

Optimizing Alzheimer's Biomarker Detection Using Superior Indicators in a Multi-Cloud Environment

Binu C. T.¹, Dr. S. Saravana Kumar², Dr. Rubini P.³

¹²³School of Engineering & Technology, CMR University

Abstract

Alzheimer's disease (AD) still poses a challenge in neurodegenerative diseases because of the multiple pathological processes involved and the presence of equivocal diagnostic features. The purpose of this research is to apply higher indicators with biomarkers to enhance the identification of Alzheimer's within a multi-cloud system. In this research, the main goal is to enhance the detection methods of Alzheimer's biomarkers by using highly advanced indicators and multi-cloud computing. The actual system, which is proposed in this paper, incorporates cloud-optimized solutions and state-of-the-art artificial neural networks, such as CNNs and RNNs. Biochemical biomarkers are essential, and thus, we carefully gathered a large amount of biomarker data and preprocessed multi-modal data, MRI images, and genetic and cognitive data. The several-cloud structure was intended to divide information jobs among the clouds, increasing the capability and dependability. The results thus establish the usefulness of the method in enhancing the accuracy of Alzheimer's biomarkers. The Superior Indicators Model reduces the validation loss equivalent to training loss. These results reflected that the preservation of the combined design received high predictive accuracy and minor loss to stage classifying of Alzheimer's disease, showing the model can also learn well and generalize. Applying it to a multi-cloud setting promoted positive results in relation to the event due to the availability of massively scalable computational frameworks for efficient processing of the data, improved security, and data redundancy. This discovery spotlights the ability of such a model incorporated with interconnected complex cloud systems and other superior machine learning algorithms in the early diagnosis and management of Alzheimer's disease. According to the author of the study, using this approach can provide more efficient diagnostic and therapeutic approaches to diseases and thus contribute to the improvement of patients' outcomes as well as the identification of the molecular mechanisms of the illness.

Keywords: Alzheimer's disease, biomarkers, multi-cloud environment, data analysis, cloud computing, machine learning, neurodegenerative diseases, data heterogeneity, diagnostic markers, cloud-native tools.

1. Introduction

1.1 Background Information and Context of the Study

Alzheimer's disease, also known as dementia of Alzheimer's type, is a chronic disease that is related to the progressive degeneration of the brain, resulting in dementia in the elderly. Alzheimer's disease is the most common cause of dementia, taking 60-70% of persons diagnosed with dementia globally (Alzheimer's Association, 2020). Now, given the continually increasing world population of people aged 60 and over, the number of people affected by AD is likely to increase drastically, which further exerts enormous pressure on healthcare frameworks to serve society's needs (World Health Organization, 2019). They have neurological symptoms that involve the build-up of amyloid-beta plaques and

neurofibrillary tangles that interfere with the normal functioning of neurons and result in cell death (Jack & Holtzman, 2013). Since the course of AD has been established to be strongly associated with early diagnosis, it is imperative to diagnose the condition early in order to enable management and early intervention that may slow the progress of the disease and improve outcomes for the patients. Biomarkers formed the bulk of the diagnostic tools with diagnostic utility because they represented aspects of AD's biological process. These biomarkers can be shed in body fluids like Plasma, serum, and CSF, as evidenced by neuroimaging (Counts et al., 2017). Modern developments in biomarker studies have given new hope for the early diagnosis of AD; nevertheless, the issue of biomarkers analysis and interpretation is still challenging because of the enormous

amount of data (Liu et al., 2014). Multi-cloud, specifically as a solution to these challenges, provides a dynamic, elastic, and robust computing environment through cloud computing. Multi-cloud helps in the sharing of data and computational jobs across different CSPs that

increase data processing and, at the same time, guarantee availability and reliability (Nguyen et al., 2020). As depicted in Figure 1 below, the given study employed the general multi-cloud framework.

