ALZHEIMER'S DISEASE PREDICTION USING AN ENHANCED FEATURE SELECTION OF MULTI-LOCAL INFORMATION SHARING BASED ON THE ABC ALGORITHM

I. Murali Krishna[1], Sri Silpa Padmanabhuni[2], Uppaluru Sai Sri Vaishnavi[3], Munagala Priyanka Balaji[4], Sampangi Tulasi[5], S L Swarnamalya Anumalasetty[6]

[1,2,3,4,5,6]Department of Computer Science and Engineering, PSCMR College of Engineering and Technology, Vijayawada, A.P.

Abstract

Alzheimer's disease, a type of dementia that prevails significantly in society, presents an unprecedented challenge for individuals and society. Its gradual progression is marked by memory loss and cognitive decline, necessitating early detection. While machine learning and *several existing ABC-based feature selection models* have been investigated for this goal, efficiency remains a major concern. This paper focuses on the pressing requirement for an effective and precise Alzheimer's disease prediction system that can reduce classification errors while improving model efficiency and minimizing complexity through feature selection techniques. It utilizes the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) dataset incorporating 94 attributes with pre-processing steps, including label encoding and outlier detection via the Interquartile Range (IQR) method. The proposed model deploys a random forest classifier to enhance accuracy alongside iterative refinement techniques. The primary purpose is to improve accuracy in predicting diseases and minimize classification errors, consequently boosting efficacy. To accomplish these objectives, a Modified Artificial Bee Colony (MABC) algorithm has been suggested for feature selection while simultaneously addressing local optima problems and reducing complexity. The MABC algorithm's innovation lies in its utilization of multiple Gaussian Distribution-based local information sharing that provides feature subset evaluation with more robustness. Random forest classification was performed iteratively after every feature selection process to optimize results. Over 25 cycles with 49 ultimate selected features, the proposed model accomplished outstanding performance by yielding an accuracy of 98%, demonstrating a remarkable improvement from pre-existing methods. Such progressiveness pushes Alzheimer's disease prediction efficiency forward by lessening required resources through reduced positivity rates and fewer designated features towards remaining efficient at scale among healthcare providers in the remotest geographies. Consequently, this paper introduces implementing MABC algorithms for enhanced Alzheimer's disease prediction accuracy without compromising on system-wide cost-effectiveness values along with rapidity concerns that are critical factors affecting low-resource settings' healthcare provision decision-making processes, thus making it incredibly impactful work deserving significant attention.

Keywords: ABC, Modified ABC, local information sharing, IQR, Random Forest, ADNII2.

1. Introduction

Alzheimer's disease is a commonly prevalent neurodegenerative disorder that has ravaged the lives of millions across the globe. The ramifications of this condition are immensely

progressive, characterized by gradual deterioration of memory and thinking capacity, and other cognitive and MRI scans from patients were analyzed as insidious threats to individuals and society due to its pervasive implications on mental well-being. The onset of Alzheimer's can usually be seen when individuals start experiencing short-term memory loss or difficulty in retaining new information. As Alzheimer's disease advances, patients may encounter significant challenges while attempting to solve complex problems or recognize familiar faces. Genetics is a major player in determining an individual's chances of succumbing to this catastrophic ailment due to specific mutations or gene variations within one's genetic makeup. Other dangerous elements contributing to its emergence are age-related incidents, traumatic head injuries, hypertension complications, cardiac diseases such as myocardial infarction or heart attacks, type 2 diabetes mellitus, and an unhealthy diet regime. Early-stage detection and prediction are crucial. To predict Alzheimer's Disease, there exist many more technologies. For instance, A previous research conducted a comparative study using different algorithms, including Linear Discriminate Analysis, Support Vector Machine, Random Forest, and Decision Tree, which was applied to Magnetic Resonance Imaging (MRI) data for early disease detection [13]. The research conducted brain image segmentation based on an Artificial Bee Colony Algorithm as a feature selector to significantly decrease model complexity without reducing performance. Thuss achieving improvement over ordinary classifiers like Convolution Neural Networks [4].

Various models, including Machine Learning and Evolutionary Algorithms in supervised/independent custom neural network coordinated Feature Constructing Attempts were taken to detect Alzheimer's Disease [14]. Another similar approach analyzed clinical and demographic features and patient MRI scans by comparing models such as Logistic Regression, K-Nearest Neighbors, and Gaussian Naive Bayes to determine mild cognitive impairment or Alzheimer's Disease [3]. One of the studies conducted experiments with multiple machine learning models, including Support Vector Machines (SVM), random forests, and ensemble methods, to diagnose Alzheimer's disease based on Magnetic Resonance Imaging data. Their results demonstrated minimal effectiveness in proving the disease [11]. One technique used for detecting Alzheimer's Disease is through proteomic analysis of serum and plasma samples. Investigation of the potential of using protein biomarkers found within these sample types for accurately detecting mild cognitive impairment (MCI), an early-stage symptom of Alzheimer's disease seen in many elderly populations, including those with Down Syndrome [5]. Another possible approach explored recently includes modular machine learning-based classification based on retinal vasculature data collected using non-invasive methods such as Optical Coherence Tomography Angiography. Their results indicated their algorithm managed diagnosis issues up to minimum accuracy [1].

However, unfortunately, all these existing technologies cannot detect Alzheimer's Disease quickly and efficiently. The development of machine learning algorithms capable of accurately predicting human diseases solely based on their symptoms is a significant advancement in medicine. Machine learning is a subset of artificial intelligence that focuses on developing algorithms capable of autonomously identifying patterns and making predictions based on data inputs. One significant advancement in this area is the development of machine learning algorithms that can accurately predict human diseases solely based on their symptoms.

However, with several existing ABC-based feature selection models such as Adaptive Local Information ABC Feature Selection, Binary ABC Feature Selection, and Modified ABC Feature Selection suffering from the refinement problem over multiple dynamic changes in higher-dimensional issues, achieving optimistic feature subsets over high-dimensionality datasets poses a challenge. To overcome this limitation, we propose an Enhanced Feature Selection using Multiple Local Information Sharing-based ABCs that facilitates selecting optimistic feature subsets. The outcomes obtained through our approach exhibit superior results for disease prediction. The main motivation for this study is that, concerning this underlying issue, there is an immense need to develop a system that can effectively work without excessive equipment and skilled medical practitioners and is also less time-consuming. Therefore, this application can be effectively used for diagnosis.

Figure 1: Technologies used to detect Alzheimer's Disease

1.1. Artificial Intelligence

Artificial intelligence is an emerging field that involves designing computers and machines that mimic human cognitive abilities such as learning, reasoning, and decision-making. AI can improve medical diagnosis and treatment outcomes in several ways. One of these applications is detecting Alzheimer's disease. AI technology can aid physicians in the early detection of Alzheimer's by analyzing various biomarkers, such as brain scans or neuroimaging, along with patient data from wearable devices or electronic medical records.

