http://doi.org/10.35784/iapgos.5775

received: 08.01.2024 | revised: 29.01.2024 | accepted: 07.03.2024 | available online: 31.03.2024

GENETIC ALGORITHM-BASED DECISION TREE OPTIMIZATION FOR DETECTION OF DEMENTIA THROUGH MRI ANALYSIS

Govada Anuradha, Harini Davu, Muthyalanaidu Karri
Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Kanuru, Vijayawada, India

Abstract. *Dementia is a devastating neurological disorder that affects millions of people globally, causing progressive decline in cognitive function and daily living activities. Early and precise detection of dementia is critical for optimal dementia therapy and management however, the diagnosis of dementia is often challenging due to the complexity of the disease and the wide range of symptoms that patients may exhibit. Machine learning approaches are becoming progressively more prevalent in the realm of image processing, particularly for disease prediction. These algorithms can learn to recognize distinctive characteristics and patterns that are suggestive of specific diseases by analyzing images from multiple medical imaging modalities. This paper aims to develop and optimize a decision tree algorithm for dementia detection using the OASIS dataset, which comprises a large collection of MRI images and associated clinical data. This approach involves using a genetic algorithm to optimize the decision tree model for maximum accuracy and effectiveness. The ultimate goal of the paper is to develop an effective, non-invasive diagnostic tool for early and accurate detection of dementia. The GA-based decision tree, as proposed, exhibits strong performance compared to alternative models, boasting an impressive accuracy rate of 96.67% according to experimental results.*

Keywords: dementia, genetic algorithm, decision tree

OPTYMALIZACJA DRZEWA DECYZYJNEGO OPARTA NA ALGORYTMIE GENETYCZNYM DO WYKRYWANIA DEMENCJI POPRZEZ ANALIZĘ MRI

Streszczenie. *Demencja jest wyniszczającym zaburzeniem neurologicznym, które dotyka miliony ludzi na całym świecie, powodując postępujący spadek funkcji poznawczych i codziennych czynności życiowych. Wczesne i precyzyjne wykrywanie demencji ma kluczowe znaczenie dla optymalnej terapii i zarządzania demencją, jednak diagnoza demencji jest często trudna ze względu na złożoność choroby i szeroki zakres objawów, które mogą wykazywać pacjenci. Podejścia oparte na uczeniu maszynowym stają się coraz bardziej powszechne w dziedzinie przetwarzania obrazu, szczególnie w zakresie przewidywania chorób. Algorytmy te mogą nauczyć się rozpoznawać charakterystyczne cechy i wzorce, które sugerują określone choroby, analizując obrazy z wielu modalności obrazowania medycznego. Niniejszy artykuł ma na celu opracowanie i optymalizację algorytmu drzewa decyzyjnego do wykrywania demencji przy użyciu zbioru danych OASIS, który obejmuje duży zbiór obrazów MRI i powiązanych danych klinicznych. Podejście to obejmuje wykorzystanie algorytmu genetycznego do optymalizacji modelu drzewa decyzyjnego w celu uzyskania maksymalnej dokładności i skuteczności. Ostatecznym celem artykułu jest opracowanie skutecznego, nieinwazyjnego narzędzia diagnostycznego do wczesnego i dokładnego wykrywania demencji. Zaproponowane drzewo decyzyjne oparte na GA wykazuje wysoką wydajność w porównaniu z alternatywnymi modelami, szczycąc się imponującym współczynnikiem dokładności wynoszącym 96,67% zgodnie z wynikami eksperymentalnymi.*

Słowa kluczowe: demencja, algorytm genetyczny, drzewo decyzyjne

Introduction

Dementia represents a progressive neurodegenerative state characterized by a slow decline in cognitive functions, affecting memory, reasoning, behavior, and the ability to carry out everyday tasks. The World Health Organization (WHO) estimates that approximately 50 million individuals globally are presently grappling with dementia, and this number is projected to triple by the year 2050 [39]. As the population continues to age, dementia has emerged as a prominent public health issue, presenting difficulties in both the timely identification and the development of effective treatments. The timely identification of dementia is pivotal in facilitating prompt interventions, tailored treatment, and enhancing patient outcomes. Over the past few years, medical imaging methods have surfaced as potent instruments for detecting dementia. Specifically, Magnetic Resonance Imaging (MRI) delivers intricate perspectives on the structural and functional modifications in the brain that are linked to dementia. By analyzing MRI data, we can identify specific biomarkers and patterns that aid in the detection and classification of dementia cases.

Similar to the research conducted by Javeed, A. et al. of [22] which demonstrates the potential of machine learning techniques for improving the accuracy of dementia diagnosis through MRI analysis, our paper seeks to leverage genetic algorithms to optimize decision tree models for more accurate and efficient detection of dementia, addressing the limitations of traditional diagnostic methods. Nowroozpoor, A. et al. [32] highlights the challenge of detecting cognitive disorders in a clinical setting can reduce the effects of dementia. So our paper aims to develop an optimized decision tree model using genetic algorithms to improve the accuracy of dementia detection through MRI analysis.

The goal is to enhance the accuracy and reliability of dementia diagnosis by leveraging advanced machine learning techniques. By optimizing the decision tree model using a Genetic Algorithm, we aim to overcome the limitations of traditional decision tree approaches and improve the accuracy of dementia detection. Traditional decision tree models often struggle to accurately classify dementia cases due to the intricate patterns and high-dimensional nature of MRI data. To address this challenge, we propose the utilization of a Genetic Algorithm, which draws inspiration from natural evolution, to fine-tune the decision tree's hyperparameters. The optimization process is designed to identify the best set of hyperparameters that enhances the model's effectiveness in detecting dementia.

