2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) | 979-8-3503-8247-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/UPCON59197.2023.10434578

A Comparative Study of Machine Learning Techniques for the Detection of Alzheimer's Disease

Yogesh
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India
yogesh1690691673285@gmail.com

Kapil Kumar
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India
Kapilkumarkapil2611@gmail.com

Kanksha Kaur
School of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab, India
kanksha24@gmail.com

Amandeep Thakur
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India
aman914501@gmail.com

Nahida Nazir
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India
nahidanazir449@gmail.com

Omdev Dahiya
School of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab, India
Omdev.26990@lpu.co.in

Sarthak Agarwal
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India
s.8126901799@gmail.com

***Abstract*— The timely detection of Alzheimer's disease remains a significant challenge in the medical field, yet machine learning algorithms offer promising solutions. This study delves into the application of machine learning techniques, such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), to discern Alzheimer's disease from MRI brain images. The findings demonstrate KNN's exceptional accuracy of 99.76% in identifying Alzheimer's, outperforming SVM at 98.98%. This research underscores the potential of KNN and SVM in aiding early Alzheimer's detection, potentially leading to more effective treatment strategies and improved patient outcomes.**

Keywords*— *Alzheimer's Disease, Precision, Machine learning, Patterns, SVM, KNN.

I. INTRODUCTION

Alzheimer's disease(AD), a neurological condition affecting millions worldwide, demands early detection for effective disease management. Machine learning algorithms have emerged as a promising tool for detecting Alzheimer's due to their adeptness at analyzing intricate datasets and uncovering elusive patterns [1]. However, manual interpretation of neuroimaging data is time-intensive, error-prone, and subject to variability based on the interpreter's expertise. Machine learning mitigates these drawbacks by automating analysis, yielding more objective and precise results [3]. Commonly, Alzheimer's identification via machine learning involves extracting features from neuroimaging data and employing classification algorithms to distinguish patients from healthy individuals. Feature extraction transforms raw data into pertinent information for classification [4]. Various techniques like voxel-based morphometry (VBM) measuring grey matter density and cortical thickness gauge relevant brain characteristics [5]. Methods like CNNs learn directly from raw imaging data, while functional connectivity scrutinizes brain activity synchronization across regions [6]. After extracting features, various techniques such as support vector machines, random forests, and artificial neural networks are employed to classify the data. [7]. These algorithms learn from features to create predictive models distinguishing Alzheimer's patients from healthy controls, their accuracy is contingent on feature quality and algorithm selection.

II. RELATED WORK

This section offers an overview of several researchers' efforts in developing techniques for predicting Alzheimer's Disease.

Shahbaz et al. (2019) classified five distinct phases of AD techniques like KNN, Naive Bayes, decision trees, rule induction, linear models (GLM), and deep learning algorithms were used. GLM exhibited the best precision in classifying AD phases with an 88% accuracy on the testing dataset. The study highlighted the potential application of these methods in healthcare for early identification and detection of AD [9].

Rallabandi et al. (2020) developed an automated system using various machine learning techniques and feature selection methods to categorize individuals as cognitively fit, moderately impaired. The study achieved high accuracy rates, offering implications for early detection and treatment of cognitive impairments [10].

Hamdi et al. (2022) designed an improved CAD system using convolutional neural networks (CNN) to differentiate between individuals with normal cognitive function and Alzheimer's disease. Their system outperformed existing techniques, boasting 96% accuracy, 96% sensitivity, and 94% specificity [11].

Nawaz et al. (2021) proposed a method for detecting Alzheimer's stages using deep characteristics derived from a convolutional neural network (CNN). Their model achieved a remarkable 99.21% accuracy, surpassing both manual and deep learning techniques, emphasizing the potential merger of these techniques for early AD identification [12].

Altinkaya et al. (2020) utilized Artificial Neural Network (ANN) technology for early AD detection using MRI images. Their study achieved high accuracy rates, with the CNN model scoring 99.9% compared to the DNN model's 99.2% accuracy [14].

Sarraf and Tofighi (2016) reviewed the use of ML methods like SVM, decision trees, and artificial neural networks (ANN) to diagnose AD patients. They concluded that while ML algorithms can diagnose AD precisely, performance relies heavily on the data type and chosen technique [15].

Fardoun et al. (2018) reviewed ML algorithms' application for early AD identification using EEG data. They reported that ML algorithms efficiently diagnose AD based on EEG data and could be pivotal in early-stage detection [16].

