Comparing Logistic Regression to the K-nearest Neighbors (KNN) technique, A Novel Pattern Discovery Based Human Activity Recognition

S.Ritesh Reddy[1], Devi.T[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:

The main objective of this research study is to improve accuracy for Human Activity Recognition using Data Analysis Techniques. **Materials and Methods:** A Logistic Regression with sample size 10 and K-nearest Neighbors (KNN) Algorithm with sample size of 10. It was iterated at different times predicting the accuracy percentage of human Activity Recognition. **Results:** Human Activity Recognition utilizing Novel Logistic Regression 95.52% accuracy compared with K-nearest Neighbors (KNN) Algorithm 90.52% accuracy. Logistic Regression seems to perform essentially better compared to K-nearest Neighbors (KNN) Algorithm ($p=0.42$) ($p<0.05$). The Logistic Regression algorithm in computer vision appears to perform significantly better than the K-nearest Neighbors (KNN) Algorithm algorithm. **Conclusion**: Within this human Activity Recognition Logistic Regression has more accuracy over K-nearest Neighbors (KNN) Algorithm

Keywords: NovelLogistic Regression, Data Analysis, Human Activity Recognition, K-nearest Neighbors (KNN) Algorithm , Accuracy.

Introduction:

An action is a set of movements of the human body and these movements tend to be sequential.Action recognition intuitively means recognizing a set of actions(Roy, Girdzijauskas, and Socolovschi 2021). From an outlook of computer vision, action recognition refers to learning a set of video sequences to identify the sequence of movements associated with a particular action and using the knowledge obtained to predict a future action based on the movements associated(Grieco et al. 2021). Action recognition is a key component in many applications, namely, human-computer interaction, surveillance, video analysis etc(Han 2021). Recognizing an action helps the system to summarize an entire event. For instance, action recognition is becoming an integral part of sports systems for summaries as seen in Han et al(Daher et al. 2021).It is additionally pertinent for typical activities for the most part connected with clinical issues help and furthermore it is likewise appropriate for gaming reason as well.

Estimation of predicting human activity recognition in machine learning has been administered by researchers and 45 related articles in IEEE and 30 articles are published within the research gate. Act prediction may be a probabilistic process of inferring ongoing activities from videos only containing onsets of the activities. The goal is to enable early recognition(Fan and Gao 2021) of unfinished activities as opposed to after-the-fact classification of completed activities.this is proposed to implement(Fu 2015) a systematic human activity recognition method to recognize basic activities (BA) and transitional activities in a continuous sensor data stream.In this work both a novel method in process prediction, which has largely relied on explicit process models, and also a novel application of deep learning methods.(Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020)

Previously our team has a rich experience in working on various research projects across multiple disciplines(Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).Existing systems(Ortiz 2015) have drawbacks of poor image resolution, less accuracy, poor lighting and low contrast, higher computational cost, lack of standards, more time consuming process. Therefore an approach is required which can handle the above issues and helps in (Ahad, Antar, and Ahmed 2021)predicting activity of a human. In this article it is proposed to design a system which will predict Activity recognition of humans through detection and tracking by using Data Analysis(Chen and Nugent 2019). Techniques thereby increasing accuracy and precision by reducing computational cost.A Novel Pattern Discovery Based Human Activity Recognition Using Data Analysis Techniques By Comparing Novel Logistic Regression Over K-nearest Neighbors (KNN) algorithm

Methods and Materials

This research study was carried out in the Image Processing Laboratory at the Department of Computer Science and Engineering, Saveetha School of Engineering, Chennai. This research study uses two groups,(Strickland, n.d.) the LR Algorithm, and the K-nearest Neighbors (KNN) algorithm. Each sample size was predicted using the g-power tool with version 3.1.10 and resulting in 64 sample sizes with 80% of g-power values and the threshold two tailed significant value is set to 0.05 and the confidence interval as 95%

The dataset used for this experiment is obtained from the Kaggle open access dataset.The Pre- test analysis has been prepared using clinical.com by having a G power of 80% and threshold 0.05%, CI 95% mean and standard deviation. Group 1 is the Novel Logistic Regression algorithm with the sample size of 10 and the KNN algorithm is group 2 with sample size of 10 and they are compared for more accuracy score and precision score values for choosing the best algorithm.Sample size has been calculated and it is identified that 10 samples/ group in total 20 samples with a standard deviation for Novel Logistic Regression algorithm = 0.4198 and KNN algorithm = 0.30102

