

Knowledge data discovery on Alzheimer's images by using image enhancement techniques and KNN approach

D. Deepa¹, Dr. Anitha Karthi²

Research Scholar, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, BIHER, BIST, Bharath University, Chennai, India¹

Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, BIHER, BIST, Bharath University, Chennai, India²

deepasaram@gmail.com¹, anithakarathi.cse@bharathuniv.ac.in²

Abstract

An identification of Alzheimer's disease using knowledge data discovery methods has exposed hopeful results, but fruitful application in clinical settings needs a collection of optimal accuracy level, less time consumption with generalizability to populations. Here, the system proposed has Image enhancement techniques with instance based classification approaches. This research works finds that image feature extraction technique such as Auto color correlogram Filter techniques on Alzheimer's images dataset by implementing the KNN classifier with linear Nearest Neighbor (LNN) search algorithm using Euclidean distance function model gives 84% of accurateness level, 0.56 of Cohen's kappa coefficient value, 0.84 of Positive Predictive Value, 0.84 of Hit Rate value, 0.84 of F1-Score value, 0.56 of Matthews connection coefficient value, 0.82 of receiver functioning characteristic curve value and 0.83 of Area Under Curve of Precision Recall and it takes time consumption as 0.01 seconds to build a model which is produced an optimal results based on their performance compare with other models.

Keywords: Auto Color Correlogram Filter, KNN, Linear NN Search, Alzheimer's disease, and Euclidean distance function

I Introduction

Alzheimer's disease is a chronic, permanent neurodegenerative illness that sources amnesia, cognitive impairment, and the gradual loss of various additional brain utilities as well as day to day living independence [1,2,3]. Alzheimer's disease is predicted the number of people to rise from 47 million to 152 million, , medical, posing major economic and societal ramifications in the year of 2050.[4].

Research of Alzheimer 's disease connected changes in the brain have been made possible recognitions to brain magnetic resonance imaging. Random forests [6], support vector machine (SVM) [7], and boosting algorithms [8] are just a few of the potential machine learning applications that have used MRI for AD prediction [5].

The following is how the rest of the paper is structured: Section 2 shows the connected works; section 3 displays the materials and techniques; section 4 provides the results and discussions; and lastly, section 5 shows the conclusion.

II Literature Survey

Manual selection is time-consuming and labor-intensive, and it is prone to subjective errors. Image features may be automatically extracted when a CNN model is trained with MRI slices, removing the requirement for manual feature selection during the learning process. [10,11] For Alzheimer's disease diagnosis utilising brain MRI data processing, Islam and Zhang [11] developed an ensemble of three deep CNNs with slightly varying topologies. 20 white matter and GM slices from MR images with major brain structures were chosen in 2019 to sequence an metae of Convolutional Neural Networks [12]. A full MR brain images into several local areas and made a number of 3D patches from each region. The scientists then used the K-Means clustering approach to group the patches from each region into various clusters.[9]

In particular, the CNN-EL was distinct from the previously described approaches, which merged deep learning with ensemble learning to evaluate MRI data for diagnosing Alzheimer's disease in the base classifiers. [13-27].

III Materials and Methods

In this segment concentrations on the Materials and methods on this research work. Alzheimer's dataset rented from Kaggle repository. The below table shows that the description of the borrowed dataset.

S.No	Category	Actual Image Size	Processed Image Size	Number of Images	Sample Size (Random with balanced data)
1	Non Demented	176x208	256x256	1792	500
2	Very Mild Demented	176x208	256x256	2560	500
3	Mild Demented	176x208	256x256	717	500
4	Moderate Demented	176x208	256x256	52	500(synthesized)
Total					2000

Table 1: Meta data of Dataset

Methods:

The following method are applied in this research work.[24-27]

- 1) Borrowed dataset
- 2) Data preprocessing
- 3) Apply Auto Color Correlogram Filter
- 4) Apply for KNN machine learning algorithms using Linear Nearest Neighbor Search in below Distance functions:
 - a) Euclidean distance function
 - b) Chebyshev distance function
 - c) Filtered distance function
 - d) Manhattan distance function
 - e) Minkowski distance function
- 5) To get Optimal results
- 6) Find a best Model

This work follows the above methods in weka 3.9.5 open source software with 10:90 fold cross validation to produce innovative model.