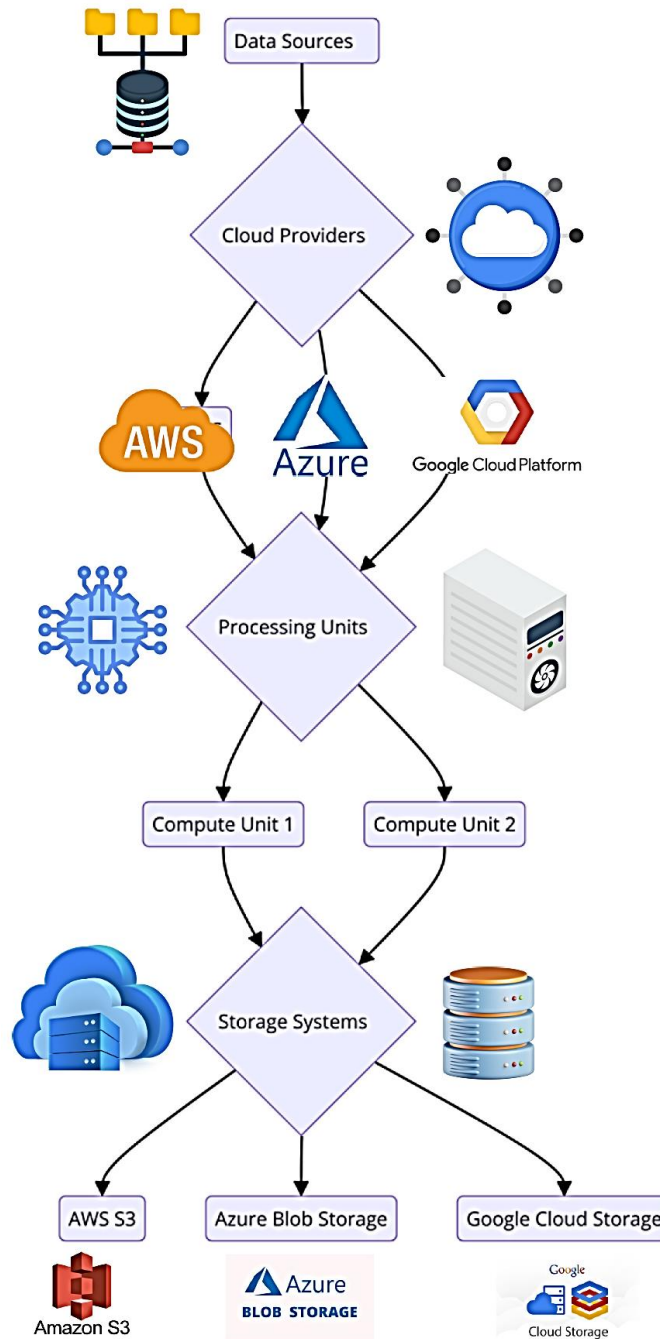


Figure 1: Multi-Cloud Architecture Diagram

The multi-cloud architecture illustrated in Figure 1 presents a suitable multi-cloud environment for the analysis of Alzheimer's biomarkers. This part of the architecture includes more than two CSPs, and each CSP provides particular services that facilitate

data processing, storage, and analysis. This information is obtained from neuroimaging scans, blood tests, and neuropsychological tests and transmitted to the cloud infrastructure. It ensures that the application has high availability, fault

tolerance, and scalability, which is essential when dealing with extensive data. This work utilizes the multi-cloud environment to enhance the identification and analysis of Alzheimer's biomarkers with the help of artificial intelligence and big data integration tools. A multi-cloud system includes the use of several CSPs, where resources and data are divided among cloud service providers. This entails the following advantages, especially in the biomedical sciences, where there is often a need for intensive computation and data handling. Combining the advantages of different cloud services, the multi-cloud approach improves the versatility, stability, and efficiency of data handling procedures (Amazon Web Services, 2020). From the perspective of the detection of Alzheimer's biomarkers, an implementation of a multi-cloud framework is beneficial, as this helps researchers to process and manage the plethora of types of data that are involved in different facets of the disease, including data from MRI scans, genetics and clinical investigations. Multi-cloud thus increases the availability since data is distributed across different cloud services and reduces the chances of data loss or downtime, which is

beneficial to extensive studies that require constant data processing (Google Cloud, 2021). Additionally, multi-cloud environments help to adopt the best computational resources and machine learning libraries. For instance, various cloud suppliers can have additional services like GPU computing, mass storage, and professional analysis tools in a research environment that can be easily merged with the existing cloud environment systems (Microsoft Azure, 2021). This integration is crucial for the analysis of biomarker data and enhancing the diagnostic models' efficacy and precision.