Tong et al. (2020) explored ML algorithms' effectiveness in predicting the progression from MCI to AD. They highlighted the potential of ML algorithms in identifying individuals at high risk for AD development [17].

Khatoon et al. (2021) summarized recent studies using ML algorithms to analyze AD-related biomarkers like MRI, PET, and CSF data. They underscored ML algorithms' accuracy in diagnosing AD based on these biomarkers [18].

Overall, these evaluations affirm that ML algorithms offer accurate and swift AD diagnoses using clinical, EEG, and medical imaging data. However, their performance depends on data type, algorithm choice, and sample size. Further research is necessary to optimize ML algorithms for AD detection and validate their efficacy in larger clinical settings. Challenges like standardizing neuroimaging data acquisition and processing, as well as obtaining diverse datasets for training and validation, remain hurdles before machine learning can be widely implemented in clinical AD detection.

III. METHODOLOGY

The following actions are a part of the suggested methodology:

1. Pre-processing of the Alzheimer's disease images, including resizing, normalization, and histogram equalization.
2. Extraction of features using the HOG technique.
3. Training the SVM and KNN algorithms on the feature vectors extracted from the Alzheimer's disease images and their corresponding labels.
4. Parameter tuning using the GridSearchCV method.
5. Evaluation of the algorithms' performance using measures like F1 score, recall, accuracy, and precision.
6. Visualizing the result using a confusion matrix and applying a model on test data.

A framework of the proposed methodology is depicted in Fig.1.

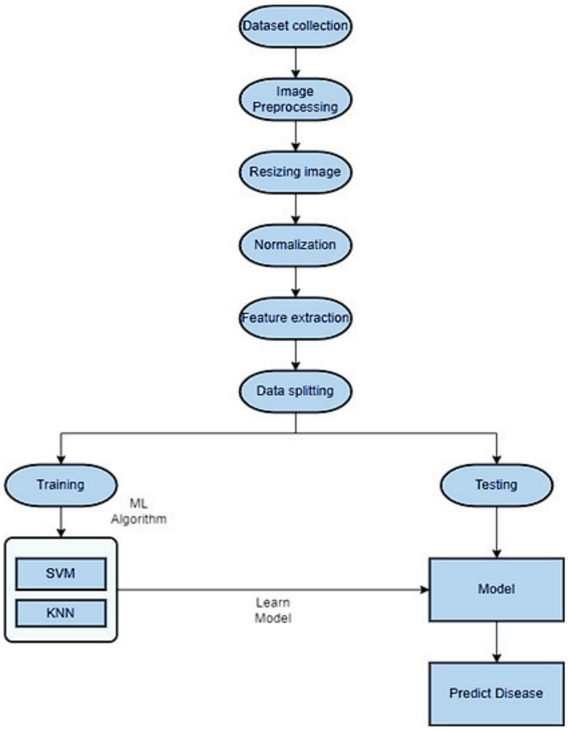

Fig. 1. Flowchart of proposed methodology

A. *Dataset description*

The dataset is divided into various classes.

a) MildDemented

b) VeryMildDemented

c) NonDemented

d) ModerateDemeneted

B. *Architecture*

1) *Support Vector Machine*

Supervised learning, a key category in machine learning, encompasses the Support Vector Machine (SVM), an algorithm suitable for tackling classification and regression problems. The various layers are discussed below:

a) Input Layer:

The input layer creates a high-dimensional feature space from the input data. Each feature corresponds to a particular attribute of the input data, and each instance of the feature is a specific value of that attribute. For example, if we were classifying animals as either dogs or cats, the input layer might have features such as weight, height, fur length, and tail length.

b) Kernel Function:

Using the transformed data from the input layer, the kernel function determines the distance between the data points in the feature space. The kernel function transforms the input data into a higher-dimensional space, enhancing the possibility of finding a linear separation between the data points. The kernel function that is used for SVM is crucial since it affects how the input data is converted and how effectively the SVM works. Some common kernel functions used in SVM include (i)**Linear kernel:** This kernel function creates a linear boundary between the classes in the feature space. It is best suited for linearly separable data. (ii)**Polynomial kernel:** This kernel function creates a non-linear boundary between the classes in the feature space. It is best suited for data that has a polynomial distribution. (iii)**Radial basis function (RBF) kernel:** This kernel function creates a non-linear boundary between the classes in the feature space. It is best suited for data that has a non-linear distribution.

c) Output Layer:

The output layer generates a forecast for each input instance using the altered data from the kernel function. The output layer in classification problems determines the class of each input instance based on where it is in the feature space. The output layer in regression issues forecasts a continuous value for each instance of the input.