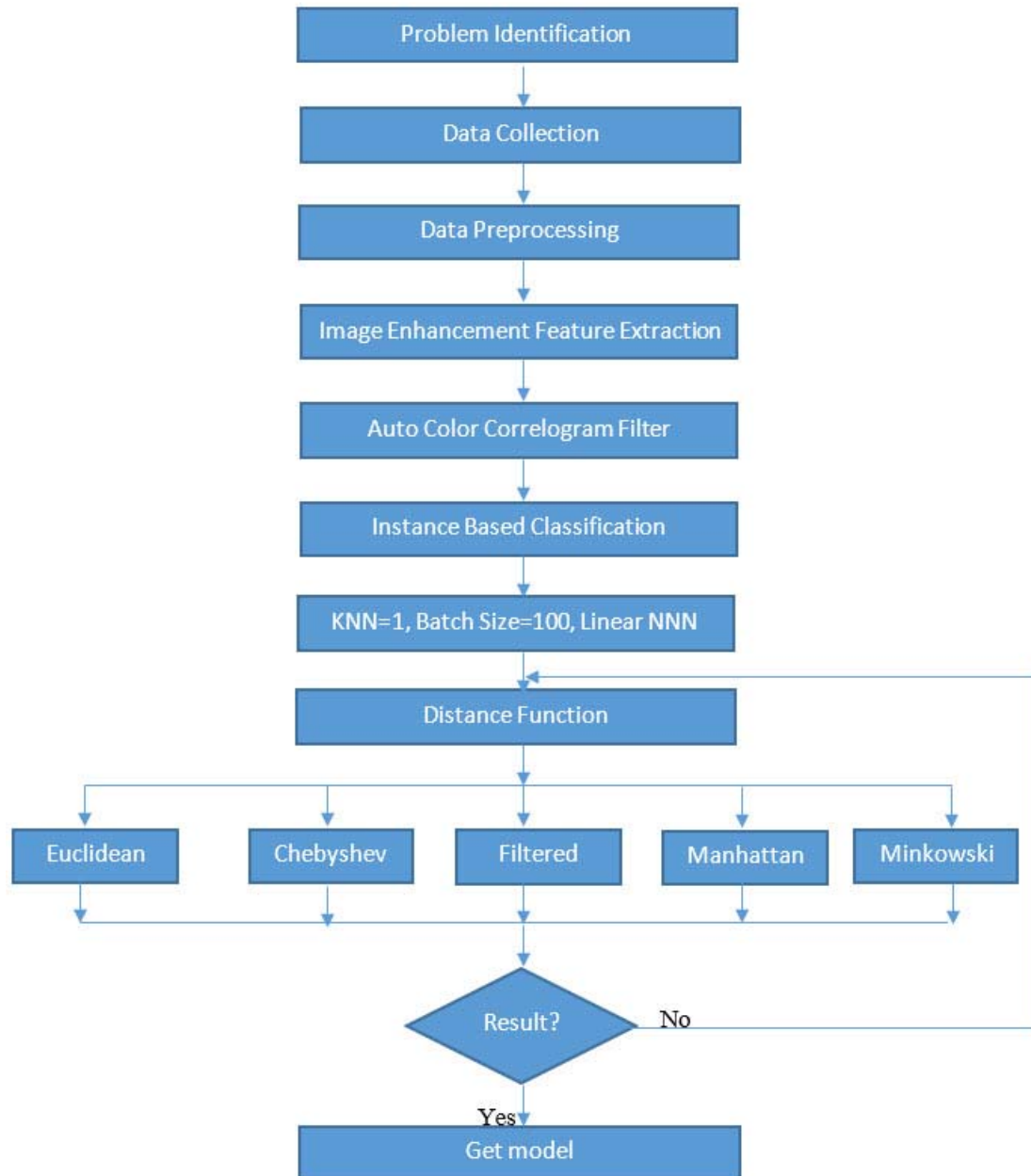


Figure 1: Proposed System

S.No	Various Distance Function on IBK	Accuracy(ACC)	Cohen's kappa coefficient	Positive Predictive Value	Hit Rate	F1 Score	Phi Coefficient	ROCAUC	PRAUC	Time Taken to build model
1	Euclidean Distance	84.00%	0.56	0.84	0.84	0.84	0.56	0.82	0.83	0.01
2	Chebyshev Distance	83.10%	0.52	0.82	0.83	0.83	0.53	0.83	0.84	0.01
3	Filtered Distance	82.85%	0.53	0.82	0.83	0.83	0.53	0.82	0.83	0.20
4	Manhattan Distance	83.75%	0.55	0.83	0.84	0.84	0.56	0.82	0.83	0.00
5	Minkowski Distance	84.00%	0.56	0.84	0.84	0.84	0.56	0.82	0.83	0.03

Table 2: Performance of various distance functions on IBK with Linear NN Search

The above table shows that Instance based classifier produces optimal measurements using Linear NN search algorithm with various distance functions.