1. Literature review

Garcia-Gutierrez et al. [19] have introduced GA-MADRID, a Python-based framework for Alzheimer's Disease and Frontotemporal Dementia diagnosis. This framework integrates data processing, feature engineering, predictive modeling, and multi-class decision-making modules. It employs Genetic Algorithms for feature selection and machine learning algorithm optimization, along with evolutionary grammars and Bayesian networks for decision-making. This work holds promise for advancing the field of medical diagnosis in neurodegenerative diseases.

In Basheer et al. study [8], the primary focus was on dementia prediction using longitudinal MRI data from the OASIS dataset. A modified capsule network, known for its efficacy in image classification, was employed. Before model development, comprehensive exploratory data analysis was carried out to identify crucial features, assess data correlations and density, and hierarchically examine various factors in the dataset. Singular Value Decomposition (SVD) matrix decomposition was utilized for feature extraction, reducing dimensionality while preserving essential information. The resulting model exhibited exceptional performance, achieving an impressive accuracy rate of 92.39%, surpassing other contemporary machine learning models. These findings hold significant promise for early dementia detection and intervention, potentially improving patient outcomes through timely treatments.

Within the realm of cognitive impairment research, Liang et al. [24] delved into the innovative application of Amazon Alexa as a Voice-Assistant System (VAS) for capturing voice commands from older adults, aged 65 and above, encompassing both individuals with healthy cognitive function and those with Mild Cognitive Impairment (MCI). The main aim of this study was to create machine learning models for classification and then evaluate their performance in comparison to cognitive assessment scores. Using a range of machine learning algorithms such as neural networks, support vector machines, decision trees, and random forests, Liang's research investigated the use of VAS data in conjunction with advanced computational methods.

In the work conducted by Perovnik et al. [33], an extensive review was carried out to identify recent English-language publications within the past five years, drawing upon databases such as PubMed and IEEE Xplore. The review primarily focused on research papers investigating the integration of Artificial Intelligence (AI) into computerized cognitive tests. Notably, the findings from these studies revealed substantial improvements in discrimination sensitivity, increasing by 4%, and specificity, by 3%, when AI-based cognitive tests were employed. Of particular significance was the integration of acoustic and linguistic features within speech, conversation, and language tests, which yielded an impressive accuracy rate of 94%. Additionally, the application of deep learning techniques in the analysis of brain scans demonstrated remarkable accuracy, approaching 92%. These results underscore the transformative potential of AI-based methodologies in enhancing the accuracy and efficiency of cognitive assessment, carrying significant implications for the early detection and intervention in cognitive impairments.

Diaz-Alvarez et al. [13] introduced an innovative methodology for precise diagnoses of Alzheimer's disease (AD), Fronto Temporal Dementia (FTD), and related disorders utilizing FDG-PET imaging. This approach entailed a comprehensive comparative analysis of genetic algorithms and principal component analysis (PCA) techniques, each tailored with K-Nearest Neighbor and Bayes Networks. To validate their effectiveness, the researchers employed K-fold cross-validation and external validation approaches. The study unequivocally affirmed the potency of genetic algorithms in accurately diagnosing AD, FTD, and related disorders based on FDG-PET imaging data. These findings underscore the significant potential of genetic algorithms as a powerful tool in neuroimaging analysis, offering promise for the precise classification and diagnosis of neurodegenerative disorders, thus advancing the field of diagnostic accuracy and precision.

Mirzaei et al. [29] conducted a thorough review of journal articles published since 2016, focusing on the application of machine learning techniques for Alzheimer's Dementia (AD) detection. This review critically analyzed various approaches, including Support Vector Machine, Random Forest, Convolutional Neural Network, and K-means, employed in these studies. While the review highlighted the absence of a single best approach for AD diagnosis due to the disease's complexity, it noted the potential of deep learning techniques, particularly Convolutional Neural Networks, in improving accuracy and effectiveness. These advanced techniques allow for intricate pattern and feature extraction from imaging data, offering the promise of more accurate and reliable AD diagnosis. The review emphasized the importance of continued research and development in the field of machine learning for AD diagnosis to advance our understanding and enhance patient care.

Javeed et al. [22] conducted a systematic literature review (SLR) that spanned the years 2011 to 2022, focusing on machine learning (ML)-based automated diagnostic systems designed for dementia prediction. Throughout this rigorous review, the selected research articles underwent critical analysis and discussion to evaluate the effectiveness of ML-driven automated diagnostic systems in predicting dementia. The review's specific emphasis was placed on comparing the performance of various data modalities, including images, clinical features, and voice data, when integrated with ML-based models for dementia prediction.

An alternative method was developed by Angelillo et al. [5] using an Automated Measurement Tool (AMT) implemented on a digitizing tablet that incorporated an electronic pen. This novel approach enabled the acquisition of additional measures beyond traditional paper-based test observations. Handwriting measures were obtained through the digitizing tablet, which served as input for machine learning algorithms aimed at automating disease detection. The ensemble approach demonstrated notable performance with an area under the curve (AUC) of 87.30% and a sensitivity rate of 86.11%. The outcomes underscore the efficacy of the AMT-based approach and underscore the potential of incorporating digitizing tablets and machine learning algorithms for the automated detection of the specified disease.