1.2. Deep Learning

Deep learning (DL) is a subfield of machine learning that has emerged as one of the most promising techniques for developing artificial intelligence applications. It involves training complex neural networks to recognize patterns and make decisions based on large datasets. DL methods can accurately diagnose Alzheimer's disease early by analyzing brain imaging data such as MRI scans or PET images.

1.3. Machine Learning

Machine learning has emerged as a powerful tool to improve healthcare delivery in recent years. The vast amount of data generated by the Healthcare sector can be utilized in various ways to improve patient outcomes. One such method is predicting future diseases and their onset times. The study aimed at using machine learning algorithms to classify whether patients

had Alzheimer's Disease or not. This was done by evaluating large-scale patient data and feeding it into an algorithm that could learn from this information and make predictions about future cases of Alzheimer's disease. One significant advantage of using machine learning for predictive analysis is its ability to deal with complexity across multiple variables. Instead of relying on single indicators, researchers could analyze numerous factors simultaneously, like medical history, genetic markers, social demographics, etc., contributing significantly towards better accuracy in prediction models. There are several swarm intelligence techniques in machine learning, such as ABC, PSO, ACO, etc.

1.4. Swarm Intelligence-Based Algorithm

Swarm intelligence is another name for collective intelligence. Scientists and ecologists have been studying the behavior of social insects due to their effectiveness in solving challenging problems like determining the simplest path between their nest and food source or organizing their nests. Although these insects lack knowledge, they interact as a hive to create supernatural happenings by relating to each other and their environment. A numerical optimization technique is utilized to imitate different swarm's behaviors to seek food or pairing. Swarm intelligence-based algorithms are simple and robust strategies that efficiently identify the most effective answer for optimization issues without requiring important mathematical struggle. The three main critical aspects of swarm intelligence are:

Behavior: It is hard to predict actions given a set of instructions.

Knowledge: Without knowing how a living thing performs, it cannot see how a colony functions.

Sensitivity: Modifications to the basic rules, no matter how modest, alter how groups function. Swarm intelligence-based algorithms and representative algorithms for discrete and continuous optimizing issues.

1.5. Artificial Bee Colony (ABC)

The Artificial Bee Colony (ABC) algorithm is a swarm-based meta-heuristic algorithm for optimizing mathematical issues. It was influenced by how honey bees used cognition in their gathering. The algorithm is mainly built on the model for honey bee colony foraging behavior. In the ABC algorithm, bees change their search phases by playing the following roles of bees: employed bees, onlooker bees, and scout bees. This algorithm is a continuous optimization problem. This algorithm is purely based on the position of the bees. The ABC approach has become prominent because of its robustness and ease of application. ABC has an early convergence fault, which causes ABC's discovery and extraction capabilities to be lost. This article taught us about the operation and likeness of the Artificial Bee Colony algorithm to a bee colony and its application in problem optimization. Several improvement techniques of ABC are BABC, HABCDE, and ibinABC.

1.6. Binary Artificial Bee Colony (BABC)

ABC must be converted into a binary version that may be used as a feature selection. Food sources are a possible practical solution in the original algorithm. These food sources were regarded as a likely optimal feature subset for the solution in the binary version and were created as bit vectors. The bit vector has the dimension N, where N represents all the features. The bit value shows if that particular feature is a part of the feature subset in this bit vector form. If the value is one, a component feature subset combines the quality provided to the critical place. The feature has been removed from the feature subset if it is irrelevant.

1.7. Hybrid Artificial Bee Colony with D.E. (HABCDE)

It provides a hybrid version of ABC and D.E., including the ABC position update solution modification. HABCDE is the name assigned to the proposed hybrid algorithm. ABC's employed, onlooker, and scout bee phases have been modified in HABCDE. During the employed bee phase, a bee will travel to another food source depending not simply upon a selected food source but also on the most current optimal food source. Gbest-guided ABC has already implemented the best solution knowledge to maintain every bee's position. During the observer bee phase, it changes the bee's position using the Differential Evolution (D.E./best/1/bin) algorithmic approach. During the scouting phase, the number of scouts increases to enable all bees that have not been maintained a predetermined number of times to re-initialize themselves.

1.8. Improved Binary Artificial Bee Colony (ibinABC)

This technique is an upgraded variant of binABC, which initially updated a single dimension, one of D, of the decision variable in every operation. However, other swarm intelligence algorithms suggest updating many variables over the entire decision-making vector. While attempting the improvements, there is a trade-off between equal amounts of discovery and extraction. ibinABC tries to balance exploration and exploitation by using an exponentially computed rate, dt, inspired by updating several dimensions per operation. At iteration t, the number of features will be updated. A dynamic normalization factor is imposed by ibinABC.

1.9. Modified Artificial Bee Colony (MABC)

The Modified Artificial Bee Colony (MABC) algorithm is a nature-inspired optimization technique widely used to solve complex engineering problems. The MABC works by simulating the foraging behavior of honeybees in search of food sources. In this case, the bees represent candidate solutions, while the nectar represents objective function values. MABC algorithm provides a competitive performance in solving high-dimensional optimization problems while being computationally efficient at the same time.

2. Discussion

Jianqiao Tian et al. [1] proposed a modular machine-learning approach for classifying Alzheimer's disease based on retinal vasculature. They used transfer learning to extract features from the retina's optical coherence tomography angiography images. Then, they combined them with demographic information to classify individuals as either Alzheimer's disease patients or healthy controls. Their results showed that this approach achieved an accuracy of 94% in identifying individuals with Alzheimer's disease, highlighting its potential as a promising tool for early detection and diagnosis. The authors also compared their method with other state-of-the-art techniques, such as deep neural networks, random forests, support vector machines, and logistic regression models. The modular approach outperformed these methods and identified important biomarkers associated with Alzheimer's disease progression. This study suggests that analyzing retinal vasculature through advanced machine learning algorithms can be a non-invasive and effective way to diagnose AD early on in patients without requiring invasive procedures like brain imaging scans. This research will pave the way further for studying neurodegenerative diseases using remote sensing at low cost while avoiding unnecessary testing or treatment strategies.

Aakash Shah et al. [2] examined the effectiveness of various machine-learning techniques in detecting Alzheimer's disease at an early stage. The authors focused on comparing different methods, including random forest models and deep learning models, to determine their accuracy and stability in identifying biomarkers associated with the onset of Alzheimer's disease. Using medical data from patients diagnosed with either normal or impaired cognitive function, Shah found that by analyzing certain brain regions, such as the hippocampus and lower medial hemisphere, using MRI scans, approximately 80% of people with AD were accurately identified. Furthermore, the study discussed how preclinical screening can help identify individuals at greater risk for developing AD based on their medical history and health status information. Early detection through these intervention strategies could delay the onset of A.D pathology identification so that treatment approaches can be improved during the initial stages. The findings demonstrated a successful application of machine learning algorithms like the Random Forest model for predicting dementia diagnosis among adults aged 65+.