The KNN (IBK) classifier with Linear NN Search in Euclidean distance produces 84% of accuracy level, 0.56 of Cohen's kappa coefficient value, 0.84 of Positive Predictive Value, 0.84 of Hit Rate value, 0.84 of F1-Score value, 0.56 of Mathews correlation coefficient value, 0.82 of Area Under Curve of Receiver Operating Characteristic Curve value, 0.83 of Area Under Curve of Precision Recall and it takes 0.01 seconds time consumption to produce this model.

The KNN classifier with Linear NN Search using Chebyshev distance function produces 84% of accuracy level, 0.56 of Cohen's kappa coefficient value, 0.84 of Positive Predictive Value, 0.84 of Hit Rate value, 0.84 of F1-Score value, 0.56 of Mathews correlation coefficient value, 0.82 of Area Under Curve of Receiver Operating Characteristic Curve value, 0.83 of Area Under Curve of Precision Recall value and 0.01 seconds takes to produce this model.

The KNN classifier with Linear NN Search by implementing Filtered distance function produces 84% of accuracy level, 0.56 of Cohen's kappa coefficient value, 0.84 of Positive Predictive Value, 0.84 of Hit Rate value, 0.84 of F1-Score value, 0.56 of Mathews correlation coefficient value, 0.82 of Area Under Curve of Receiver Operating Characteristic Curve value, 0.83 of Area Under Curve of Precision Recall value and 0.20 seconds takes to give this model.

The KNN classifier with Linear NN Search by utilizing Manhattan distance function gives 84% of accuracy level, 0.56 of Cohen's kappa coefficient value, 0.84 of Positive Predictive Value, 0.84 of Hit Rate value, 0.84 of F1-Score value, 0.56 of Mathews correlation coefficient value, 0.82 of Area Under Curve of Receiver Operating Characteristic Curve value, 0.83 of Area Under Curve of Precision Recall Value and zero seconds takes to build this model.

The KNN classifier with Linear NN Search in Minkowski Distance function produces 84% of accuracy level, 0.56 of Cohen's kappa coefficient value, 0.84 of Positive Predictive Value, 0.84 of Hit Rate value, 0.84 of F1-Score value, 0.56 of Mathews correlation coefficient value, 0.82 of Area Under Curve of Receiver Operating Characteristic Curve value, 0.83 of Area Under Curve of Precision Recall Value and 0.03 seconds takes for building this model.

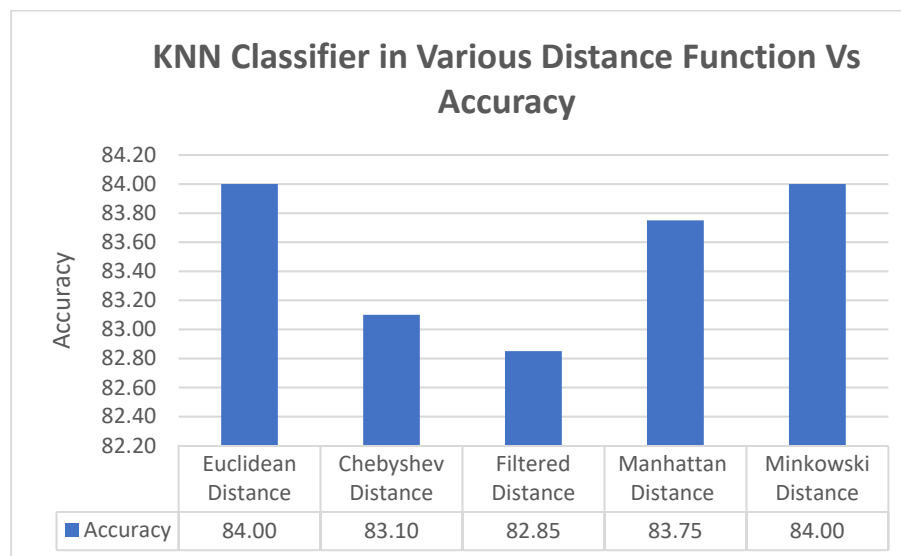


Figure 2: Performance of distance functions on IBK(KNN) with their accuracies

The above diagram shows that the KNN classifier with linear NN Search algorithm has 84% of accuracy level while applying Euclidean and Minkowski distance functions which is highest accuracy compare with other models.