Logistic Regression:

Logistic Regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variables in dichotomous, which means there would be only two possible classes

Pseudocode for the LR algorithm

Input:
Training algorithm L (logistic regression)
Sample matrix X
Labels vector Y=[1,.....K]
Initial regression parameters vector θ_i
Main:
For i=1:K
Create a new binary vector $y_i$for each label where y_i=1
if it belongs to the label and y_i=0 if it does not belong.
Apply L to X to findθ_i

Output:
θ_iParameters vector for each regressor

K-nearest Neighbors (KNN) Algorithm:

The K-nearest neighbors (KNN) algorithm is a simple , supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand,but has a major drawback of becoming significantly slows as the size of the data in use grows

Pseudocode for the KNN Algorithm:

1. Calculate "d(x, xi)" i=1, 2, ..., n; where d denotes the Euclidean distance between the points.

2. Arrange the calculated n Euclidean distances in non-decreasing order.

3. Let k be a +ve integer, take the first k

distances from this sorted list. 4. Find those k-points corresponding to these k-distances.

5. Let k_i denotes the number of points belonging to the ith class among k points i.e. k >= 0

6. If ki >kj Vij then put x in class i.

Statistical Analysis

All analyses were performed using Software Package for Social Sciences (SPSS) for descriptive statistical analyses such as were carried out for each group.The analysis was done using IBM SPSS version 21. It is a statistical software tool used for data analysis. For both proposed and existing algorithms 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing accuracy. The test was performed to compare variables across study groups. Hence independent variables in this study are sitting, standing, walking and laying etc. The dependent variable is prediction of accuracy of Human activity recognition. Independent sample t-test is performed to compare performance of both group-1 and group-2 using Novel Logistic Regression and KNN algorithms.

Result

The datasets is provided by google colab, The KNN machine learning algorithm and the Logistic Regression machine learning algorithm are compared with 10 samples by applying various 70% of training data and 30% of testing datasets by varying the number of records of the dataset and the outcomes are depicted in Table 2 and the dataset consists of 1112 rows where the accuracy of both the KNN and the LR algorithm are obtained for 10 samples (iterations). In Table 1, observed after performing statistical analysis of 10 samples, Novel Logistic Regression obtained .4198 standard deviations with .18776 standard error while the KNN algorithm obtained .301202 standard deviations with .13476 standard error. The independent t-test two tailed significance value is smaller than 0.05 showed that our hypothesis holds a good value. For changes in the input values the corresponding output values also change and it is depicted in Table 1. From Table 2, the independent sample t-test with two tailed significance and standard error is determined. The “p” value is less than 0.05 considered to be statistically significant and 95% confidence intervals were calculated. The independent sample t-test was used to compare the accuracy of two algorithms and a statistically significant difference was noticed $p < 0.001$ The logistic regression model obtained 95.52% accuracy in Fig1. When compared with the other algorithm’s performance, the proposed LR classifier is significantly better than KNN.

Discussion:

The predicted accuracy for human activity recognition by the LR algorithm has found 95.52% accuracy over KNN algorithm has found 90.528% of accuracy and also observed that the LR algorithm seems to be significantly better than the KNN algorithm with ($p<0.05$)(Ahad, Lago, and Inoue 2020). Table 1 has a calculated two tailed significant value of p is 0.001, the results corresponding to equal variances assumed are considered for analysis(Sinha et al. 2021). The negative value of ‘t’ implies that the mean of the LR appears to be statistically greater than the mean of the KNN. (Song et al. 2021) Hence this research study found that the LR algorithm seems to be significantly better than the KNN algorithm

To oppose this study some other research findings used modern techniques but using machine learning algorithms for analysis. In the other existing studies, various types of human activity recognition are used for learning(Brownlee 2018). A Neural Network is also used to find human activity recognition in the existing

study(Brownlee 2018; Hu et al. 2021) . This research study is limited to the LR algorithm applied on this dataset having limited features such as only applicable to specific regions with preprocessed dataset also compared only with two groups of data.(Kim et al. 2021)

The limitations in our study is, it's not economical, and takes a longer consuming process. Early prediction of the act isn't possible. Less accuracy rate & can't be implemented altogether in images and videos. In future, it is proposed to implement a HAR system to be trained to be ready to detect more activities, create a system supported user's age bracket , in order that their activity patterns are often want to show similarity, and add more features and perform advanced feature engineering techniques for a far better distinguished activity system. It is proposed to implement a true - time system on a smartphone which may be used for real time activity recognition. Specification is that our approach is able to estimate an individual's actions with high accuracy based only on the knowledge.