Morshedul Bari Antor et al. [3] compared machine learning algorithms to predict Alzheimer's disease (A.D.). They analyzed the effectiveness and accuracy of the Support Vector Machine, Naïve Bayes, Decision Tree C5.0, K-Nearest Neighbor with Leave One-Out Cross-validation (LOOCV), Principal Component Analysis dimension reduction followed by k-means clustering algorithm with LOOCV (PCA+k-Means), Artificial Neural Network algorithms in predicting A.D. using magnetic resonance imaging data from 228 subjects. The authors found that PCA+k-Means had the highest predictive power in identifying A.D., based on an Area Under Receiver Operating Character Curve value of 0.9787 and a sensitivity/specificity rating of 97%. Moreover, they explored the significance and usefulness of each feature for predicting A.D., where they discovered that smaller hippocampal volumes were common among people affected by this disease compared to controls, along with changes that occurred during Aging, including neuropsychological decline.

Mumine Kaya Keles et al. [4] used the Artificial Bee Colony algorithm as a feature selector in classifying Alzheimer's disease using brain volumetric data. They found that this method achieved better results than other feature selection methods, such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE). The study collected magnetic resonance imaging (MRI) images from 202 participants with Alzheimer's disease and 194 healthy controls to identify differences in the hippocampus volume. Statistical analysis revealed significant differences between the two groups regarding left, right, and total hippocampus volumes. Keles also found that Artificial Bee Colony performs better-concerning classification accuracy than other algorithms such as Support Vector Machine, K-Nearest Neighbors, and Decision Tree Classifier. Additionally, combining the ABC algorithm's performance with different classifiers leads to higher precision, where they reported an accuracy of 94% for SVM on their dataset. Therefore, Keles concluded that using the Artificial Bee Colony classifier determined which features contribute to identifying Alzheimer's Disease patients more effectively than PCA or RFE techniques alone.

Melissa E. Petersen et al. [5] analyzed proteomic profiles for Alzheimer's disease (AD) and mild cognitive impairment (MCI) among adults with Down syndrome, spanning serum and plasma. AD is prevalent in individuals with Down syndrome, often leading to early-onset dementia. The researchers used an Alzheimer's Biomarker Consortium-Down Syndrome (ABC-DS), which collected data from 254 participants across four clinical sites. They

measured levels of 184 proteins from both serum and plasma samples using the SOMAscan assay. The results showed that several differentially expressed proteins could be utilized as potential biomarkers for AD or MCI diagnosis in adults with Down syndrome, including amyloid-beta peptides, complement factors, apolipoproteins E and J, etc. Furthermore, it discovered proteome differences between APOE ε2 carriers versus non-carriers within the DS-MCI group.

Malavika G et al. [6] explore various machine learning algorithms that can be used to predict Alzheimer's disease incidence and diagnosis. The study highlights how preclinical screening of individuals at risk for AD may lead to improved treatment approaches, and early identification of AD pathology can delay onset. The article notes that machine learning models can identify biomarkers associated with Alzheimer's, such as the hippocampus area, emotions, and lower medial hemisphere. The Random Forest model performed similarly to deep-learning models in identifying persons with good cognitive function from those with Alzheimer's. Moreover, magnetic resonance imaging (MRI) was noted as an effective diagnostic tool that determines the volume of each brain region, making accurate diagnosis possible through an auto-encoder model built using CNN, demonstrating a classification accuracy range of 86.6%.

Ji Hwan Park et al. [7] studied different operational definitions of incident Alzheimer's disease: "definite A.D." with diagnostic codes and dementia medication and "possible A.D." with mere diagnosis. Preclinical screening of persons, according to medical and health data, at risk for Alzheimer's disease may lead to Improved treatment approaches, and earlier A.D. pathology identification can delay the onset of AD1-3. We identified 614 different people with A.D. incidence in 2002 among 40,736 people over the age of 65. This database's extensive, longitudinal administrative healthcare data was used to forecast the occurrence of Alzheimer's disease in the future.

D. Chitradevi et al. [8] analyzed optimizing techniques to diagnose this debilitating condition. They explored methods like particle swarm optimization (PSO), firefly algorithm (FA), and decision tree analysis. The results showed that PSO outperformed other algorithms in classifying brain images affected by AD-related disorders. Furthermore, they found that decision trees could accurately organize most cases based on the features extracted from magnetic resonance imaging (MRI). Similar to previous research, Devi and Prabha highlighted the importance of preclinical screening for early identification of Alzheimer's disease. By detecting risk factors, individuals can benefit from improved treatment options while delaying onset as much as possible. Identification using MRI is also emphasized as an accurate means for assessing brain regions affected by AD pathology.

C. Kavitha et al. [9] investigated the effectiveness of machine learning models in predicting early-stage Alzheimer's disease (AD). They used a dataset consisting of medical and demographic data from 10,000 patients. The authors found that age, depression history, hypertension, and diabetes significantly contribute to AD onset. The study aims to optimize predictive accuracy while minimizing feature complexity by comparing several supervised classification methods such as logistic regression model (LRM), support vector machines (SVMs), K-nearest neighbors(K-NN) algorithm with ensemble-based classifiers like Random Forest (RF) and Gradient Boosting (GB). Among these five algorithms, RF performed better than others for diagnosing AD, with an overall accuracy rate of 84%. The findings suggest that early diagnosis can be made by screening individuals with high-risk factors through regular

health checkups or primary care settings using advanced algorithms trained on Electronic Medical Records as part of clinical workflow integration strategies.

Minseok Song et al. [10] Verified that the Random Forest model, which had fewer features and higher stability and robustness, performed similarly to deep learning models. It shows that Alzheimer's disease biomarkers include the area of the hippocampus, the emotions, and the lower medial hemisphere. We verified that the Random Forest model, which had fewer features and higher stability and robustness, performed similarly to deep learning models. Using total-tau, phosphor-t Tau, and amyloid, more than 80% of persons with Alzheimer's disease and 90% of people with good cognitive function can be differentiated. A.D. can be accurately diagnosed using magnetic resonance imaging (MRI), which determines the volume of each brain region. An auto-encoder model built on CNN demonstrated 86.6% classification accuracy for separating controls from A.D. in a test using the MRI data of 695 participants.

M. Tanveer et al. [11] explored using machine learning techniques to diagnose Alzheimer's disease (AD). The study discussed different aspects of AD, including operational definitions, preclinical screening methods, and biomarkers that can aid in early detection and accurate diagnosis. Machine learning algorithms were used to analyze medical imaging data, such as magnetic resonance imaging (MRI) scans, to detect changes in brain structure associated with AD pathology at an early stage. The authors also reviewed various studies that employed machine learning models for predicting the probability of developing AD based on demographic factors like age and gender. The article highlights recent advancements in deep neural network architectures, such as convolutional neural networks (CNNs), which have shown promising results in detecting AD pathology from MRI images with high accuracy rates. Random forest models were also useful due to their fewer features but increased stability and robustness compared with deep learning models.