Varoquaux et al. research [38] critically examines challenges in computer analysis of medical images. Focusing on data limitations and research incentives, the paper discusses potential biases at every step. Drawing evidence from the literature and data challenges, the study offers insights into the pervasive nature of biases in the field. Despite challenges, the paper highlights ongoing efforts to counteract these issues and provides recommendations for future improvement. This work contributes to a nuanced understanding of obstacles and potential solutions in advancing computer analysis for medical imaging and patient health.

Nori et al. [31] conducted research focusing on predicting incident mild cognitive impairment, Alzheimer's Disease, and related dementias using structured data from administrative and electronic health records. The study, encompassing a cohort of patients and controls, employed gradient boosting machine, lightGBM, for model fitting. The 2-year model demonstrated a sensitivity of 47% and an impressive area-under-the-curve of 87%. Even in the 3-year model, the sensitivity remained notable at 24%. While this dropped to 15% in year 8, the model maintained a commendable AUC of 72%. The study suggests the model's potential as a valuable tool for screening patients in trial recruitment and management. Nori, V. S.'s methodology showcases the application of machine learning in predicting dementia-related conditions from diverse healthcare data sources.

Bansal et al. [7] contributes to the emerging field of applying machine learning to neurodegenerative disorders, with a particular focus on dementia. The study presents a comparative analysis of four machine learning algorithms – J48, Naïve Bayes, Random Forest, and Multilayer Perceptron – utilizing CFSSubsetEval for attribute reduction. The results highlight J48 as the most effective algorithm for dementia detection among the tested methods. This comparative assessment underscores the potential of machine learning techniques, particularly J48, in the early detection of dementia – a crucial step in addressing this global health issue. The last page should be filled at least 50%.

In the study by So et al. [36], a two-layer model inspired by dementia support centers is proposed for early dementia diagnosis. Utilizing data from dementia screenings at the Gangbuk-Gu center in the Republic of Korea, the model incorporates patient information, MMSE-KC dementia screening results, and CERAD-K for precise testing. The two-stage classification involves initial categorization of MMSE-KC data into normal and abnormal, followed by the use of CERAD-K data for dementia and mild cognitive impairment classification. Comparative analysis with various algorithms, including Naive Bayes, Random Forest, and Multilayer Perceptron (MLP), shows that MLP achieves the highest F-measure values for normal cases, while SVM exhibits the highest values for mild cognitive impairment and dementia. The proposed early diagnosis model is positioned as a time and cost-effective solution, simplifying the dementia diagnostic process.

Al-Badarneh et al. [3], presents a hybrid approach for the classification of Magnetic Resonance Imaging (MRI) brain images. The study utilizes decision tree (DT) and genetic algorithms (GA) to construct a binary classifier categorizing MRIs as normal or abnormal. The methodology comprises three stages: texture features extraction, features reduction through principal component analysis, and MRI classification using the proposed DT/GA approach. The classifier is evaluated on a benchmark MRI dataset of 710 brain images from Harvard Medical School, demonstrating significant accuracy improvements. Results indicate an accuracy of 96.31% with the decision tree and a notable improvement to 98.55% with the proposed DT/GA approach. This approach showcases the efficacy of combining DT and GA for accurate classification of MRI brain images.

Miled et al. [27], explores the transformation of medical notes into dementia risk classifiers, emphasizing their potential in predicting the risk of developing chronic diseases. The study utilizes TF-ICF for keyword selection, employing two encodings for summarizing the notes: the average vector embedding using BERT or Clinical BERT, and aggregation based on UMLS concepts. Misspellings of keywords are considered for enhanced predictive performance. A neural network is applied to the first encoding, while a gradient boosted trees model is used for the second encoding. Results indicate successful identification of dementia risk one year ahead with an AUC of 75%, though challenges in generalizability to other healthcare institutions are noted. The study underscores the promise of medical notes for risk prediction models, considering factors such as note length.

Mirheidari B. [28] addresses the challenge of dementia detection by proposing an automatic classification system using an intelligent virtual agent (IVA). The study focuses on conversation analysis (CA) during interviews between patients and neurologists, aiming to distinguish between progressive neurodegenerative memory disorder (ND) and non-progressive functional memory disorders (FMD). The manual CA process is resource-intensive, prompting the introduction of an IVA-led approach. Analyzing neurologist-patient and IVA-patient conversations, the study achieves impressive ND/FMD classification rates of 90.0% and 90.9%, respectively. Mirheidari B.'s work underscores the potential of leveraging IVA technology to automate and enhance the efficiency of dementia detection through conversational analysis, providing valuable insights into distinguishing features between human and IVA-led interactions.

Zhao et al. [40] introduced a voice recognition-based digital cognitive screener (DCS) for community-based dementia screening. This proof-of-concept study assesses the reliability, validity, and feasibility of the DCS among Chinese older adults. Participants underwent demographic, clinical, and DCS assessments. Dementia and mild cognitive impairment (MCI) diagnoses were based on the Montreal Cognitive Assessment. The DCS exhibited promising results, showing good internal consistency, test-retest reliability, and inter-rater reliability. Receiver operating characteristic analyses indicated high discriminative validity, with the DCS demonstrating excellent sensitivity (100%) and good specificity (80%) for dementia detection. The findings suggest that the DCS is a reliable and valid tool for digital dementia screening in a community setting, warranting further investigation in larger-scale screenings.