The KNN classifier gives lowest accuracy level is 82.85% which is produced by while applying filtered distance function. The Chebyshev distance function and Manhattan distance function of instance based classifier with linear nearest neighborhood algorithm gives 83.10% of accuracy performance level and 83.75% of accuracy performance level respectively.

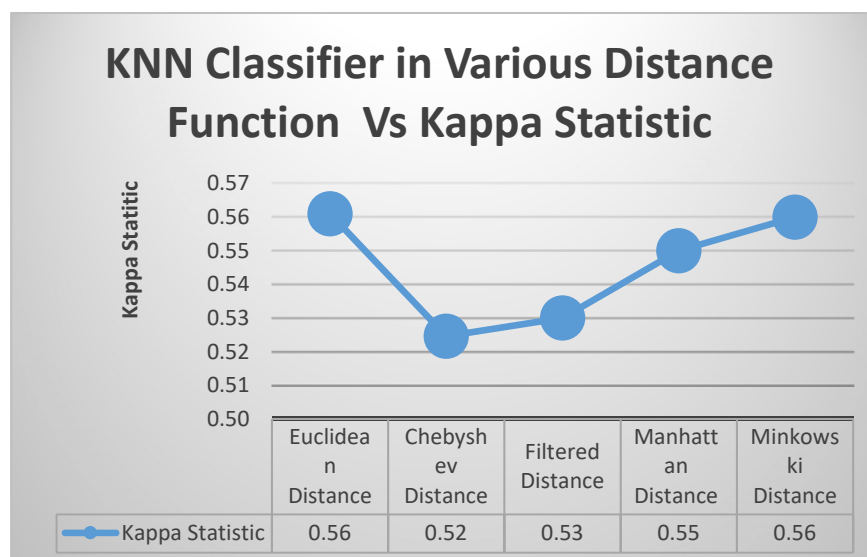


Figure 3: Performance of distance functions on IBK(KNN) with their Cohen's kappa coefficient values

The above diagram shows that the minimum Cohen's kappa coefficient values is 0.52 which is produced by Chebyshev distance function of instance based classifier algorithm with Linear Nearest Neighborhood search classification approach. The filtered distance and Manhattan distance functions are 0.53 and 0.55 of Cohen's kappa coefficient values respectively. The Linear Nearest Neighborhood search classification using Euclidean distance function and Minkowski distance function of instance based classification produce same Cohen's kappa coefficient values which is 0.56 of Cohen's kappa coefficient value.

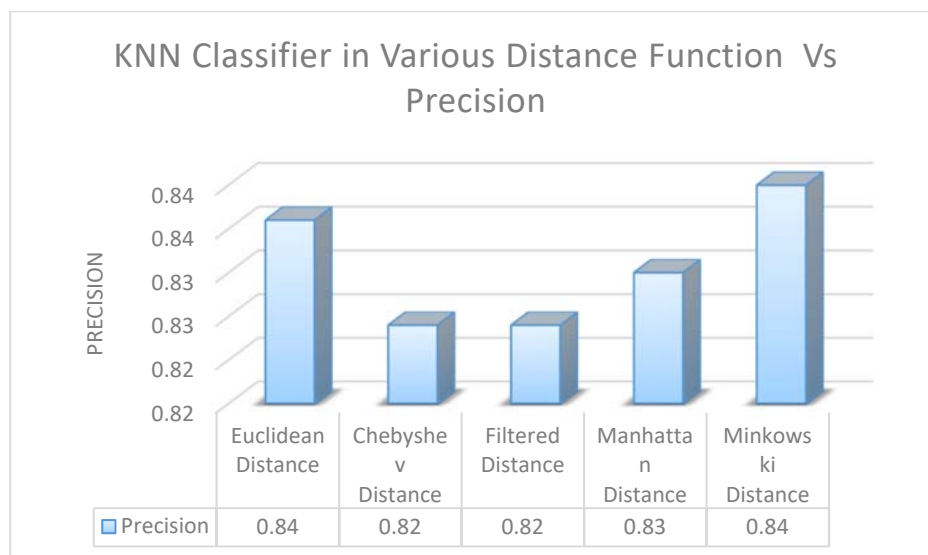


Figure 4: Performance of distance functions on IBK(KNN) with their Positive Predictive Value values

The above diagram shows that the KNN classifier with linear NN Search using Euclidean distance function, and Minkowski distance function produce same Positive Predictive Value values as well highest Positive Predictive Value value which is 0.84 of Positive Predictive Value value.

The least Positive Predictive Value value is 0.82 which is produced by Chebyshev distance function and filtered distance function of KNN classifier with Linear NN search. The Manhattan distance function is 0.83 of Positive Predictive Value value.

