0.07 with a 95 percent confidence interval), showing that our hypothesis is valid. The average accuracy difference was calculated at 13.18600.

Table 3 shows the results. The statistical analysis of ten samples yielded a standard deviation of 4.92822 and a standard error of 1.55844 for Logistic Regression. The KNN method yielded a standard deviation of 2.68501 and a standard error of 0.84907. The related output values (dependent variables) varied in response to changes in the input values (independent variables).

Figure 3 depicts a bar graph analysis based on two algorithms' accuracy. The KNN algorithm and the Novel Logistic Regression algorithm's mean accuracies. The pseudocode for Logistic Regression is shown in Figure 1, while the pseudocode for KNN is shown in Figure 2.

4. DISCUSSION

The accuracy of forecasting leaf moisture was investigated using Logistic Regression and the KNN method in this study. When compared to KNN, Logistic Regression appears to have a higher accuracy (91.89 percent). The Novel sigmoid function enhances accuracy by transferring predictions to the probability of Leaf Wetness based on the dataset's Relative Humidity and temperature variables. The results reveal that the Logistic Regression and KNN algorithms have statistically significant differences. In this work, researchers employed a combination of machine

learning algorithms and spectral reflectance data to detect fusarium disease in peppers (Karadag 2017). Our organisation is committed to conducting elevated, substantial proof investigation and has had success in a number of fields(Confalonieri et al. 2011; Bregaglio et al. 2011). We hope this proposed work adds to this rich legal system. Relative Humidity and Temperature were the two characteristics that had the greatest impact on the accuracy percentage. When compared to earlier study publications, Logistic Regression looks to be more accurate.
One of the features in the Meteorological dataset utilised for forecasting the accuracy of leaf wetness is "Ceil hgt," which does not predict the exact accuracy of leaf wetness and is not supported for larger datasets, which is a restriction of the proposed work. If the dataset contains other parameters like precipitation and evaporation, there may be a potential to forecast leaf wetness more accurately in future studies. Future work could be used in universities, colleges, and online course platforms to provide students with thorough guidance in determining the accuracy of the Leaf Wetness.

5. CONCLUSION

In this proposed work, when compared to the KNN method, the accuracy of prediction of leaf wetness using Novel Logistic Regression based on temperature and relative humidity in the atmosphere is enhanced by 91.89%.

DECLARATIONS

Conflict of interests

No conflict of interest in this manuscript.

Authors Contribution

Author TM was in charge of data collection, data analysis, and manuscript writing. The manuscript's conceptualization, data validation, and critical review were all done by author KSR.

Acknowledgements

The authors would like to express their gratitude to Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science (Formerly Saveetha University) for providing the necessary support to complete this work effectively.

Funding:

We are grateful to the following organisations for their financial contributions, which enabled us to complete the study

1. Qbec Infosol.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Bregaglio, Simone, Marcello Donatelli, Roberto Confalonieri, Marco Acutis, and Simone Orlandini. 2011. "Multi Metric Evaluation of Leaf Wetness Models for Large-Area Application of Plant Disease Models." *Agricultural and Forest Meteorology*. https://doi.org/10.1016/j.agrformet.2011.04.003.

Confalonieri, Roberto, Simone Bregaglio, Alexandra Stella Rosenmund, Marco Acutis, and Igor Savin. 2011. "A Model for Simulating the Height of Rice Plants." *European Journal of Agronomy*. https://doi.org/10.1016/j.eja.2010.09.003.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. "Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics." *Bioresource Technology* 350 (April): 126904.

Giannakopoulou, Olga, Kuang Lin, Xiangrui Meng, Mei-Hsin Su, Po-Hsiu Kuo, Roseann E. Peterson, Swapnil Awasthi, et al. 2021. "The Genetic Architecture of Depression in Individuals of East Asian Ancestry: A Genome-Wide Association Study." *JAMA Psychiatry* , September. https://doi.org/10.1001/jamapsychiatry.2021.2099.