CONCLUSION

The main aim of study is to measure accuracy in human activity recognition. In the proposed system Logistic Regression algorithm is implemented for human activity recognition. The parameters achieved have been compared with the K-nearest Neighbors algorithm. The results obtained show that the LR algorithm has found 95% of accuracy on human activity recognition than 90% of the KNN algorithm.

Declaration

Conflict of Interests

No conflict of interest

Authors Contributions

Author SRR was involved in data collection, data analysis, and manuscript writing. Author ASK was involved in conceptualization, data validation, and verifying best accuracy in human activity recognition

Acknowledgements

The authors would like to express their sincere 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. Surya Informatics
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

References

1. Ahad, Md Atiqur Rahman, Anindya Das Antar, and Masud Ahmed. 2021. *IoT Sensor-Based Activity Recognition: Human Activity Recognition*. Springer.
2. Ahad, Md Atiqur Rahman, Paula Lago, and Sozo Inoue. 2020. *Human Activity Recognition Challenge*. Springer Nature.
3. Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.
4. Brownlee, Jason. 2018. *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.
5. Chen, Liming, and Chris D. Nugent. 2019. *Human Activity Recognition and Behaviour Analysis: For Cyber-Physical Systems in Smart Environments*. Springer.
6. Daher, Ali Walid, Ali Rizik, Marco Muselli, Hussein Chible, and Daniele D. Caviglia. 2021. "Porting Rulex Software to the Raspberry Pi for Machine Learning Applications on the Edge." *Sensors* 21 (19). https://doi.org/10.3390/s21196526.
7. Fan, Changjun, and Fei Gao. 2021. "Enhanced Human Activity Recognition Using Wearable Sensors via a Hybrid Feature Selection Method." *Sensors* 21 (19). https://doi.org/10.3390/s21196434.
8. Fu, Yun. 2015. *Human Activity Recognition and Prediction*. Springer.
9. Grieco, Fabrizio, Briana J. Bernstein, Barbara Biemans, Lior Bikovski, C. Joseph Burnett, Jesse D. Cushman, Elsbeth A. van Dam, et al. 2021.

"Measuring Behavior in the Home Cage: Study Design, Applications, Challenges, and Perspectives." *Frontiers in Behavioral Neuroscience* 15 (September): 735387.
10. Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed IsaqaliKarobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.
11. Han, Yang. 2021. "A Virtual Reality Algorithm for the Study of Clinical Efficacy of Sports Injury Rehabilitation Training." *Journal of Healthcare Engineering* 2021 (September): 6725625.
12. Hu, Jianyu, Hongjie Song, Jing Zhou, Rui Liu, and Yi Lv. 2021. "Metal-Tagged CRISPR/Cas12a Bioassay Enables Ultrasensitive and Highly Selective Evaluation of Kanamycin Bioaccumulation in Fish Samples." *Analytical Chemistry*, October. https://doi.org/10.1021/acs.analchem.1c03094.
13. Kim, Hyejoo, Hyeon-Joo Kim, Jinyoon Park, Jeh-Kwang Ryu, and Seung-Chan Kim. 2021. "Recognition of Fine-Grained Walking Patterns Using a Smartwatch with Deep Attentive Neural Networks." *Sensors* 21 (19). https://doi.org/10.3390/s21196393.
14. Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.
15. Ortiz, Jorge Luis Reyes. 2015. *Smartphone-Based Human Activity Recognition*. Springer.
16. Reddy, Poornima, JogikalmatKrithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.
17. Roy, Debaditya, Sarunas Girdzijauskas, and Serghei Socolovschi. 2021. "Confidence-Calibrated Human Activity Recognition." *Sensors* 21 (19). https://doi.org/10.3390/s21196566.
18. Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.
19. Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.
20. Sinha, Vikas Kumar, Kiran Kumar Patro, Paweł Pławiak, and Allam Jaya Prakash. 2021. "Smartphone-Based Human Sitting Behaviors Recognition Using Inertial Sensor." *Sensors* 21 (19). https://doi.org/10.3390/s21196652.
21. Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.
22. Song, Liwen, Changcheng Xiang, Huafeng Guo, and Shiqiang Chen. 2021. "Activity Recognition Using Multiplex Limited Penetrable Visibility Graph." *Journal of the Mechanical Behavior of Biomedical Materials* 124 (October): 104880.
23. Strickland, Jeffrey. n.d. *Logistic Regression Inside and Out*. Lulu.com.
24. Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