1.2 Problem Statement and Research Questions

Nevertheless, various issues persist in the identification and evaluation of biomarkers related to Alzheimer's disease, even with the progress made in biomarker research and cloud computing. The challenges are mainly related to data heterogeneity, data security, and the fusion of multiple data modalities from different sources. These challenges inhibit the attainment of precise and timely diagnosis and restrict the scaling of potential solutions (Fisher et al., 2019). Figure 2 shows that the described data flow is in operation across different multi-cloud scenarios.

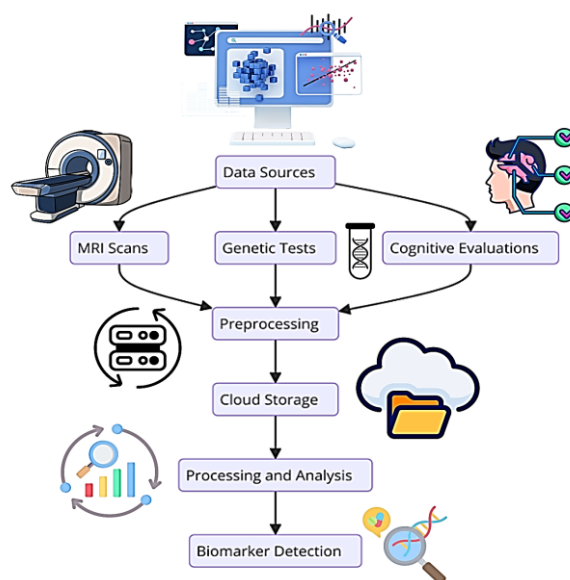


Figure 2: Data Flow Diagram

Taken together, the multi-cloud framework that governs the archiving and sharing of data within the organization is discussed in reference to Figure 2 below, which lays out the path the data takes from the time of collection and processing all the

way to analysis. Typically, the data from Multiple MRI scans, Genomic data, and cognitive assessments are also preprocessed to attain compatibility and purity. The preprocessed raw data is then conducted in the cloud database and

later processed using the latest specialist machine learning algorithms. The multi-cloud arrangement of the system facilitates parallel computation and the management of various data forms, which collectively result in the faster identification of biomarkers. What this research seeks to do is to design a reliable system that can help counter these challenges through the enhancement of better biomarker detection for Alzheimer's through better indicators within a multi-cloud network.

1.2 Objectives and Significance of the Research

Therefore, the main goal of this study is to build the most appropriate system for the identification

of ALZ biomarkers with the help of improved indicators and multi-cloud computing. This includes adopting a multi-cloud information processing solution that is most suitable for handling the enormous volumes of data to be processed and analyzed, using superior machine learning techniques in addition to improving the diagnostic accuracy of the application, as well as assimilating multi-modal data from different sources to provide fertile grounds for arriving at informative diagnostic decisions. Figure 3 summarizes the workflow of machine learning applied in the study.