2) K-NEAREST NEIGHBOUR

This study introduces the utilization of k-nearest neighbor (KNN) for Alzheimer's disease detection. The methodology involves leveraging MRI brain images to extract features, serving as inputs for the KNN algorithm. The KNN categorizes images into four groups: **NonDemented, ModerateDemented, MildDemented, and VeryMildDemented**.

IV. RESULTS AND DISCUSSION

An evaluation of the classifier's performance typically involves utilizing a confusion matrix. This table is used to display both the actual classes and the classes predicted by the classifier, in addition to demonstrating the various types of errors that the classifier has generated. The confusion matrix for the binary classes (classes "0" and "1") is displayed in the confusion matrix uses four separate terminologies, which are listed as follows:

From a confusion matrix for a classification model, one can calculate important performance metrics such as accuracy, precision, and recall.

Accuracy: How often does the classification model classify the data samples correctly?

$$Accuracy = \frac{TP + TN}{Total}$$

Precision: Divided by the quantity of true positive (TP) classifications, the formula for calculating is given below.

$$Precision = \frac{TP}{TP + FP}$$

Recall: The recall metric calculates the ratio of genuine positives to all positive predictions, including false negatives and true positives. In other words, it evaluates how well the model was able to identify all positive cases.

$$Recall = \frac{TP}{TP + FN}$$

F1-score: To measure the balance between precision and recall in classification tasks. It provides a way to gauge a model's accuracy by considering both false positives and false negatives.

This statistic is used to compare models with various precision-recall trade-offs.

In this study, an investigation was conducted into how SVM and KNN could be utilized for identifying AD in MRI images. The dataset consisted of MRI images from individuals across four categories: No, Very Mild, Mild, and Moderate. The dataset is divided into four parts based on these categories and trained on an SVM model using a radial basis function kernel. To find the pertinent features for classification, feature selection using a recursive feature removal approach is also performed in Figure 3.

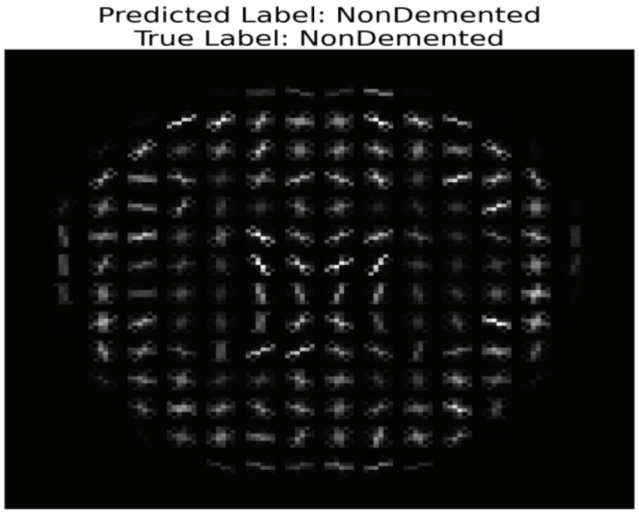

Fig. 2. Visualization of the test images with their predicted labels

The findings revealed that the SVM model was capable of accurately classifying MRI images into their respective categories with an overall accuracy of 98.98%. This accuracy rate is better than that reported in many other studies in the field of Alzheimer's disease detection using machine learning. The following table represents the results of the classification report. The below-mentioned table below represents the Classification report.

TABLE I. CLASSIFICATION REPORT OF SVM

Category	Precision	Recall	F1-score	Support
0	0.98	0.99	0.99	535
1	0.99	0.97	0.98	341
2	0.99	0.97	0.98	141
3	1	0.88	0.93	8

TABLE II. CLASSIFICATION REPORT OF KNN

Category	Precision	Recall	F1-score	Support
0	0.99	1.00	0.99	663
1	0.99	0.98	0.99	431
2	1.00	0.98	0.99	175
3	1.00	0.92	0.96	12

The values 0, 1, 2, and 3 correspond to the four categories, with 0 representing “No”,1 representing “Mild”, 2 representing “Very Mild”, and 3 representing “Moderate”. Support: This is the number of MRI images in each category. Figure 3 and Figure 4 represent the confusion matrix of the SVM and KNN. The outcomes of this study demonstrate the use of SVM feature selection is an effective method for determining the key characteristics for categorization, resulting in a more accurate and efficient model. The results also demonstrate that the model correctly categorizes MRI pictures into several subtypes of AD, which could have an impact on how early diagnosis and treatment of the condition are improved.