TABLES AND FIGURES

Table 1. Statistical results of KNN and LR algorithms. Mean accuracy value, Standard Deviation and Standard Error mean for KNN and LR algorithms are obtained for 10 iterations. It is observed that the LR algorithm performed better than the KNN algorithm

	Group	N	Mean	Std Deviation	Std Error Mean
Accuracy	1	5	95.52820	.419845	.1877604
	2	5	90.52820	.301028	.134736
Loss	1	5	1.64200	.216494	.096819
	2	5	1.64200	.217554	.097293

Table 2. Independent Sample Test of Accuracy and loss (Calculate P-value = 0.001 and Significant value = .98, Mean Difference = 4.9460 and confidence interval = (0.2311-0.1375).
Logistic Regression algorithm are significantly different from KNN algorithm
Independent Samples Test

	Levene's Test for Equality of variances							**T-test or Equality of Means**			
						Significance				**95% Confidence Interval of the Difference**	
		F	**Sig**	**t**	**df**	**One-Sided p**	**Two-Sided P**	**Mean Difference**	**Std.Error Difference**	**Lower**	**Upper**
Accuracy	Equal variances assumed	.711	.424	21.40	8	<.001	<.001	4.9460	.23110	4.4130	5.47892
	Equal variances not assumed			21.40	7.25	<.001	<.001	4.9460	.23110	4.40341	8.488582
Loss	Equal variances assumed	.040	.846	-8.53	8	<.001	<.001	-1.172	.13725	-1.4885	-.8554804
	Equal variances not assumed			-8.53	8.0	<.001	<.001	-1.172	.137258	-1.4885	-.855479

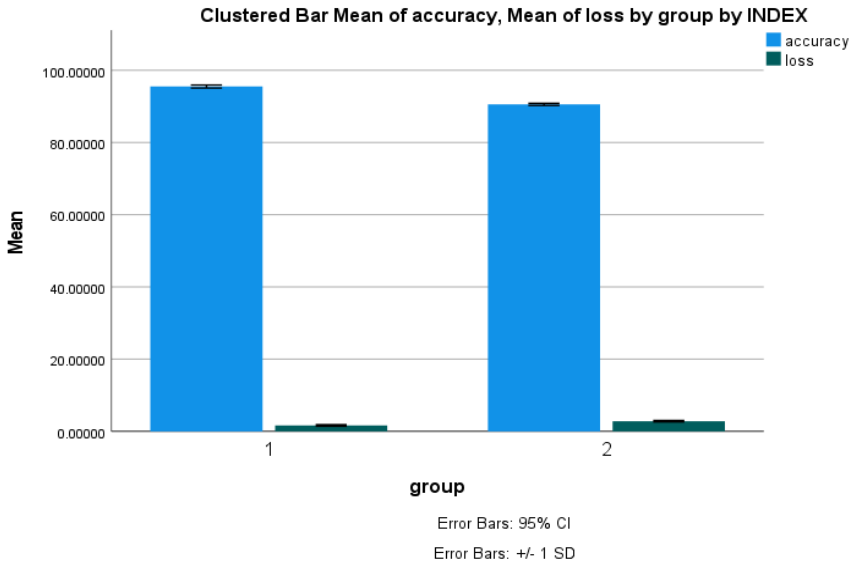

Fig. 1. Comparison of Logistic regression(group1) and K-nearest Neighbors algorithm(group2) in terms of accuracy. The mean accuracy of the group1 is better than group2 and standard deviation of group1 is slightly better than group2. X Axis : LR vs KNN algorithm Y Axis: Mean accuracy of detection±1 SD.