K.G. Achilleos et al. [12] conducted research to extract explainable assessments of Alzheimer's disease (AD) through machine learning on brain MRI imaging data. They used the Alzheimer's Disease Neuroimaging Initiative public dataset and extracted features from two regions of interest - hippocampus and entorhinal cortex - linked with early stages of AD progression. They then employed four different classifiers: Support Vector Machines, k-nearest Neighbor classifier, Random Forests algorithm, and Gradient Boosting Trees to classify patients as normal or having mild cognitive impairment or AD according to their neuropsychological features derived from Mini-Mental State Examination scores such as attention, language usage, etc. Gradient Boosted Trees had the best performance among these algorithms, indicating an average accuracy rate of 87% for this particular problem setting. Furthermore, they also employed a visually interpretable decision tree model called Bayesian Rule List (Rule Fit), which achieved comparable results while providing transparent, rule-based explanations regarding how individual factors contribute linearly to risk prediction compared with classical non-transparent models like neural networks without any clear understanding between inputs-output relationships.

Afreen Khan et al. [13] developed a three-tiered cognitive hybrid machine-learning algorithm to diagnose Alzheimer's disease (AD) effectively. The proposed approach has three tiers: symptom-based screening using demographic variables and Neuropsychological tests, imaging-backed testing with traditional machine learning models, and incorporating deep learning techniques in Tier 3. Their study identified significant predictors that can effectively

diagnose AD within Tier 1 using gender-related clinical demographics and age/education-dependent composite scores of cognition assessments on MMSE-2. In addition, their results showed significant correlations between specific areas, such as hippocampal volume reductions or emotional connectivity index deterioration touched upon by previous research studies by Khan and on extensive analysis of personal healthcare data among more than 10 thousand socially demographically heterogeneous participants exceeded an excellent AUC value at above 92%.

Moolchand Sharma et al. [14] explored using Machine Learning and Evolutionary Algorithms for diagnosing and detecting Alzheimer's disease (AD). They highlighted that early detection can aid in improving treatment approaches and delaying the onset of AD pathology identification. The authors identified various biomarkers such as cognitive tests, brain imaging techniques like Magnetic Resonance Imaging (MRI), cerebrospinal fluid analysis, and genetic testing to predict an individual's risk for developing AD. The study also investigated different machine learning algorithms such as Random Forests, Convolutional Neural Networks (CNNs), and evolutionary algorithms like Genetic Algorithm (GA) to tackle challenges in accurately predicting AD diagnosis. It was found that CNN-based models demonstrated excellent performance compared to traditional models using MRI data from 695 participants. Sharma et al.'s work emphasizes the importance of a multi-dimensional approach toward identifying individuals at high risk for developing AD while highlighting how technology can play a vital role in achieving this goal efficiently and effectively.

Yusera Farooq Khan et al. [15] developed an ensemble model for the diagnostic classification of Alzheimer's disease (A.D.) based on brain anatomical magnetic resonance imaging (MRI). The study aimed to investigate the diagnostic efficacy of MRI biomarkers in identifying A.D. patients and differentiating them from healthy controls using machine learning techniques. The researchers collected data from 100 participants, including 50 A.D. patients and 50 age-matched controls. The results showed that using a leave-one-out cross-validation technique, the proposed ensemble model achieved high accuracy in classifying A.D. versus control subjects with a mean area under the curve (AUC) value of 0.951 ± 0.038. The ensemble method combines several machine learning algorithms by taking their average prediction to improve accuracy and reduce risks associated with single models' variations. This study compared three other non-ensemble systems: Random Forest Tree, Gradient Boosting Decision Trees, and K-nearest Neighbour algorithm against the Ensemble Model. The results demonstrated superior performance by the ensemble methods, achieving higher accuracy than all individual models considered during the analysis.

Table 1: Merits and demerits of various conventional models in the detection of AD

S.no	Author	Techniques	Merits	Demerits
1	Jianqiao Tian [1]	The author proposed a modular machine learning classification model to predict the A.D.	The study is successful in differentiating images of AD from healthy control images.	The model accuracy needs to be increased and more efficient to predict the A.D.

2	Aakash Shah [2]	The author proposed machine learning models to predict the A.D.	This study has found that the accuracy of Alzheimer's disease prediction improved significantly.	The main limitation is that the model has excessive features, which may lead to improper results.
3	Morshedul Bari Antor [3]	The author proposed a machine learning model to predict the disease effectively.	This study has found that the accuracy of the model is increased.	The main limitation of this model is the limited sample size.
4	Mumine Kaya Keles [4]	The author proposed a Binary Artificial Bee Colony to predict Alzheimer's Disease.	The suggested model is very effective in giving a low complexity rate.	This method has the limitation of local optima problems over a high dimensionality dataset.
5	Melissa E. Petersen [5]	The author proposed proteomic analysis of serum and plasma samples to predict symptoms of A.D.	The proposed framework is very successful and efficient at predicting Alzheimer's disease.	The complexity levels of a model are very high.

The key incentive of the largest cause of Alzheimer's disease is that it can lead the person towards complete memory impairment and exposure to carcinogens if an early prediction does not occur. Concerning this underlying issue, there is an immense need to develop a system that can effectively work without excessive equipment and skilled medical practitioners and is also less time-consuming. As a result, clinicians in rural areas can also efficiently use this application for their diagnosis. Observing various conventional methods faces the local optima problem in feature selection over high-dimensional datasets using artificial bee colony swarm intelligent algorithms and classification models, which leads to low accuracy and high error rates due to improper subsets of features. It is also necessary to achieve the best results with our proposed model.

3. Proposed Methodology

The main spotlight of the proposed model is optimized feature selection for building a classification model in ways of early detection of AD. The proposed model consists of three stages. In the initial stage, pre-processing techniques are applied to remove noise and unwanted data from the dataset. Then, we implemented a Multiple local sharing based MABC feature selection algorithm to select the optimistic features over a high dimensional AD biomarkers data set. In the final stage, Build the convergency-based Random Forest classification model to predict AD locations. In the Fig 2 depicts the Alzheimer's Disease Prediction Using an Enhanced Feature Selection of Multi-Local Information Sharing Based on The ABC Algorithm.
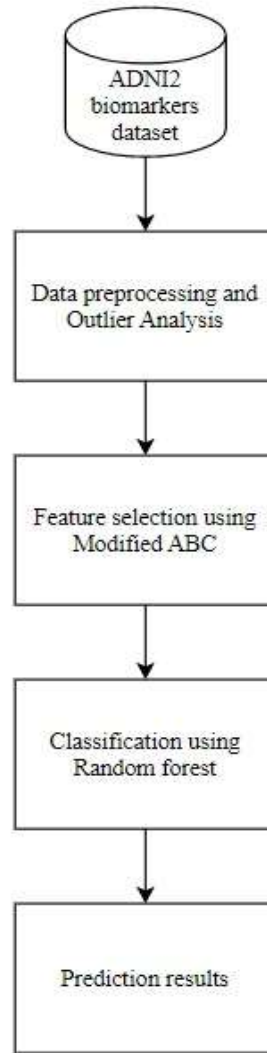
Figure 2: Proposed classification model with enhanced Feature Selection of Multi-Local Information Sharing Based on The ABC Algorithm

Stage 1: Pre-processing and Outlier Analysis

Stage 2: proposed Multiple local sharing based MABC for optimistic feature selection.