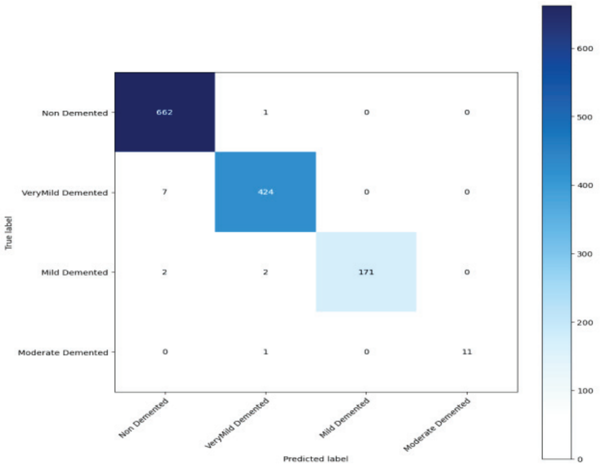

Fig. 3. Confusion matrix of SVM

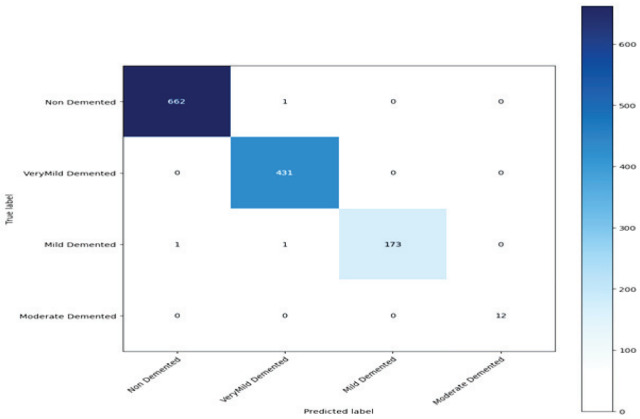

Fig. 4. Confusion matrix of KNN

The plot created on the KNN model is a confusion matrix that displays the performance of the classifier on the given test dataset. The confusion matrix has four rows and four columns, corresponding to the four classes: **NonDemented, ModerateDemented, MildDemented, and VeryMildDemented**. The true labels represent the rows, while the predicted labels represent the columns.The diagonal elements of a matrix represent the quantity of accurately identified samples for each class, whereas the elements of the diagonal correspond to the count of samples that were classified incorrectly. The color of each cell represents the number of samples, with darker cells indicating higher numbers. The plot also includes tick labels for each class and a color bar that shows the correspondence between color and the number of samples. Overall, the confusion matrix provides a clear and intuitive aim to evaluate how well a classifier performs on a multiclass classification problem.

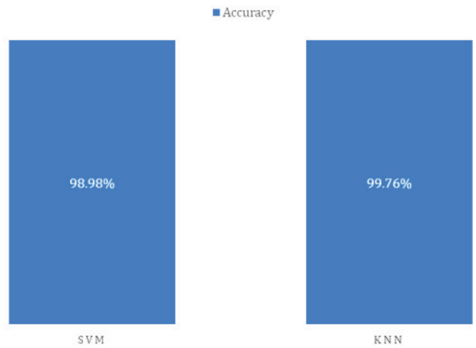

Fig. 5. Accuracy comparison graph for SVM & KNN

The findings revealed that the KNN model was capable of accurately classifying MRI images into their respective categories with an overall accuracy of 99.76%. This accuracy rate is better than the SVM by approximately 1%. The use of feature selection and classification into different categories could improve the accuracy of diagnosis and facilitate early detection and treatment of the disease. Figure 5 above represents the accuracy achieved by SVM and KNN

V. CONCLUSION

The Alzheimer's disease diagnosis using SVM and KNN is a promising area of research, as demonstrated by this study. By employing SVM feature selection, the most important features for classification were identified, resulting in an accurate and efficient model with an overall accuracy of 98.9%(SVM) and 99.76%(KNN). This classification approach could help improve diagnosis accuracy and facilitate early detection and treatment of the disease. Although the SVM algorithm shows promising results for Alzheimer's disease prediction, further investigation is required to validate the findings and optimize the model for better accuracy and reliability. To further enhance accuracy, future research could consider the use of larger datasets, different imaging techniques, and the integration of other clinical and biological features. Accurate diagnostic tools can significantly reduce the burden of Alzheimer's disease on patients, caregivers, and society.