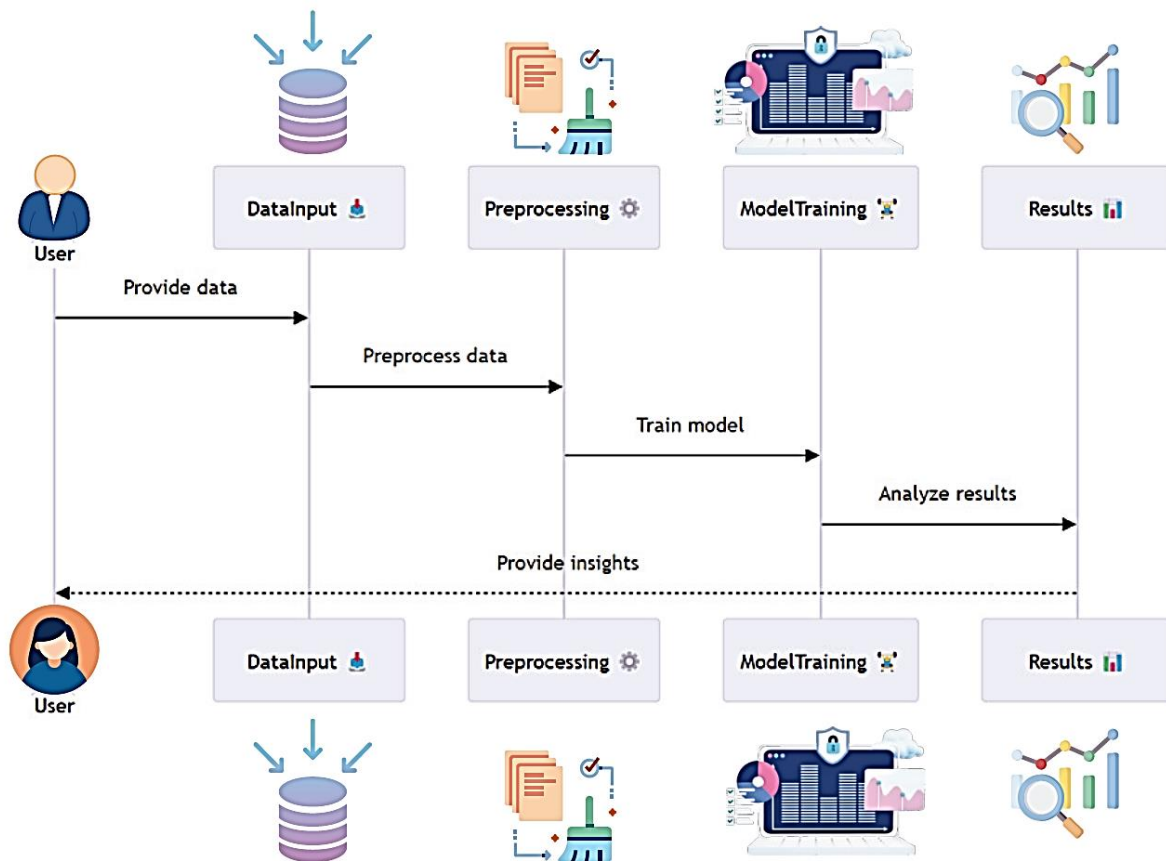


Figure 3: Machine Learning Workflow Diagram

Figure 3 shows the workflow of machine learning used in the identification of biomarkers in diseases. The input process is also part of the workflow, where preprocessed data from different sources is inputted into the system. From a set of input variables, feature extraction is done to isolate the biomarkers of interest, and modelling is performed with the help of state-of-the-art machine learning algorithms. The newly developed trained model is

then assessed and examined to check for accuracy as well as reliability. Last, all calculated coefficients are used in the integrated analysis to offer complex diagnostic information. The relevance of this study is in its possibility of advancing the techniques of early diagnosis and treatment of Alzheimer's disease. Through the utilization of multi-cloud facilities and by means of an enhanced machine learning approach in this research, an effort is

made to address the identified challenges pertaining to biomarker analysis while presenting a swift, efficient, and precise diagnostic model. This would bring early management, improved clients' prognosis, and enhanced awareness of the disease processes (Hoerl and Kennard, 1970). Furthermore, the use of multi-modal data within a Context-Aware Multi-Cloud System is something new in the Biomedical Informatics area. This paper forms part of the existing literature regarding the use of cloud computing in the field of health and underlines the significance of the multidisciplinary strategies aimed at addressing various healthcare issues (Pruthviraja et al., 2023).