Stage 3: Proposed Iterative Random Forest Classification

3.1. Pre-processing and Outlier Analysis stage

In this pre-processing stage, noisy data are removed, and the categorical data values are converted into numerical using standard encoding in a system-understandable format. Machine learning algorithms can subsequently prefer the functionality of such labels. It is an essential step in the defined dataset to training data pre-processing. The raw data was

transformed from object format to string format to accommodate the data with the proposed classification model.

Algorithm:

```
for each data sample as df
do
        for each column as c in df
        do
                if df[c]==obj then
                        df[c]=str(df[c])
                else if iscategorical(df[c])==True then
                        df[c]=Num(df[c])
        End_for
```

Outliers are data points that are different from the rest of the data in some way. Outliers in the data could make model fitting difficult. The IQR method computes lower and upper bounds to detect and remove outliers. After detecting the outliers, we used the IQR score to remove the outliers from the data.

3.2. Multi-Local Information Sharing Based Modified ABC feature selection

In the proposed model, we suggest A Modified Multiple Local Information Sharing Based Artificial Bee Colony Algorithm for Effective Feature Selection. Instead of using a simple random partner solution to update evaluation value, we propose information sharing based on multiple partners using Gaussian Distribution. The effectiveness of the proposed methodology and measure regarding the length of feature subset, limit, iterative runtime and various objects. The artificial bee colony (ABC) algorithm is an optimization approach implemented for several real-world issues. It builds honey bees' foraging behavior. Karaboga proposed ABC in 2005, making it one of the swarm intelligence algorithms. Bees modify various evaluation steps in the ABC algorithm by choosing among three bee parts: employed bees, onlooker bees, and scout bees. The Artificial bee colony algorithm evaluates bee food sources as possibilities, determining that the most crucial food source is the optimal solution. The given objective function assesses the significance of food sources.

Using the simple algorithm, the employed bees begin looking for food sources. According to evaluations of the employed bees, the onlooker bees select an essential food source. If the employed bee can't detect a food source near its area, it becomes the scout bee. Scout bees randomly select their new sources of food in the target region.

Algorithm:
Step 1: Initial phase
In the initial step, create food sources for the employed bees. The food source produces a random generator that generates a Gaussian value between lower and upper bounds or between {0,1}. $\quad X_i = L + guass(U - L) \rightarrow (1)$

Where Xi = X1, X2, X3…., L = lower bound = -7,

U = upper bound = 7

Gauss value = upper bound – lower bound (randomly chosen)

$$it = \frac{1}{1 + f(x)} \rightarrow (2)$$

Where f(x) is an algebraic equation, where we should substitute our X_i values in the equation.

Table 2: Initial phase of MABC

f(x)	Population size	No of cycles	Dimension	Limit
$X_1^2 + X_1X_2 + X_2^2$ where - 7<=X_1,X_2<=7	14	10	2(X1,X2)	1

F(x) = function of algebraic expression = $X_1^2 + X_1X_2 + X_2^2$ where -7<=X_1,X_2<=7

Dimension D = X1. X2 = 2

Limit = 1 (should always be 1)

number of employed bees = number of onlooker bees = food source = 7

Population size = 14

No of cycles (iteration) = 10

Step 2: Employee bee phase

The bees look around the surroundings during this process, and when they locate a source of food that contains extra honey, they switch to the newer, better food source.

$$X_{new} = X_i + \emptyset(X_i - X_p) \rightarrow (3)$$

where $\emptyset$ is the $\max\{Gaussian\ distribution(\mu, \sigma), roulettewheel\_rand(0,1)\}$

μ is the mean and σ is the standard deviation of the ASF Biomarker feature

Trial updation: if the solution improves, then set trial=0. If the solution does not improve, then set trial =1.

Table 3: Computation in the Employee Bee Phase

X_1	X_2	f(x)	fit	trial
6.9110053	-2.948792	76.83648678	0.012847445	0
5.485536	-1.663606	23.73291953	0.040431943	0
-4.673956	6.046124	30.14216256	0.032110807	0
-4.319966	5.044774	22.31859879	0.042884223	0
-4.84029	6.073172	68.70773912	0.014345609	0
3.306352	0.810768	14.2699927	0.065407784	0
1.98168	3.197152	20.48653605	0.046540778	0

Since we perform iteration for 1 employed bee, we must perform iterations for another 6 employed bees.

Step 3: Onlooker bee phase

Bees that are busy convey the information on their food sources to bees waiting outside the colony, and they use this information to select their food sources probabilistically. According to the probabilities computed using the fitness values of employed bees, an onlooker bee in ABC selects a food source based on the multiple random samples by sharing their information.

Calculate probabilities using

$$probabilities = \frac{fit_x}{\sum fit} \rightarrow (4)$$

Table 4: Computation in Onlooker Bee phase

X₁	X₂	f(x)	fit	trial
6.9110053	-2.948792	76.83648678	0.012847445	3
5.485536	-1.663606	23.73291953	0.040431943	2
-4.673956	6.046124	30.14216256	0.032110807	2
-4.319966	5.044774	22.680199	0.042229	1
-4.84029	6.073172	118.433117	0.008373	0
3.306352	0.810768	14.2699927	0.06548791	2
1.98168	4.370122	31.685257	0.030595	1

Since we perform iteration for 1 employed bee, we must perform iterations for another.

Step 4: Scout bee phase

The food source was eliminated when the surrounding areas had been thoroughly researched. When a source of food is searched, the trial counter is increased. If the trial count exceeds the fixed value, discard the food sources and choose the latest random food source. Scouts don't have to make sure the random sources of food are superior to the old sources of food or because it's within the surroundings of the previous one; as an initial phase, they choose a random food source.

Trial (1)>limit, so we apply the scouting phase.

3>1-true discard the trail(1) and update the new solution. So again, we have to find the X₁, X2, f(x) and fit values.

Table 5: Computations in the Scout bee phase

X₁	X₂	f(x)	fit	trial
6.9110053	-2.948792	76.83648678	0.012847445	3
5.591802	-1.663606	24.733280	0.040431943	3
-5.848435	6.046124	35.382502	0.027486	3
-4.417879	5.044774	22.680199	0.042229	3
-5.877153	6.795677	120.661386	0.008220	1
4.173125	0.810768	21.455753	0.044532	3
1.98168	4.370122	31.685257	0.030595	3

Again, we have to perform iterations.

Step 5: Best fitness of the food source
After completing the scout bee phase, it memorizes the best fitness solution to the
food source.