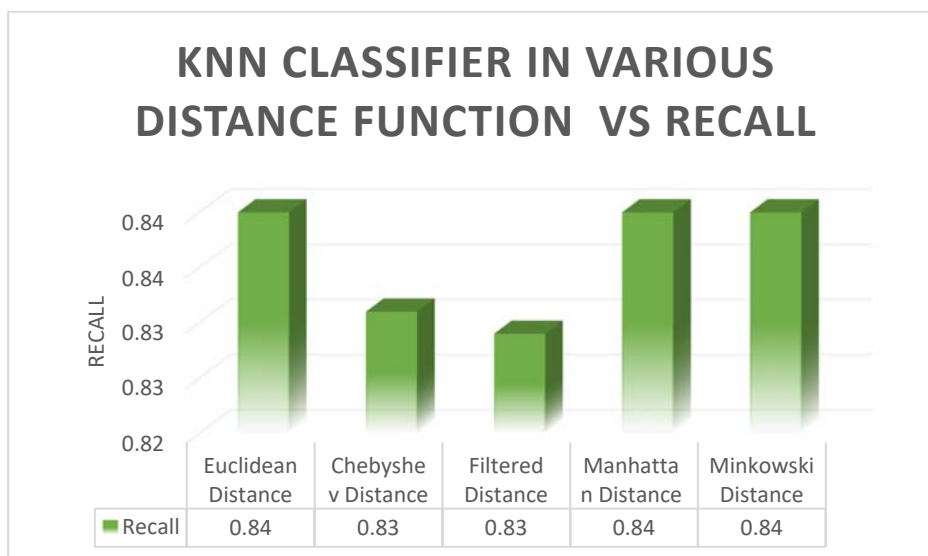


Figure 5: Performance of distance functions on IBK (KNN) with their Hit Rate values

The above diagram shows that the KNN classifier with linear NN Search using Euclidean distance function, Chebyshev distance function and Minkowski distance function produce same Hit Rate values as well highest Hit Rate value which is 0.84 of Hit Rate value.

The least Hit Rate value is 0.83 which is produced by Chebyshev distance function and Filtered distance function of KNN classifier with Linear NN search.

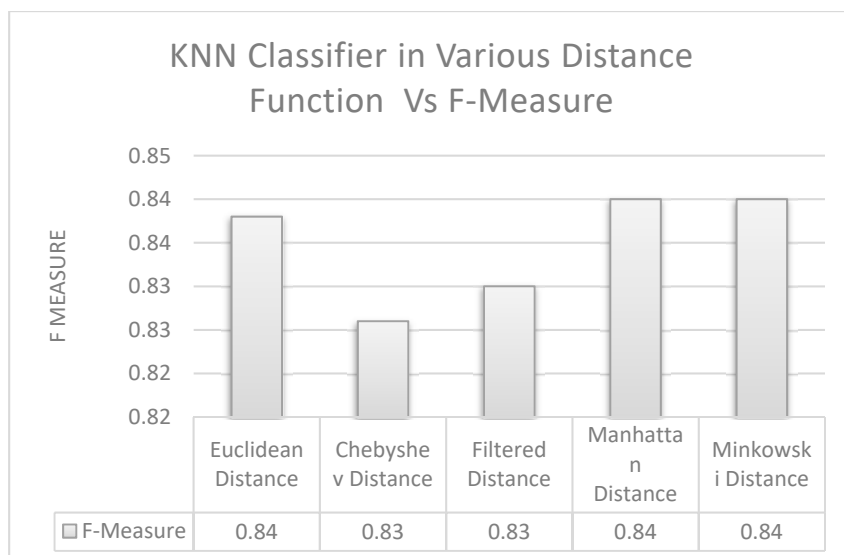


Figure 6: Performance of distance functions on IBK (KNN) with their F1-Score values

The above diagram shows that the KNN classifier with linear NN Search using Euclidean distance function, Manhattan distance function and Minkowski distance function produce same F1-Score values as well highest F1-Score value which is 0.84 of F1-Score value.

The least F1-Score value is 0.83 which is produced by Chebyshev distance function and Filtered distance function of KNN classifier with Linear NN search.