Gorji et al. [20] proposes an effective method for early Alzheimer's disease (AD) diagnosis, crucial for increased eligibility in clinical trials and better planning for patients. Focusing on mild cognitive impairment (MCI) as a key AD risk factor, the study introduces a novel diagnostic approach utilizing pseudo Zernike moments (PZMs) on structural MRI data. PZMs extract discriminative information from AD, MCI, and healthy control (HC) groups. Artificial neural networks, specifically pattern recognition and learning vector quantization (LVQ) networks, classify the extracted information. With 500 MRIs from the ADNI database, the method achieves impressive results, notably 94.88% accuracy, 94.18% sensitivity, and 95.55% specificity for AD vs. MCI, and 95.59% accuracy, 95.89% sensitivity, and 95.34% specificity for MCI vs. HC. This approach showcases significant potential for accurate early AD diagnosis using advanced image analysis and neural networks.

2. Proposed methodology

The methodological flowchart of our suggested model is depicted in Fig. 1. In our study, we utilize the OASIS Longitudinal dataset for dementia detection through MRI analysis. Data preprocessing involves mean imputation for missing values and one-hot encoding for categorical variables. Principal Component Analysis (PCA) is employed for feature extraction, mitigating high-dimensionality challenges. Our unique approach involves optimizing Decision Tree hyperparameters using Genetic Algorithm (GA). The GA-tuned Decision Tree is then used for model training, resulting in improved performance.

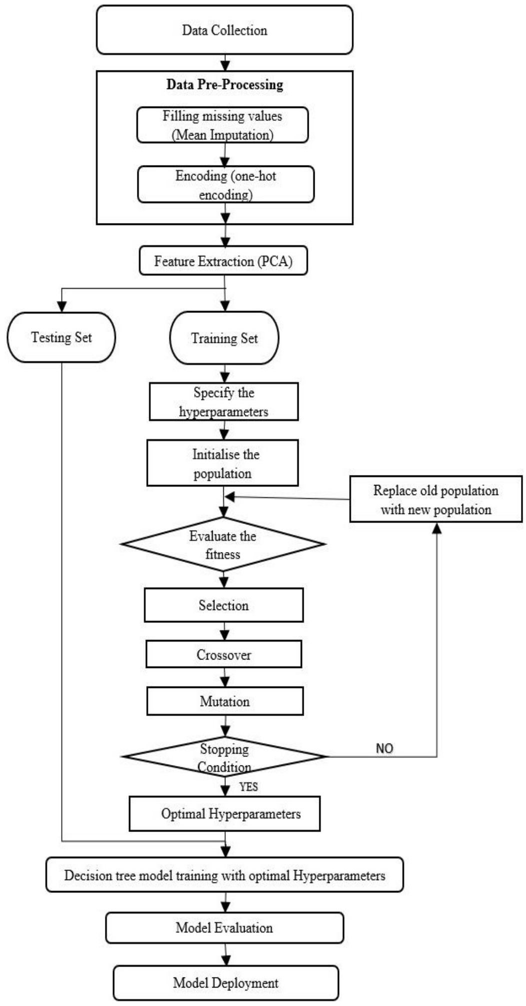

Fig. 1. Proposed system flow diagram

2.1. Data collection

Our data collection revolves around the Oasis dataset [30], encompassing longitudinal MRI and PET images accumulated over a 15-year period from clinical patients. The dataset encompasses 150 subjects aged between 60 and 96, each having undergone multiple scans, resulting in a total of 373 imaging sessions. Each subject has access to 3 or 4 T1-weighted MRI scans. The cohort includes both right-handed men and women. Within this group of 150 subjects, 72 consistently remained nondemented, 64 remained persistently demented, and 14 underwent a change in diagnostic status during the study. In summary, our dataset provides a longitudinal MRI perspective on 150 subjects, capturing diverse diagnostic trajectories. Table 1 outlines diagnoses, including counts for demented, non-demented, and converted individuals, along with gender distribution.

Table 1. Class-labels of Oasis dataset

Diagnosis row labels	Number of individuals	Male	Female
Demented	146	86	60
Non-Demented	190	61	129
Converted	37	37	34

2.2. Data pre-processing

Within our study, we encountered null values within the Oasis dataset during data pre-processing. To ensure data quality, we employed a simple algorithm that systematically identified and replaced these null values with the respective column's mean value. This meticulous approach resulted in a clean dataset devoid of null values, setting the stage for reliable and noise-free analysis as we delve into dementia prediction through MRI analysis.

Data encoding stands as a crucial step in the data pre-processing phase. In our paper, the one-hot encoding technique plays a pivotal role in effectively utilizing the OASIS Longitudinal dataset. It serves as a vital step in the data pre-processing phase, where categorical variables within the dataset are transformed into a numerical format compatible with machine learning models. Through one-hot encoding, each category within a categorical feature is systematically converted into binary variables, ensuring that our dataset is primed for seamless integration with machine learning algorithms. This transformation enhances data compatibility, simplifies computations, and facilitates the subsequent analysis as we embark on dementia prediction through MRI analysis.

2.3. Feature extraction using PCA

In this module, we leverage Principal Component Analysis (PCA) to distill essential information from our pre-processed MRI data. PCA stands as a potent technique for dimensionality reduction, effectively addressing the challenges posed by high-dimensional feature spaces. By transforming our original data into a lower-dimensional representation, PCA retains maximum information while shedding redundant or less informative features. This dimensionality reduction not only enhances computational efficiency but also aids in subsequent optimization and model training phases. It allows us to identify influential features, improving the accuracy and reliability of our dementia detection model by focusing on the most discriminative aspects of the MRI data. The formula [13] for the principal component is characterized as linear combinations of the fundamental variables, which is given by eq. (1).

$$PCA_n = b_{n1} * y_1 + b_{n2} * y_2 + \cdots + b_{nr} * y_r \quad (1)$$

where PCA_n represents the n^{th} principal component, b_{nj} denotes the weights or coefficients associated with the j^{th} original variable y_j, and y_j represents the j^{th} original variable. The formula illustrates how each principal component is formed as a linear combination of the original variables with specific weights.