In the Fig 3 shows about the brief the proposed enhanced Feature Selection of Multi-
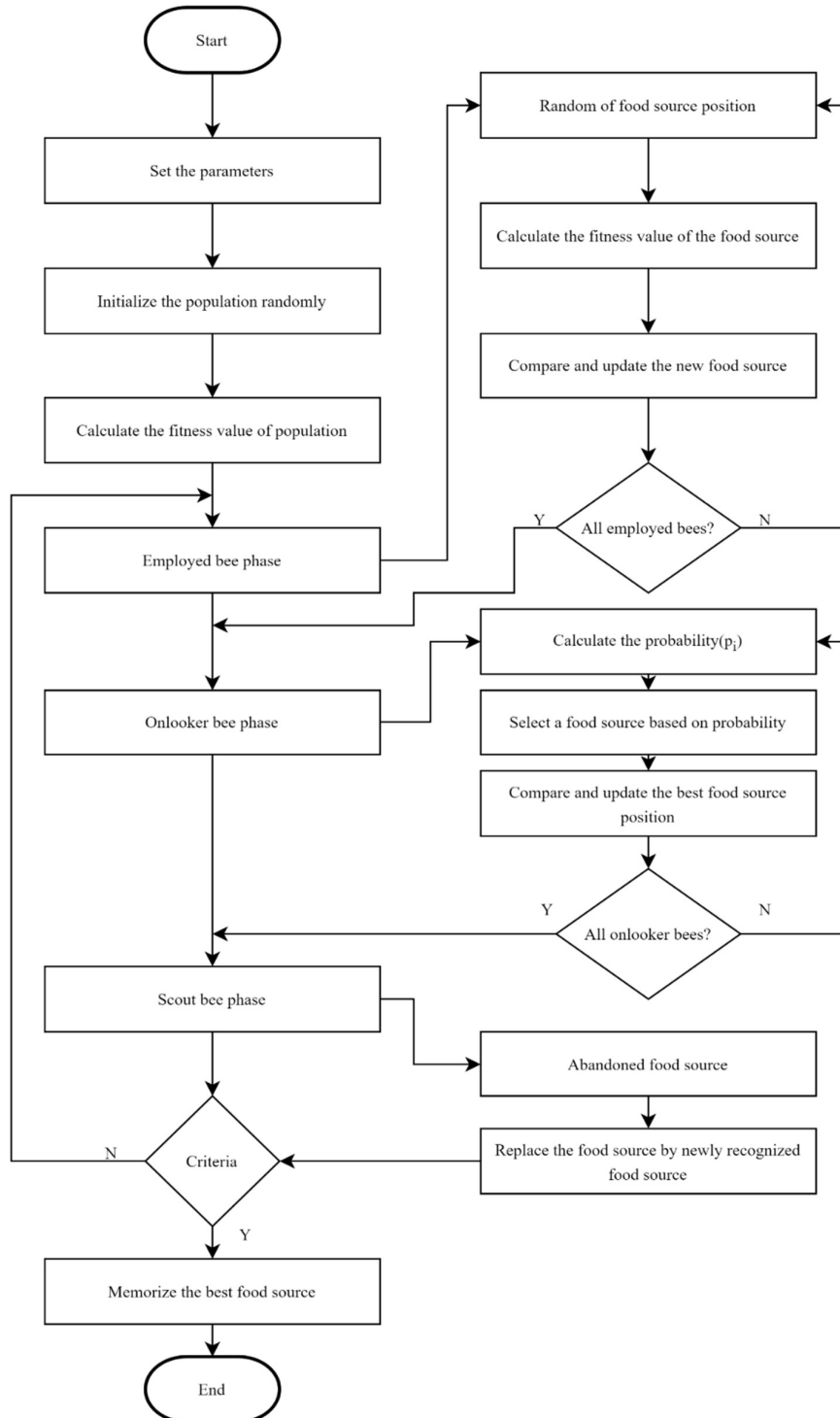Local Information Sharing Based on The ABC Algorithm

Figure 3: proposed enhanced Feature Selection of Multi-Local Information
Sharing Based on The ABC Algorithm

3.3. Proposed Ensemble Classification Model

A supervised machine learning classification method involves grouping input data into classes using one or more factors. Classification prediction modelling uses data as training input. We use a random forest classifier, i.e., it enhances the accuracy level of a dataset by aggregating the outcomes among various decision trees applied to different dataset samples. Here, we use the proposed iteration of the random forest classifier to get more accurate results. In iteration, it selects the subset of features for the biomarker high dimensional features and builds the classifier to predict the disease labels. The model is evaluated using test data, and the performance of the model is improved and reduces the risk factor of overfitting.
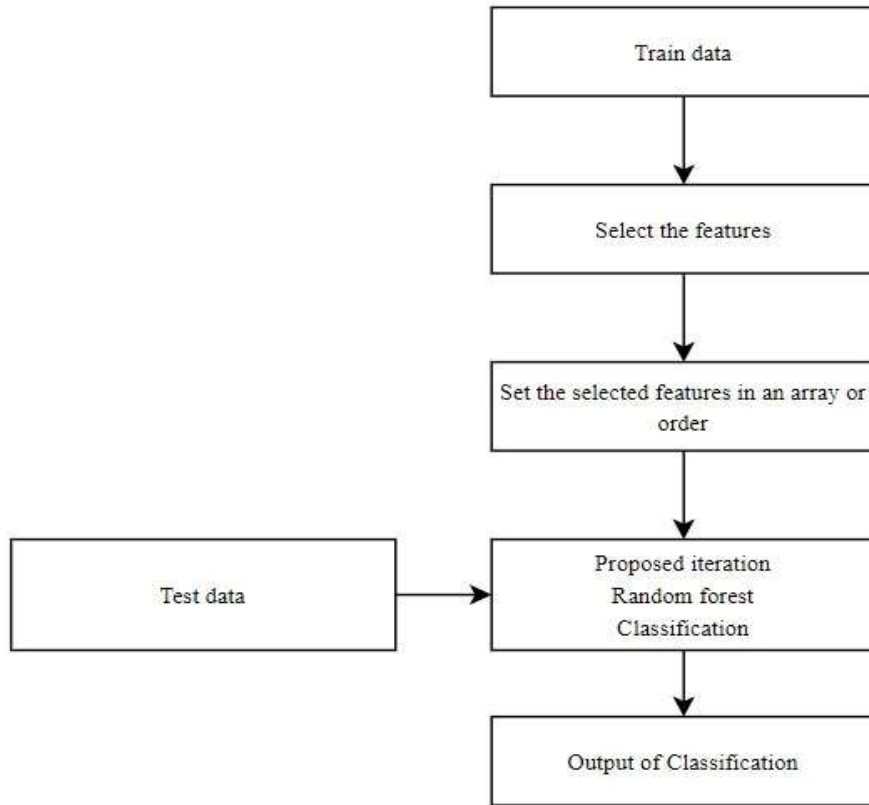

Fig 4: Proposed Iterative Random Forest Classification

4. Materials and Methods

Alzheimer's Disease Neuroimaging Initiative2 (ADNI2) is a pioneering initiative that brings together medical experts, neurologists, and researchers to document the progression of Alzheimer's disease. It aims to develop early diagnosis measures and evaluation methods for Alzheimer's Disease using various parameters such as genetic markers, imaging processing techniques, biological biomarkers, etc. This longitudinal survey aims to create a comprehensive

dataset with 94 attributes, including genetics data and other clinical metrics. The ADNI2 data provides an opportunity to study how specific genetic variants are associated with mental impairment in patients who have Alzheimer's disease. The ultimate goal of this research project is not just the identification of indicators for early detection but also to measure the progress or severity of disease progression over time.