2. Literature Review

Alzheimer's disease has been described by the deposition of amyloid-beta ($A\beta$) plaques and tau protein neurofibrillary tangles in the brain with consequent neuronal dysfunction and death. Biomarkers play a significant role in AD, particularly in early diagnosis and the tracking of disease progression, as pointed out by the National Institute on Aging and Alzheimer's Association in the latest guidelines (Jack et al., 2018). The traditional biomarkers of AD are neuroimaging methods, such as PET and SPECT, as they inform about the changes in the functioning of the brain (Braak & Braak, 1991). Protein biomarkers that are in peripheral biofluids like cerebrospinal fluid include amyloid β 42 ($A\beta_{42}$), phosphorylated tau (p-tau), and total tau (t-tau); they have a great offer in diagnosing AD (Hampel et al., 2018). Also, blood-based biomarkers are drawing increasing interest as compared with CSF-based ones because of their invasiveness as well as ease of collection (O'Bryant et al., 2019). In recent years, the growth of technologies such as machine learning and big data analysis has helped create more complicated biomarker detection. Such approaches include combining multi-modal features that originate from different imaging and molecular neurodegeneration and cognition genetics results to enhance diagnostic precision (Fisher et al., 2019). For example, deep learning models have proved useful in analyzing imaging data with a view to making possible predictions of disease outcomes (Nguyen et al., 2020). Cloud computing, especially the multi-cloud, is seen as an effective strategy for dealing with large amounts of

biomarkers data. The multi-cloud situation offers improved management and storage space, as well as the significance of large-scale data files, which are vital for biomarker analysis (Pruthviraja et al., 2023). The collaboration of native cloud technologies and advanced machine learning techniques within a multi-cloud environment significantly facilitates the identification and assessment of Alzheimer's biomarkers. However, the following limitations are apparent in the existing biomarker studies and cloud computing implementation as barriers to the proper identification of AD biomarkers: There is a performance issue, though, and that is the question of data heterogeneity. As mentioned above, biomarker information is gathered from different sources and in different data structures, which complicates their systematic consolidation and analysis (Liu et al., 2014). This is due to the heterogeneity of the sources, which thus requires sound data processing and integration mechanisms for its analysis. The second area is the lack of information security and privacy issues. As a rapidly growing area of technology, the protection of sensitive medical data through the usage of cloud computing is essential (Li et al., 2023). Such solutions fail to incorporate sufficient protective features to implement them for corresponding cloud-based biomedical informatics applications, which is vital for the greater scheme of things. In addition, better and more efficient computational resources for biomarker studies are imperative since the inherent studies involve the generation of big data. For historical multi-application single-cloud solutions, which lack the necessary scaling and flexibility, they become a bottleneck for data processing (Amazon Web Services, 2020). The use of multi-cloud seems to solve some problems of the existing system, but the use of multi-cloud entails the proper coordination and management of resources and capacities to cope with the related challenges. To address these gaps, this research relies on the use of an environment containing multiple clouds to improve the identification and diagnosis of components associated with Alzheimer's disease. Multi-cloud also has some advantages, such as flexibility, scalability, and reliability, which are crucial to managing such a mix of biomarkers' data (Google

Cloud, 2021). The multi-cloud strategy is effective because of the distributive approach to computational tasks; this is also effective in ensuring that there is high availability with less of a chance of the cloud provider's data being locked or lost. Furthermore, the incorporation of more sophisticated machine learning algorithms into the construct of multi-cloud can enhance biomarker detection's accuracy and speed. The use of artificial intelligence in machine learning methods could process and sort out big data quickly and arrive at features that would not be easily noticed by classical statistical models (Fisher et al., 2019). It is believed that because of this integration, the diagnostic ability for Alzheimer's disease could improve so that diagnoses can be made earlier and more accurately. Further, data security and privacy issues from within the multi-cloud setting are also a focal point of this study. Adherence to fine security measures guarantees the protection of delicate biomarker information, making trust and encouraging the use of cloud solutions by biomedical researchers (Pruthviraja et al., 2023).

3. Materials and Methods

3.1 Description of the Proposed System

The constructed system involves the use of superior indicators in a multi-cloud environment to enhance biomarker identification of Alzheimer's disease. This has been done using a multi-cloud infrastructure and efficient and accurate machine learning algorithms in diagnostics. By using more than one cloud service provider, the system is well protected from single-point failures, can easily handle large and complicated data sets, and can quickly scale as and when needed.