2.4. Optimization of decision tree hyperparameters using Genetic Algorithm

Our approach involves the optimization of Decision Tree hyperparameters using the Genetic Algorithm (GA) module, a pivotal component in our MRI-based dementia detection framework. Decision Trees, a foundational machine learning technique, rely heavily on hyperparameter settings to determine their structure and predictive performance. These hyperparameters encompass critical factors such as maximum depth, minimum samples required for splits, maximum leaf nodes, and maximum features, each exerting a substantial influence on model accuracy and generalization capabilities. The task of identifying the ideal hyperparameter configuration can be formidable, with traditional methods proving computationally expensive or inadequate is described in Algorithm 1. GA offers an efficient alternative, systematically exploring hyperparameter combinations, uncovering intricate associations, and elevating the Decision Tree's accuracy and reliability in dementia detection is given by eq. (2).

$$Population_{new} = Mutation\left(Crossover\left(Selection(Population_{old})\right)\right) \quad (2)$$

Algorithm 1: Hyper parameter tuning using Genetic Algorithm

```
Begin
  Population ← Create the initial population by employing
   random selection.
  Loop: For each iteration
  Begin
    Calculate fitness values for each individual in the population
    in the array.

    p1, p2 ←  Choose parents through bit tournament selection.
    o1, o2 ← single-point crossover between selected parents
    m_o1 ←  offspring1 should be subjected to bit-flip mutation
    m_o2 ←  offspring2 should be subjected to bit-flip mutation
    replace the least suited individuals with their mutant progeny

  End

  Hyper_parameters ← select the most fit people from population

End
```

2.5. Model training with optimized Decision Tree

In this module, the optimized Decision Tree obtained from the previous step is utilized for model training. The optimized Decision Tree refers to the Decision Tree that has undergone hyperparameter tuning using the Genetic Algorithm, resulting in improved performance. Algorithm 2 describes the process of model training.

Algorithm 2 serves as a comprehensive guide, encapsulating the procedural intricacies entailed in this model training endeavor. This extensive process extends beyond the initial optimization through the Genetic Algorithm, encompassing the subsequent training phase where the Decision Tree engages with the dataset. In this phase, the Decision Tree dynamically learns and adapts based on the refined parameters achieved through the optimization process. The nuanced interplay between the Genetic Algorithm and the Decision Tree stands as a sophisticated approach aimed at enhancing the model's discernment and classification accuracy. This holistic methodology underscores a meticulous and iterative training framework, ensuring that the model refines its understanding of intricate patterns within the data for more precise and reliable decision-making.

By harnessing the capabilities of the optimized Decision Tree, the overarching objective is to attain an elevated level of accuracy and reliability in the subsequent stages of the project. This module holds paramount significance as a critical step in the evolution of the model. The synergistic integration of advanced optimization techniques and model training principles establishes the groundwork for robust and effective decision-making in the classification of dementia stages. The refined Decision Tree, a product of meticulous tuning, acts as a pivotal component in shaping the model's ability to discern intricate patterns within the data. This strategic approach ensures a more precise and reliable classification process, marking a pivotal advancement in the overall effectiveness of the model for dementia diagnosis.

Algorithm 2: GA-Decision Tree Model ()

```
Begin
    ID3(Instances, Desired_feature, Attributes)
     Input: Instances - Training examples
     Desired_feature - Attribute for prediction
     Attributes - List of attributes for evaluation
     Output: Decision tree accurately classifying the provided
      examples.

   1. Create a new tree node called root.
   2. If all instances are positive:
        2.1 Set the root node label to '+'.
        2.2 Return the root.
   3. If all instances are negative:
        3.1 Set the root node label to '-'.
        3.2 Return the root.
   4. If Attributes is empty:
        4.1 Set the root node label to the most frequent value of
             Desired_feature in Instances.
        4.2 Return the root.
   5. Otherwise:
        5.1 Select the best attribute (A) from Attributes that most
             effectively classifies the instances.
        5.2 Set the decision attribute for the root node as A.
        5.3 For each potential value (vi) of attribute A:
          5.3.1 Create a new tree branch beneath the root,
                 associated with the condition A=vi.
          5.3.2 Consider Examples_vi as the subset of instances
                 within Instances having the value vi for attribute.
          5.3.3 If Examples_vi is empty:
            5.3.3.1 Create a leaf node beneath the branch with
                     the label set to the most prevalent value.
          5.3.4 Else:
            5.3.4.1 Add the subtree ID3(Examples_vi,
           Desired_feature, Attributes - {A}) beneath the branch.
      6. Return the root.

 End
```

2.6. Performance evaluation

In our research, a comprehensive evaluation and comparison are undertaken, focusing on two distinct models: the GA-Decision Tree model and the standalone Decision Tree model. This meticulous analysis aims to gauge the efficacy of these classification models in the context of dementia detection through MRI analysis. Key performance metrics, including precision, recall, F1-score, and accuracy, serve as fundamental benchmarks in this assessment. The detailed results of this comparative analysis are meticulously presented in table 2, offering nuanced insights into the individual and relative performance levels of both models. This thorough examination provides a robust foundation for understanding the strengths and limitations of each approach, contributing valuable knowledge to the ongoing discourse in the field of dementia diagnosmodel and the standalone Decision Tree model. To assess the efficacy of these classification models, we utilize fundamental performance metrics such as precision, recall, F1-score, and accuracy. The results of this comparative analysis are presented in Table 2, providing valuable insights into the respective performance levels of both models in the context of dementia detection through MRI analysis.