Gillespie, Michael Boyce. 2016. "B.A.D. (Black Abstraction Dreaming): A Conversation with Kevin Jerome Everson." *Black Camera*. https://doi.org/10.2979/blackcamera.8.1.0155.

Gillespie, Richard M., and Joshua L. Gary. 2009. "Principles of Fractures and Dislocations." *Parkland Trauma Handbook*. https://doi.org/10.1016/b978-0-323-05226-9.50042-6.

Karadag, Engin. 2017. *The Factors Effecting Student Achievement: Meta-Analysis of Empirical Studies*. Springer.

Kerstiens, G. 1996. *Plant Cuticles: An Integrated Functional Approach*. Taylor & Francis.

Knapik, Derrick M., Jensen G. Kolaczko, Robert J. Gillespie, Michael J. Salata, and James E. Voos. 2020. "Complications and Return to Activity After Arthroscopic Repair of Isolated Type II SLAP Lesions: A Systematic Review Comparing Knotted Versus Knotless Suture Anchors." *Orthopaedic Journal of Sports Medicine* 8 (4): 2325967120911361.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish,

T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. "A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123594.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. "A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations." *Journal of Water Process Engineering*. https://doi.org/10.1016/j.jwpe.2022.102712.

Matti, Bashar, Fairleigh Reeves, Matthew Prouse, and Kamran Zargar-Shoshtari. 2021. "The Impact of the Extent and Location of Positive Surgical Margins on the Risk of Biochemical Recurrence Following Radical Prostatectomy in Men with Gleason 7 Prostate Cancers." *The Prostate*, September. https://doi.org/10.1002/pros.24240.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. "Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites." *Bioinorganic Chemistry and Applications* 2022 (January): 8559402.

Nishiura, Hiroshi, Sung-Mok Jung, Natalie M. Linton, Ryo Kinoshita, Yichi Yang, Katsuma Hayashi, Tetsuro Kobayashi, Baoyin Yuan, and Andrei R. Akhmetzhanov. 2020. "The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020." *Journal of Clinical Medicine Research* 9 (2). https://doi.org/10.3390/jcm9020330.

Pais, Amélia Pinto. 1982. *Para compreender "Os Lusíadas."*

Peng, Cindy Q., Kenny D. Lawson, Mark Heffernan, Geoff McDonnell, Danny Liew, Sean Lybrand, Sallie-Anne Pearson, et al. 2021. "Gazing through Time and beyond the Health Sector: Insights from a System Dynamics Model of Cardiovascular Disease in Australia." *PloS One* 16 (9): e0257760.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102102.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." *Journal of Healthcare Engineering* 2022 (February): 2761847.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123798.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model." *Journal of Healthcare Engineering* 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. "Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management." *Science of The Total Environment*. https://doi.org/10.1016/j.scitotenv.2022.153897.

Sutton, Mark, Stefan Reis, and Samantha Baker. 2008. *Atmospheric Ammonia: Detecting Emission Changes and Environmental Impacts. Results of an Expert Workshop under the Convention on Long-Range Transboundary Air Pollution*. Springer Science & Business Media.

Uy, Kim, Oy Heang, Vanna Keo, Yong June Kim, Jim Gollogly, and Ling Hong Lee. 2021. "Surgical Management of Patients with Snakebite-Related Musculoskeletal Complication-A Single Institution Experience in Cambodia." *World Journal of Surgery*, September. https://doi.org/10.1007/s00268-021-06307-w.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. "Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach." *International Journal of Energy Research*. https://doi.org/10.1002/er.7780.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. "A

Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants.” *Environmental Pollution* 301 (May): 119035.