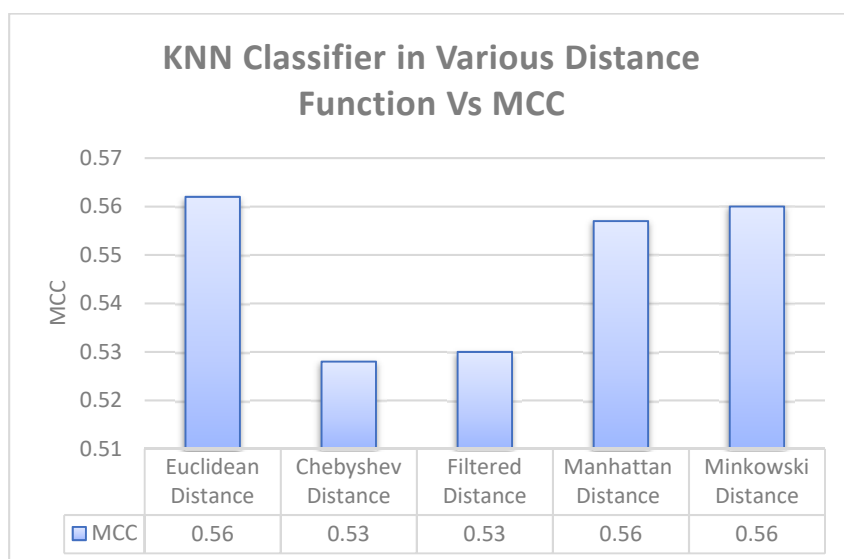


Figure 7: Performance of distance functions on IBK(KNN) with their MCC values

The above diagram shows that the KNN classifier with linear NN Search using Euclidean distance function, Manhattan distance function and Minkowski distance function produce same Matthews Correlation Coefficient values as well highest Matthews Correlation coefficient value which is 0.56 of MCC value.

The least MCC value is 0.53 which is produced by Chebyshev distance function and Filtered distance function of KNN classifier with Linear NN search.

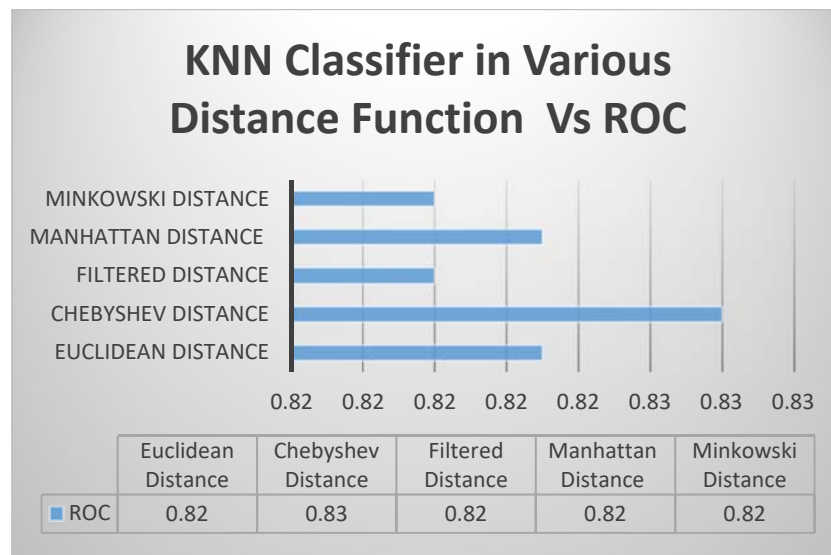


Figure 8: Performance of distance functions on IBK(KNN) with their ROC values

The above diagram shows that the KNN classifier with linear NN Search using Euclidean distance function, Filtered distance function, Manhattan distance function and Minkowski distance function produce same ROC values as well lowest ROC value which is 0.82 of ROC value.

The highest Area Under Curve of Receiver Operating Characteristic Curve value is 0.83 which is produced by Chebyshev distance function of KNN classifier with Linear NN search.

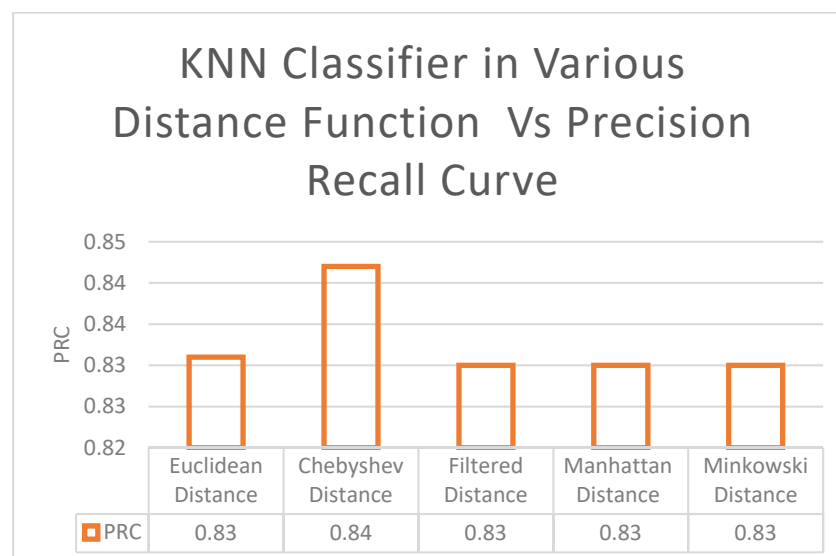


Figure 9: Performance of distance functions on IBK(KNN) with their PRC- AUC values