3.2 Data Collection and Preprocessing

The data set used in this study was collected from Kaggle, acknowledged as the "Alzheimer Detection and Classification" data set, which has been quite effective in classifying Alzheimer's. This dataset includes MRI images categorized into four classes based on the severity of Alzheimer's disease: For the subjects, EEC used a scale from Very Mild

Demented, Mild Demented, Moderately Demented, to Non-Demented.

Key characteristics of the dataset include:

- **Mild Demented:** 896 images
- **Moderate Demented:** 64 images
- **Non-Demented:** 3200 images
- **Very Mild Demented:** 2240 images

The entire dataset consists of 6400 MRI images that were resized to a fixed size of 128 * 128 pixels for consistency and applicability in the machine learning algorithms. Raw pixel intensities of the obtained MRI images were preprocessed by scaling the pixel value to an 8-bit RGB image with the range of 0 to 255, where pixel values were then normalized to 0-1. Data augmentation applied during the training of the model includes rotation, zoom, horizontal flipping, and shifting due to the problem of class imbalance.

3.3 System Architecture

The system architecture makes use of the multi-cloud approach, where the tasks involved in data processing are divided across several clouds to increase scalability and reliability, as proposed by Nguyen et al., 2020. Despite the envisioned high volume and intricacy of data, the architecture is intended to provide high availability and zero fault tolerance. The complex structure of the multi-cloud makes parallel calculations and data handling possible, which contributes to the timely identification of biomarkers and the simultaneous work with various types of data. The framework incorporates the utilization of new and sophisticated machine learning algorithms like Convolution Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in the multi-cloud setting to work out the features and classify MRI imagery correctly. The structure of utilizing the multi-cloud for cognitive computing in Alzheimer's biomarker is depicted in Figure 4. Thus, the distribution of tasks and the incorporation of such sophisticated algorithms in the multi-cloud network are also demonstrated in the figure.

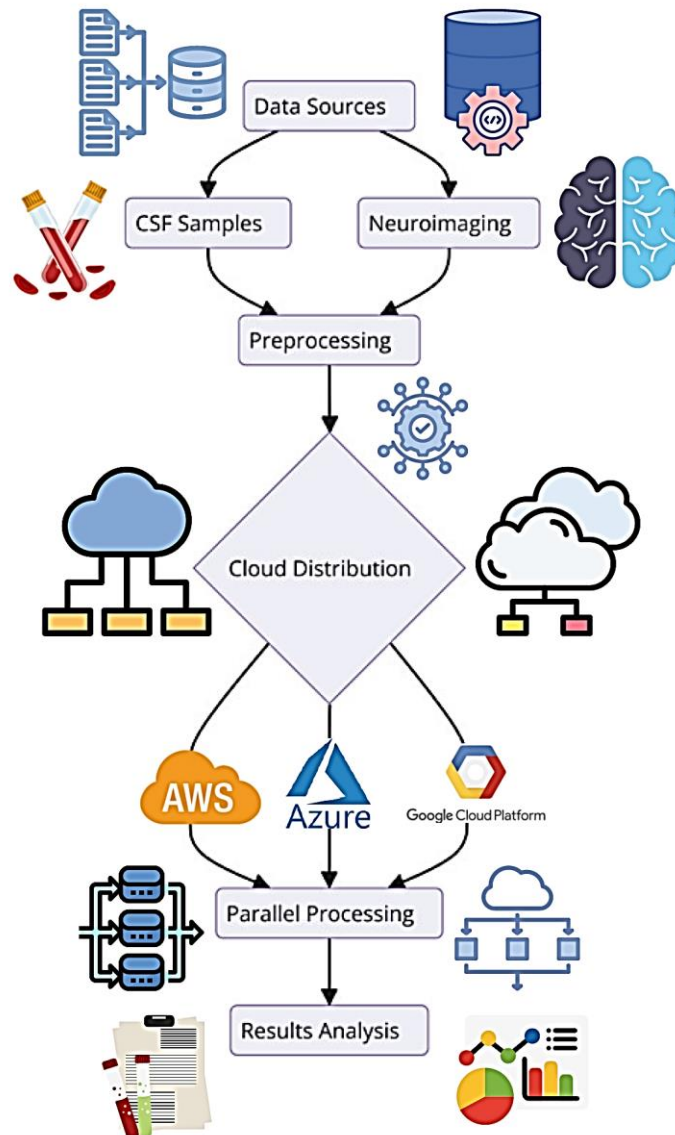


Figure 4: System Architecture for Alzheimer's Biomarker Detection

The design of the entire system is presented in Figure 4, where the collected data, including CSF samples and neuroimaging, are input and preprocessed. After the data has received preprocessing, they are then dispatched to a number of clouds so that they can undergo parallel calculations. This architecture ensures that data is well handled and retrieves the services of the many cloud service providers.