5. Result & Discussion

The proposed model is designed and tested using Python libraries, and various performance metrics evaluate the model's performance. Our proposed model produces greater accuracy and lower runtime than traditional classification approaches. Model performance is assessed in accuracy, specificity, recall, precision, and f1 score. General Artificial bee colony algorithm faces the local optima problem during the bees computation to reduce the risk of local optima problem in ABC Computation in our proposed model Multi-Local Information Sharing Based Modified Artificial Bee Colony algorithm processes the data by randomly selecting multiple partners and sharing their optimistic computations and on each iteration produce a set of optimized features to builds the multi-class-based random forest. The model is tested over the test data and obtained better accurate results than conventional ML-based AD models, and it is observed that the tree count in the forest is increased. The Proposed model is built in an iterative approach using convergence to improve accuracy over the disease prediction rate.

Table 6: Depicts the No of Cycles and Accuracy of the Proposed Iterative Approach of Random Forest Classifier with an Artificial Bee Colony Algorithm.

Cycles	No of the features selected	Random forest Accuracy
0	36	0.9771750805585392
1	49	0.978249194414608
2	44	0.9511278195488722
3	30	0.9723415682062299
4	47	0.952207933404941
5	43	0.9774436090225563
6	41	0.943875570204082
7	42	0.9575725026852846
8	51	0.9742212674543801
9	39	0.9527389903329753
10	45	0.970130397421267
11	49	0.9742212674543801
12	43	0.9669709989258861
13	40	0.9487110633727115
14	56	0.9691192266380236
15	42	0.9575725026852846

16	44	0.924818030098783
17	34	0.9401181525241675
18	49	0.9618689581098896
19	40	0.9747583243823545
20	49	0.9616004296455424
21	43	0.9750268528464017
22	55	0.9683136412459729
23	44	0.9578410311493017
24	51	0.9701933404940923
25	49	0.9830827067669173

Based on the iterative cycles report, it observed that when the proposed model produces less than 42 features within 96 features, it is generated with low performance in terms of accuracy.

In each iteration cycle, it chooses a subset of features and generates an accuracy of 98.30% in the 25th cycle with 49 subsets of features. Based on the iterative process, by observing various cycles produced, more than 45 to 49 features are produced for better performance in terms of accuracy.

Table 7: Classification Report proposed in the 25th iteration over 49 optimal selected features

	Precision	Recall	F1-Score	Support
	1	0.56	0.71	9
	0.96	0.95	0.96	236
	0.97	1	0.99	1260
	0.99	0.99	0.99	799
	0.99	0.98	0.98	1279
	1	0.89	0.94	141
Accuracy			0.98	3724
Macro Average	0.99	0.89	0.93	3724
Weighted Average	0.98	0.98	0.98	3724

In Table 8, The proposed classification model performance is evaluated using a classification report, and the proposed model achieve better accuracy, 98.308 in the 25th cycle of model building with 49 biomarker features. The perform the better. How many estimations are correct, and how many need to be corrected? Performance metrics can evaluate the classification report. The formulae to calculate performance metrics are as follows.

Table 8: Depicts the Existing technique and Proposed approach analysis.

| Author	Technique	Number of Features	Classification Accuracy
Jianqiao Tian[1]	Modular machine learning-based classification methods like Optical Coherence Tomography Angiography.	5	90%
Akash Shah[2]	Machine learning algorithms	6	81%
Morshedul Bari Antor[3]	Machine learning algorithms	8	81.3%
Mumine Kaya Keles[4]	BABC	109	83.9%
Melissa E. Petersen[5]	Proteomic analysis of serum and plasma samples	6	85%
Malavika.G[6]	Machine learning algorithms	3	86.8%
Ji Hwan Park[7]	SABC	21	89.13%
D. Chitradevi[8]	ABC	5	90%
C. Kavitha[9]	ML framework and EHR-produced data	12	92%
Minseok Song [10]	Machine learning algorithms	63,29,22	90.2%
M. Tanveer[11]	RF-RSVM	9	90.53%
K.G.Achilleos[12]	Data Augmentation	10	91.4%
Afreen Khan[13]	CML	113	93.90%
Moolchand Sharma[14]	ABC	15	95.71%
Yusera Farooq Khan[15]	XGB + D.T. + SVM using Grid Search	3	95.75%
Our Proposed Approach	MABC	49	98.30%
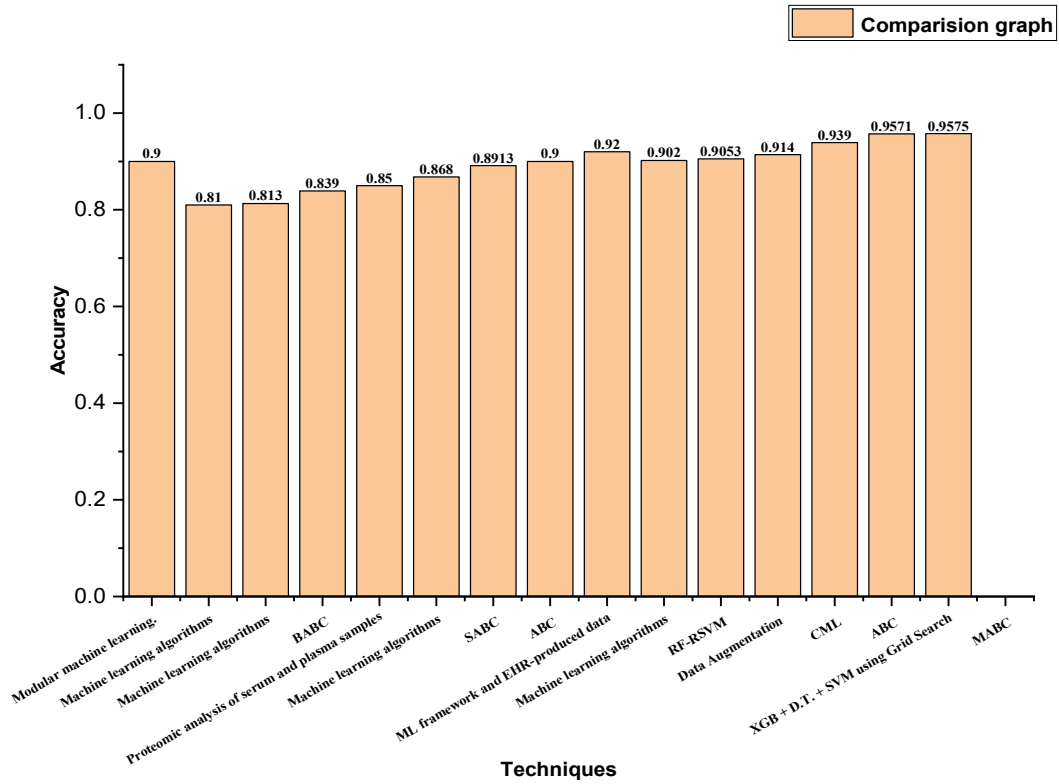
Figure 5: Graph for comparison of existing systems and proposed system

In Fig 5, the Modified Artificial Bee Colony (MABC) achieved greater accuracy than other systems. The x-axis indicates techniques, and the y-axis indicates accuracy. A bar graph may show some of the advantages of the model.