The above diagram shows that the KNN classifier with linear NN Search using Euclidean distance function, Filtered distance function, Manhattan distance function and Minkowski distance function produce same AUC of PRC values as well lowest AUC of ROC value which is 0.83 of PRC value.

The highest Positive Predictive Value Recall Curve value is 0.84 which is produced by Chebyshev distance function of KNN classifier with Linear NN search.

V Conclusion

This research works concludes that the KNN classifier with linear NN Search algorithm has 84% of accuracy level while applying Euclidean and Minkowski distance functions which is highest accuracy compare with other models. The KNN classifier with linear NN Search using Euclidean distance function and Minkowski distance function produce same Cohen's kappa coefficient values which is 0.56 of Cohen's kappa coefficient value. The KNN classifier with linear NN Search using Euclidean distance function, and Minkowski distance function

produce same Positive Predictive values as well highest Positive Predictive Value which is 0.84 of Positive Predictive Value. The KNN classifier with linear NN Search using Euclidean distance function, Chebyshev distance function and Minkowski distance function produce same Hit Rate values as well highest Hit Rate value which is 0.84 of Hit Rate value. The KNN classifier with linear NN Search using Euclidean distance function, Manhattan distance function and Minkowski distance function produce same F1-Score values as well highest F1-Score value which is 0.84 of F1-Score value. The KNN classifier with linear NN Search using Euclidean distance function, Manhattan distance function and Minkowski distance function produce same Matthews Correlation Coefficient values as well highest Matthews Correlation coefficient value which is 0.56 of MCC value. The KNN classifier with linear NN Search using Euclidean distance function, Filtered distance function, Manhattan distance function and Minkowski distance function produce same ROCAUC values as well lowest AUC of ROC value which is 0.82 of ROCAUC value. The KNN classifier with linear NN Search using Euclidean distance function, Filtered distance function, Manhattan distance function and Minkowski distance function produce same PRC values as well lowest ROCAUC value which is 0.83 of PRC value.

This research work recommends that the KNN classifier with linear NN search using Euclidean distance function from performance of various measurements compare with other models