Table 2. Performance Comparison of GA-Decision Tree and Decision Tree Models

Model	Accuracy, %	Precision, %	Recall, %
Decision tree	86.66	50.00	42.85
GA-Decision tree	96.67	97.46	96.25

3. Results and discussion

Within the context of our investigation, we conducted a comparative assessment of two models: the standalone Decision Tree model and the GA-Decision Tree model. The Decision Tree model exhibited an accuracy rate of 86.66%, demonstrating its competence in dementia detection through MRI analysis. In contrast, the GA-Decision Tree model showcased a remarkable accuracy rate of 96.67%, signifying a significant performance enhancement. This substantial improvement in accuracy observed in the GA-Decision Tree model underscores the efficacy of employing Genetic Algorithms (GAs) for hyperparameter tuning within the decision tree framework. The GA-Decision Tree model, by optimizing hyperparameters, elevates its capacity to capture intricate patterns, resulting in more precise dementia prediction. The integration of the Genetic Algorithm (GA) not only refines hyperparameter settings but also significantly enhances the model's ability to accurately identify essential diagnostic patterns. This synergy between GA optimization and the Decision Tree architecture represents a substantial improvement in the model's overall performance, emphasizing its potential as a robust tool for nuanced dementia prediction through the meticulous extraction and utilization of crucial features within the MRI data.

3.1. Experimental results

In this section, we present the experimental findings of our study, placing specific emphasis on evaluating the performance of both the Decision Tree and GA-based Decision Tree models in the domain of dementia detection using MRI data. To ensure a comprehensive assessment of the models' effectiveness, we utilized Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) analysis, a widely acknowledged metric for binary classification tasks. The ROC curve provides a visual representation of the true positive rate (sensitivity) plotted against the false positive rate (1-specificity) across various classification thresholds. This approach enables a nuanced exploration of the models' performance characteristics, offering valuable insights into their capacity for accurate dementia classification.

Figure 2 vividly displays the Receiver Operating Characteristic (ROC) curve corresponding to the Decision Tree model in our study. Calculating the Area Under the Curve (AUC) for this model yields a noteworthy value of 0.87, attesting to its commendable discriminative power. The ROC curve visually portrays the model's proficiency in effectively distinguishing between dementia and non-dementia cases, with higher AUC values indicative of a heightened level of diagnostic accuracy. This outcome underscores the robust performance of the Decision Tree model in delineating nuanced patterns within the MRI data and substantiates its potential as an effective tool for dementia detection.

Figure 3 depicts the ROC curve for the Decision Tree model optimized with Genetic Algorithm in our investigation. The noteworthy AUC of 0.97 emphasizes the model's strong discriminative ability. This high AUC underscores the proficiency of the GA-based Decision Tree model in accurately identifying dementia cases using MRI data.

Figure 4 illustrates the confusion matrix of the Decision Tree model when applied to testing data. This matrix provides valuable information about the model's accuracy in classification. The model demonstrates a higher proportion of correct predictions than incorrect ones. Specifically, it effectively identifies the presence of the condition and accurately recognizes its absence.