6. Conclusion

The major cause of the disease of Alzheimer's is that the individual gradually loses memory. In the early stages, traditional machine and deep learning methods cannot predict the disease effectively or efficiently. The model's findings have significant consequences for healthcare expenditures. Our methodology can help healthcare providers detect high-risk patients and build strategies for treatment for them. There is a considerable need to develop an approach that can perform well in the scope of the underlying issue in skilled medical partitioners. So, this software might well be successfully used by specialists who work in remote areas, and the cost expenditure of this model is low. Because the dataset contains few selected features, the algorithm can predict the disease quickly and precisely. The proposed methodology uses a modified artificial bee colony for the feature selection. The set features have been selected from the dataset. We can get the accuracy for the selected features using the random forest classifier. The proposed system achieved an accuracy of 98%, which is increased by 3% compared to the existing model, which is noteworthy in this field. Therefore, future work will concentrate on integrating with IoT technology.

References

[1]. Tian, J., Smith, G., Guo, H., Liu, B., Pan, Z., Wang, Z., Xiong, S., & Fang, R. (2021). Modular machine learning for Alzheimer's disease classification from retinal vasculature. In Scientific Reports (Vol. 11, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-020-80312-2

[2]. Shah, A., Lalakiya, D., Desai, S., Shreya, & Patel, V. (2020). Early Detection of Alzheimer's Disease Using Various Machine Learning Techniques: A Comparative Study. In 2020, the 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184). 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE. https://doi.org/10.1109/icoei48184.2020.9142975.

[3]. Bari Antor, M., Jamil, A. H. M. S., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., Singh, P., & Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. In J. Kang (Ed.), Journal of Healthcare Engineering (Vol. 2021, pp. 1–12). Hindawi Limited. https://doi.org/10.1155/2021/9917919.

[4]. Kaya Keles, M., & Kilic, U. (2022). Classification of Brain Volumetric Data to Determine Alzheimer's Disease Using Artificial Bee Colony Algorithm as Feature Selector. In IEEE Access (Vol. 10, pp. 82989–83001). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2022.3196649.

[5]. Petersen, M. E., Zhang, F., Schupf, N., Krinsky-McHale, S. J., Hall, J., Mapstone, M., Cheema, A., Silverman, W., Lott, I., Rafi, M. S., Handen, B., Klunk, W., Head, E., Christian, B., Foroud, T., Lai, F., Rosas, H. D., Zaman, S., ... Ances, B. M. (2020). Proteomic profiles for Alzheimer's disease and mild cognitive impairment among adults with Down syndrome spanning serum and plasma: An Alzheimer's Biomarker Consortium-Down Syndrome (ABC-DS) study. In Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring (Vol. 12, Issue 1). Wiley. https://doi.org/10. 1002/dad2. 12039

[6]. G, M. (2020). Alzheimer Disease Forecasting using Machine Learning Algorithm. In Bioscience Biotechnology Research Communications (Vol. 13, Issue 11, pp. 15–19). Society for Science and Nature. https://doi.org/10.21786/bbrc/13.11/4.

[7]. Park, J. H., Cho, H. E., Kim, J. H., Wall, M. M., Stern, Y., Lim, H., Yoo, S., Kim, H. S., & Cha, J. (2020). Machine learning prediction of the incidence of Alzheimer's disease using large-scale administrative health data. In npj Digital Medicine (Vol. 3, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41746-020-0256-0.

[8]. Chitra Devi, D., & Prabha, S. (2020). Analysis of Alzheimer's Disease using Optimization Techniques. In 2020, the Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII). 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII). IEEE. https://doi.org/10.1109/icbsii49132.2020.916755.

[9]. Kavitha, C., Mani, V., Srividhya, S. R., Khalaf, O. I., & Tavera Romero, C. A. (2022). Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. In Frontiers in Public Health (Vol. 10). Frontiers Media SA. https://doi.org/10.3389/fpubh.2022.853294.

[10]. Identification of Alzheimer's Disease with Random Forest Algorithm. In Brain Sciences (Vol. 11, Issue 4, p. 453). MDPI AG. https://doi.org/10.3390/brainsci11040453.

[11]. Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine Learning Techniques for the Diagnosis of Alzheimer's Disease. In ACM Transactions on Multimedia Computing, Communications, and Applications (Vol. 16, Issue 1s, pp. 1–35). Association for Computing Machinery (ACM). https://doi.org/10.1145/3344998.

[12]. Achilleos, K. G., Leandrou, S., Prentzas, N., Kyriacou, P. A., Kakas, A. C., & Pattichis, C. S. (2020). Extracting Explainable Assessments of Alzheimer's disease via Machine Learning on brain MRI imaging data. In 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE. https://doi.org/10.1109/bibe50027.2020.00175.

[13]. Khan, A., & Zubair, S. (2022). Development of a three-tiered cognitive hybrid machine learning algorithm for effectively diagnosing Alzheimer's disease. In Journal of King Saud University - Computer and Information Sciences (Vol. 34, Issue 10, pp. 8000–8018). Elsevier B.V. https://doi.org/10.1016/j.jksuci.2022.07.016.

[14]. harma, M., Pradhyumna, S. P., Goyal, S., & Singh, K. (2021). Machine Learning and Evolutionary Algorithms for diagnosing and Detecting Alzheimer's Disease. In Data Analytics and Management (pp. 229–250). Springer Singapore. https://doi.org/10.1007/978-981-15-8335-3_20.

[15]. Khan, Y. F., Kaushik, B., Chowdhary, C. L., & Srivastava, G. (2022). Ensemble Model for Diagnostic Classification of Alzheimer's Disease Based on Brain Anatomical Magnetic Resonance Imaging. In Diagnostics (Vol. 12, Issue 12, p. 3193). MDPI AG. https://doi.org/10.3390/diagnostics12123193.

[16]. Famila, S., Arulmurugan, A., Mahendar, A., Kalaiyarasan, R., Supriya, N. and Maridu, B., 2024. Transfer learning based feature extraction with metaheuristic optimization algorithm for detecting gastric cancer using optoelectronic sensor in endoscope. Optical and Quantum Electronics, 56(4), p.509.

[17]. Mahendar, A., and Dr K. Shahu Chatrapati. "Detection and Prevention of Cyber Attacks on Cloud-Based Data Centers using Machine Learning." International Journal of Computing and Digital Systems 12.1 (2022): 1063-1070.