Tables And Figures

Table 1**.** Improved accuracy for predicting the Leaf Wetness (Logistic Regression accuracy of 91.89% and KNN accuracy of 79%)

Iteration No.	Logistic Regression Accuracy (%)	KNN Accuracy (%)
1	91.89	79.00
2	91.84	78.84
3	91.79	78.23
4	91.78	78.15
5	91.76	78.09
6	91.24	77.76
7	91.05	77.54
8	90.46	77.35
9	90.23	76.72
10	90.14	76.26

Table 2. T-test with independent samples The confidence interval for the dataset was set to 95%, while the level of significance was set to 5% (p<0.05) (With a p=0.07 value, Logistic Regression looks to perform massively better than K-Nearest Neighbour)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the difference	
		F	sig.	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	Lower	Upper
Accuracy	Equal Variances assumed	3.603	0.07	7.43	18	.001	13.18	1.77	9.45	16.91
	Equal variances not assumed			7.43	13.9	.001	13.18	1.77	9.45	16.99

Table 3. KNN group statistical analysis with logistic regression. After 10 iterations, the mean accuracy value, standard deviation, and standard error mean for the Logistic Regression and KNN algorithms are obtained. According to the findings, the Logistic Regression algorithm outperformed the KNN algorithm.

Group	N	Mean	Std.deviation	Std.Error Mean
LR	10	86.6930	4.92822	1.55844
KNN	10	73.5070	2.68501	0.84907

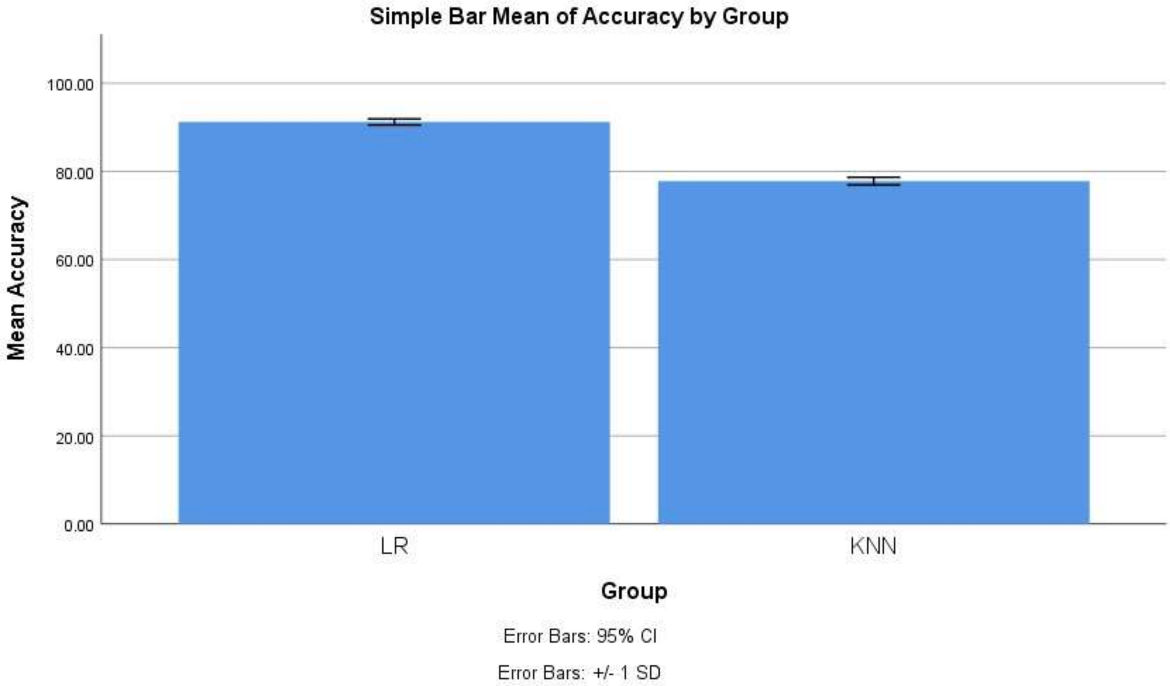

Fig. 1. In terms of mean accuracy, Logistic Regression and KNN are compared. The mean accuracy of Logistic Regression is higher than that of KNN, and the standard deviation is slightly lower than that of GLM. X Axis: Logistic Regression vs. KNN Algorithm; Y Axis: Mean detection accuracy within 1 SD.