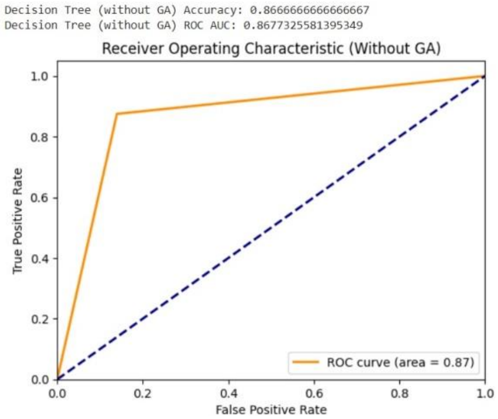

Fig. 2. ROC curve for the Decision Tree model

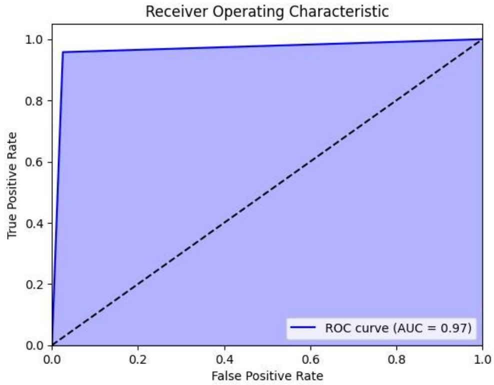

Fig. 3. ROC curve of GA-Decision tree model

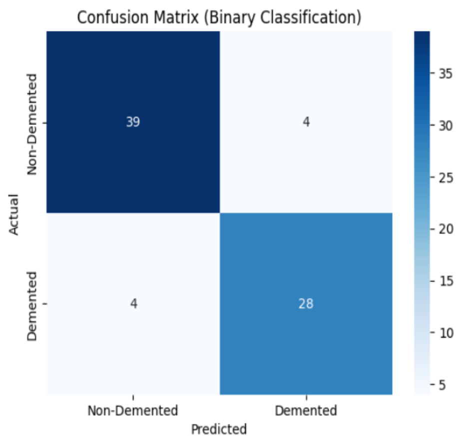

Fig. 4. Confusion Matrix of Decision tree model

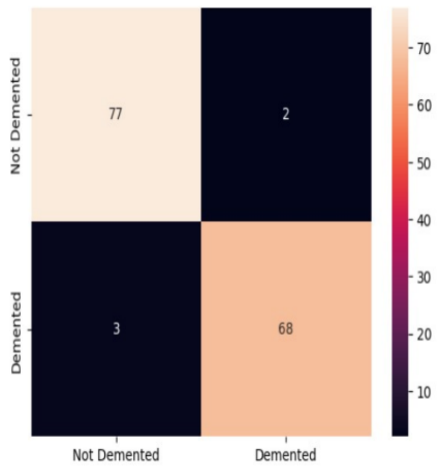

Fig. 5. Confusion Matrix of GA-Decision tree model

Figure 5 displays the confusion matrix for the GA-Decision Tree model, showcasing a significant presence of accurate positive and negative predictions. This suggests adeptness in accurately predicting both positive cases and non-diseased samples. Furthermore, the visualization emphasizes the equilibrium in addressing both facets of the classification task. Additionally, the matrix accentuates the skill in minimizing erroneous positive and negative predictions, enhancing its overall dependability in discerning between positive and healthy samples.

3.2. Discussion

Table 2 provides a comprehensive comparison between our proposed GA-Decision Tree model and previous work in the field. Dementia prediction through MRI analysis has been explored using various machine learning approaches. The table summarizes the performance of both our model and previously implemented methods, offering insights into the advancements achieved through the integration of Genetic Algorithms for hyperparameter tuning. This comparison stands as a useful reference for researchers aiming to assess their approaches in the current landscape of this domain.

Table 3. Comparative analysis of prior approaches and our proposed model

Methods	Accuracy, %	Specificity, %	Recall, %
Deep Neural Networks [5]	92.39	79	74
Neural networks, SVM, Random Forest [6]	68	77	71
Deep Learning [7]	92	-	91.67
KNN, Bayes Networks [8]	86.7	71	86
SVM, Random Forest, CNN, K-means [9]	85	68	84
Decision tree	81.08	90.0	42.85
GA-Decision tree	96.67	97.14	96.25

4. Conclusion

This research work presents a comprehensive comparative analysis between the Decision Tree and GA-Decision Tree models for the task of dementia detection using MRI data. The results demonstrate the superior performance of the GA-Decision Tree model, which achieved remarkable accuracy at 96.67%, precision at 97.46%, recall at 96.25%, and an impressive F1-score of 96.85%. These findings underscore the potency of Genetic Algorithms (GAs) in optimizing hyperparameters and elevating the overall predictive capabilities of the model. Notably, our experiments stress the critical role of parameter fine-tuning, encompassing learning rate and epochs, to unlock the full potential of the model. The GA-Decision Tree's ability to converge swiftly during training, courtesy of its adaptive learning rate mechanism, has significantly contributed to its superior predictive accuracy within a reduced number of iterations. This study reveals the promising prospects of incorporating GAs into the Decision Tree framework, marking a pivotal step toward advancing dementia detection through MRI analysis. The potential implications include early diagnosis and the capacity to revolutionize patient care, representing a significant advancement in imaging and diagnostic practices.

References

[1] Abdollahi J., Nouri-Moghaddam B.: Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran Journal of Computer Science 5(3), 2022, 205–220.
[2] Adeli H., Ghosh-Dastidar S., Dadmehr N.: Alzheimer's disease and models of computation: Imaging, classification, and neural models. Journal of Alzheimer's Disease 7(3), 2005, 187–199.
[3] Al-Badarneh A., Najadat H., Alraziqi A.: Brain Images Classifier: A Hybrid Approach Using Decision Trees and Genetic Algorithms. JINT 7(2), 2016.
[4] Aminizadeh S. et al.: The applications of machine learning techniques in medical data processing based on distributed computing and the Internet of Things. Computer methods and programs in biomedicine 241, 2023, 107745.
[5] Angelillo M. T. et al.: Attentional pattern classification for automatic dementia detection. IEEE Access 7, 2019, 57706–57716.
[6] Azad R. et al.: Medical image segmentation on MRI images with missing modalities: a review [http://arxiv.org/abs/2203.06217].