References

- [1] Pan Dan, Zeng An, Jia Longfei, Huang Yin, Frizzell Tory, Song Xiaowei, Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning, *Frontiers in Neuroscience* ,14,2020, doi: 10.3389/fnins.2020.00259
- [2] Jong Bin Bae et al, Identification of Alzheimer's disease using a convolutional neural network model based onT1-weighted magnetic resonance imaging, *Scientific Reports*, (2020) 10:22252 | <https://doi.org/10.1038/s41598-020-79243-9>
- [3] Ulep, M. G., Saraon, S. K., and McLea, S. (2018). Alzheimer disease. *J. Nurse Practit.* 14, 129–135.
- [4] Christina, P. (2018). *World Alzheimer Report 2018 – The State of the Art of Dementia Research: New Frontiers*. London: Alzheimer's Disease International.
- [5] Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., and Evans, A. C. (2018). Structural neuroimaging as clinical predictor: a review of machine learning applications. *NeuroImage Clin.* 20, 506–522. doi: 10.1016/j.nicl.2018.08.019
- [6] Tripoliti, E. E., Fotiadis, D. I., and Argyropoulou, M. (2011). A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment. *Artif. Intell. Med.* 53, 35–45. doi: 10.1016/j.artmed.2011.05.005
- [7] Leemput, K., Van Maes, F., Vandermeulen, D., and Suetens, P. (2002). Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* 18, 897–908. doi: 10.1109/42.811270
- [8] Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., and Johnson, S. C. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* 48, 138–149. doi: 10.1016/j.neuroimage.2009.05.056
- [9] Li, F., and Liu, M. Alzheimer's Disease and Neuroimaging Initiative (2018). Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *Computer. Med. Imag. Graph.* 70, 101–110. doi: 10.1016/j.compmedimag.2018.09.009
- [10] Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., et al. (2018). Convolutional neural networks-Based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front. Neurosci.* 12:777. doi: 10.3389/fnins.2018.00777
- [11] Islam, J., and Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Inform.* 5:2.
- [12] Ji, H., Liu, Z., Yan, W. Q., and Klette, R. (2019). "Early diagnosis of alzheimer's disease using deep learning," in *Proceedings of the 2nd International Conference on Control and Computer Vision* (New York, NY: Association for Computing Machinery), 87–91.
- [13] Ortiz, A., Munilla, J., Gorriz, J. M., and Ramirez, J. (2016). Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *Int. J. Neural Syst.* 26:1650025. doi: 10.1142/s0129065716500258
- [14] Hwang, E. J. et al. Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw. Open* 2, e191095. <https://doi.org/10.1001/jamanetworkopen.2019.1095> (2019).
- [15] Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410. <https://doi.org/10.1001/jama.2016.17216> (2016).
- [16] Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. <https://doi.org/10.1038/nature21056> (2017).
- [17] Li, H. et al. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimers Dement.* 15, 1059–1070. <https://doi.org/10.1016/j.jalz.2019.02.007> (2019).
- [18] Li, F. & Liu, M. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *Comput. Med. Imaging Graph* 70, 101–110. <https://doi.org/10.1016/j.compmedimag.2018.09.009> (2018).
- [19] Luo, S., Li, X. & Li, J. Automatic Alzheimer's disease recognition from MRI data using deep learning method. *J. Appl. Math. Phys.* 05, 1892–1898. <https://doi.org/10.4236/jamp.2017.59159> (2017).
- [20] Liu, M., Zhang, J., Adeli, E. & Shen, D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Med. ImageAnal.* 43, 157–168. <https://doi.org/10.1016/j.media.2017.10.005> (2018).
- [21] Basaia, S. et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin.* 21, 101645. <https://doi.org/10.1016/j.nicl.2018.101645> (2019).
- [22] Lu, D. et al. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci. Rep.* 8, 5697. <https://doi.org/10.1038/s41598-018-22871-z> (2018).
- [23] Petersen, R. C. et al. Alzheimer's disease neuroimaging initiative (ADNI): Clinical characterization. *Neurology* 74, 201–209 (2010).
- [24] Liu, Y. et al. Education increases reserve against Alzheimer's disease-evidence from structural MRI analysis. *Neuroradiology* 54,929–938. <https://doi.org/10.1007/s00234-012-1005-0> (2012).
- [25] Lee, H. et al. Construction and validation of brain MRI templates from a Korean normal elderly population. *Psychiatry Investig.* 13, 135–145. <https://doi.org/10.4306/pi.2016.13.1.135> (2016).
- [26] Tang, Y. et al. Te construction of a Chinese MRI brain atlas: A morphometric comparison study between Chinese and Caucasian cohorts. *Neuroimage* 51, 33–41. <https://doi.org/10.1016/j.neuroimage.2010.01.111> (2010).
- [27] Suk, H. I., Lee, S. W., Shen, D. & Alzheimer's Disease Neuroimaging, I. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med. Image Anal.* 37, 101–113. <https://doi.org/10.1016/j.media.2017.01.008> (2017).

- [28] Ayyappan.G et al. “*Heart Disease Data Set Classifications: Comparisons of Correlation Co Efficient by Applying Various Functions*”, Indian Journal of Computer Science and Engineering (IJCSE), Volume No.9 Issue No.5 Oct-Nov 2018, Page No: 135-140, e-ISSN: 0976-5166, p-ISSN: 2231-3850.
- [29] S.M. Manimegalai et al. “Image Classification on Covid 19 CXR Images Using Auto Color Correlogram Filter”, Indian Journal of Computer Science and Engineering (IJCSE), EISSN: 0976-5166 P-ISSN: 2231-3850 Vol-12, No-5, and Sep-Oct 2021.

Author Profile



Ms. Deepa.D, received B.E degree in 2006 from Anna University, obtained M.Tech degree from BIHER in 2009 and doing research in Computer Science and Engineering in Bharath Institute of Higher Education and Research, Chennai, India. My current research interests are in Data Mining, Image Processing, Machine Learning, Artificial Intelligence, Data science.



Dr.ANITHA KARTHI received her B.E degree from the University of Madras, Chennai, India in 1998. She obtained her M.E. degree (First Class with Distinction) from CEG, Anna University, Chennai, India in 2007 and her Ph.D. degree from CEG, Anna University, Chennai, India in 2013. She is currently working as a Professor in the Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India. Her current research interests are in data science, machine learning, and IoT applications in the field of Electrical, Electronics, and Computer Science Engineering.