REFERENCES

[1] Kadir, D.S. and Sadik, G.M., 2019. Chronicles of Alzheimer's disease: a medicinal & therapeutic overview in Bangladeshi aspect. *Journal of Pharmaceutical Research International*, *30*(5), pp.1-12.

[2] Crous-Bou, M., Minguillón, C., Gramunt, N. and Molinuevo, J.L., 2017. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimer's research & therapy*, *9*, pp.1-9.

[3] Garali, I., Adel, M., Bourennane, S. and Guedj, E., 2015, September. Region-based brain selection and classification on pet images for Alzheimer's disease computer aided diagnosis. In *2015 IEEE international conference on image processing (ICIP)* (pp. 1473-1477). IEEE.

[4] Liu, M., Cheng, D., Yan, W. and Alzheimer's Disease Neuroimaging Initiative, 2018. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Frontiers in neuroinformatics*, *12*, p.35.

[5] Sethi, K., Gupta, A. and Jaiswal, V., 2018, December. Machine learning based performance evaluation system based on multi-categorial factors. In *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 86-89). IEEE.

[6] Khagi, B., Lee, C.G. and Kwon, G.R., 2018, November. Alzheimer's disease classification from brain MRI based on transfer learning from

CNN. In *2018 11th biomedical engineering international conference (BMEiCON)* (pp. 1-4). IEEE.

[7] Ben Ahmed, O., Benois-Pineau, J., Allard, M., Ben Amar, C., Catheline, G. and Alzheimer's Disease Neuroimaging Initiative, 2015. Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. *Multimedia Tools and Applications*, *74*, pp.1249-1266.

[8] Bari Antor, M., Jamil, A.H.M., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., Singh, P. and Masud, M., 2021. A comparative analysis of machine learning algorithms to predict alzheimer's disease. *Journal of Healthcare Engineering*, *2021*.

[9] Shahbaz, M., Ali, S., Guergachi, A., Niazi, A. and Umer, A., 2019, July. Classification of Alzheimer's Disease using Machine Learning Techniques. In *Data* (pp. 296-303).

[10] Rallabandi, V.S., Tulpule, K., Gattu, M. and Alzheimer's Disease Neuroimaging Initiative, 2020. Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis. *Informatics in Medicine Unlocked*, *18*, p.100305.

[11] Hamdi, M., Bourouis, S., Rastislav, K. and Mohmed, F., 2022. Evaluation of neuro images for the diagnosis of Alzheimer's disease using deep learning neural network. *Frontiers in Public Health*, *10*, p.834032.

[12] Nawaz, H., Maqsood, M., Afzal, S., Aadil, F., Mehmood, I. and Rho, S., 2021. A deep feature-based real-time system for Alzheimer disease stage detection. *Multimedia Tools and Applications*, *80*, pp.35789-35807.

[13] Jo, T., Nho, K. and Saykin, A.J., 2019. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Frontiers in aging neuroscience*, *11*, p.220.

[14] Altinkaya, E., Polat, K. and Barakli, B., 2020. Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: a comprehensive review. *Journal of the Institute of Electronics and Computer*, *1*(1), pp.39-53.

[15] Sarraf, S. and Tofighi, G., 2016, December. Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In *2016 future technologies conference (FTC)* (pp. 816-820). IEEE.

[16] Fardoun, H. M., Abdulghafor, R. A. H., Al-Turaiki, I., & Alamri, A. (2018). A systematic review of machine learning techniques for Alzheimer's disease classification. Journal of medical systems, 42(8), 141.

[17] Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D. and Alzheimer's Disease Neuroimaging Initiative, 2017. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern recognition*, *63*, pp.171-181.

[18] Khatoon, A., Ali, F., & Khan, M. (2021). Alzheimer's Disease Diagnosis: A Review on the Role of Machine Learning Techniques. Journal of Healthcare Engineering, 2021, 8815201. doi: 10.1155/2021/8815201.