[7] Bansal D. et al.: Comparative analysis of various machine learning algorithms for detecting dementia. Procedia computer science 132, 2018, 1497–1502.
[8] Basheer S., Bhatia S., Sakri S. B.: Computational modeling of dementia prediction using deep neural network: analysis on OASIS dataset. IEEE access 9, 2021, 42449–42462.
[9] Biswal A.: What Is Principal Component Analysis? Simplilearn.com [www.simplilearn.com/tutorials/machine-learning-tutorial/principal-component-analysis] (avaible 7.11.2023).
[10] Bukhari S. N. H., Webber J., Mehbodniya A.: Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates. Scientific Reports 12(1), 2022, 7810.
[11] Deng W. et al.: An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. Information Sciences 585, 2022, 441–453.
[12] Dhiman G. et al.: A novel machine-learning-based hybrid CNN model for tumor identification in medical image processing. Sustainability 14(3), 2022, 1447.
[13] Díaz-Álvarez J. et al.: Genetic algorithms for optimized diagnosis of Alzheimer’s disease and Frontotemporal dementia using Fluorodeoxyglucose positron emission tomography imaging. Frontiers in aging neuroscience 13, 2022, 983.
[14] Drouka A. et al.: Dietary and nutrient patterns and brain MRI biomarkers in dementia-free adults. Nutrients 14(11), 2022, 2345.
[15] Elhazmi A. et al.: Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU. Journal of infection and public health 15(7), 2022, 826–834.
[16] Elyan E. et al.: Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artificial Intelligence Surgery 2, 2022.
[17] Emam M. M., Houssein E. H., Ghoniem R. M.: A modified reptile search algorithm for global optimization and image segmentation: Case study brain MRI images. Computers in Biology and Medicine 152, 2023, 106404.
[18] Fang L., Wang X.: Brain tumor segmentation based on the dual-path network of multi-modal MRI images. Pattern Recognition 124, 2022, 108434.
[19] García-Gutierrez F. et al.: GA-MADRID: Design and validation of a machine learning tool for the diagnosis of Alzheimer’s disease and frontotemporal dementia using genetic algorithms. Medical & Biological Engineering & Computing 60(9), 2022, 2737–2756.
[20] Gorji H. T., Haddadnia J.: A novel method for early diagnosis of Alzheimer’s disease based on pseudo Zernike moment from structural MRI. Neuroscience 305, 2015, 361–371.
[21] Haug C. J., Drazen J. M.: Artificial intelligence and machine learning in clinical medicine. New England Journal of Medicine 388(13), 2023, 1201–1208.
[22] Javeed A. et al.: Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions. Journal of Medical Systems 47(1), 2023, 17.
[23] Leocadi M. et al.: Awareness impairment in Alzheimer’s disease and fronto-temporal dementia: a systematic MRI review. Journal of Neurology 270(4), 2023, 1880–1907.
[24] Liang X. et al.: Evaluating voice-assistant commands for dementia detection. Computer Speech & Language 72, 2022, 101297.
[25] Li R. et al.: Applications of artificial intelligence to aid early detection of dementia: a scoping review on current capabilities and future directions. Journal of biomedical informatics 127, 2022, 104030.
[26] Liu H. et al.: NeuroCrossover: An intelligent genetic locus selection scheme for genetic algorithm using reinforcement learning. Applied Soft Computing 146, 2023, 110680.
[27] Miled Z. B. et al.: Feature engineering from medical notes: A case study of dementia detection. Heliyon 9(3), 2023.
[28] Mirheidari B. et al.: Dementia detection using automatic analysis of conversations. Computer Speech & Language 53, 2019, 65–79.
[29] Mirzaei G., Adeli H.: Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia. Biomedical Signal Processing and Control 72, 2022, 103293.
[30] Marcus D. S. et al.: Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. Journal of Cognitive Neuroscience 22(12), 2010, 2677–2684.
[31] Nori V. S. et al.: Machine learning models to predict onset of dementia: a label learning approach. Alzheimer's & Dementia: Translational Research & Clinical Interventions 5, 2019, 918–925.
[32] Nowroozpoor A. et al.: Detecting cognitive impairment and dementia in the emergency department: a scoping review. Journal of the American Medical Directors Association 23(8), 2022, 1314–1315.
[33] Perovnik M. et al.: Automated differential diagnosis of dementia syndromes using FDG PET and machine learning. Frontiers in Aging Neuroscience 14, 2022, 1005731.
[34] Ramos D. et al.: Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. Energy Reports 8, 2022, 417–422.
[35] Shehab M. et al.: Machine learning in medical applications: A review of state-of-the-art methods. Computers in Biology and Medicine 145, 2022, 105458.
[36] So A. et al.: Early diagnosis of dementia from clinical data by machine learning techniques. Applied Sciences 7(7), 2017, 651.
[37] Squires M. et al.: A novel genetic algorithm based system for the scheduling of medical treatments. Expert Systems with Applications 195, 2022, 116464.
[38] Varoquaux G. Cheplygina V.: Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digital Medicine 5(1), 2022, 48.
[39] World Health Organization: WHO and World Health Organization: WHO. Dementia [www.who.int/news-room/fact-sheets/detail/dementia/?gclid=CjwKCAiA-P-rBhBEEiwAQEXhHzn09E8kOoRAOoS8xNltu1svCep3MzGBPB363AJ20n3XF3v9M9C9axoCS7QQAvD_BwE] (avaible 15.03.2023).
[40] Zhao X. et al.: A voice recognition-based digital cognitive screener for dementia detection in the community: Development and validation study. Frontiers in Psychiatry 13, 2022, 899729.

Prof. Govada Anuradha
e-mail: ganuradha@vrsiddhartha.ac.in

Dr. G. Anuradha, received her Ph.D. from Andhra University, Visakhapatnam, Andhra Pradesh. She is working as an Associate professor in Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Krishna district. She published 25 papers in national and international journals and more than 10 papers published in national and international conferences. She has guided several projects at M. Tech. level. Her research interests are artificial intelligence and machine learning.

https://orcid.org/0000-0002-0999-0376

Eng. Harini Davu
e-mail: davuharini@gmail.com

Harini Davu is currently pursuing B.Tech. degree in Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College. She is interested in doing projects in the machine learning domain. She contributed to the project "Crow detection in peanut field using Raspberry Pi".

https://orcid.org/0009-0008-6187-1797

Eng. Muthyalanaidu Karri
e-mail: karrimuthyalanaidu3802@gmail.com

Muthyalanaidu Karri is currently pursuing B.Tech. degree in Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College. He contributed to the project "Morning Start: An Android Application for Milk Delivery".

https://orcid.org/0009-0006-5850-3761