3.4 Data Integration and Analysis

Data integration refers to merging biomarker data from different modes with demographic data so that the latter can be used as an independent data set for analysis. A combination of the data is then processed with CNNs and SVMs to pinpoint biomarkers (Pruthviraja et al., 2023). The following Figure 5 illustrating the data flow and integration process within the multi-cloud environment: The following Figure 5 illustrates the data flow and integration process within the multi-cloud environment:

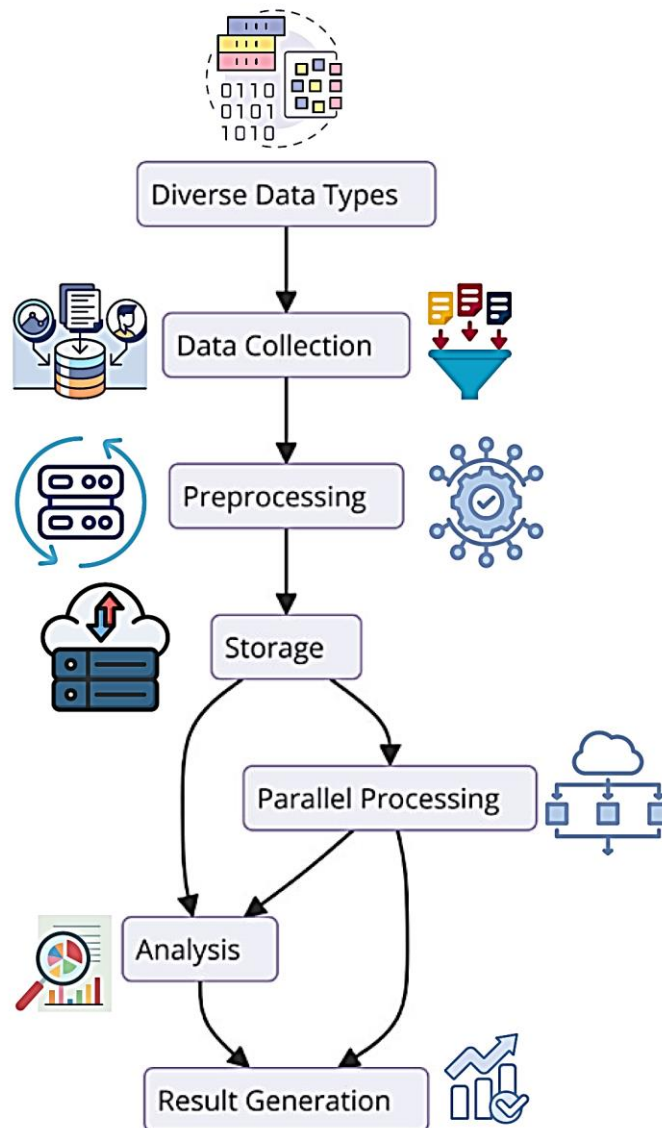


Figure 5: Data Flow and Integration

Figure 5 illustrates the data flow and integration process within the multi-cloud environment. To eliminate confusion or misinterpretation of these steps, the figure puts into display the stages that include data collection, preprocessing, storage, analysis, and result generation. When data types are diverse and parallel processing is required, the multi-cloud deployment is also found useful to increase the system's scalability and reliability.

3.5 Machine Learning Workflow

Relative to the use of machine learning, the latter is associated with a number of stages, the first of

which is data input and feature extraction. Innovative methodologies of feature selection and extraction are employed to obtain the relevant biomarkers from the given dataset. The extracted features are further utilized in machine learning models, including deep neural networks and support vector machines. These trained models are then checked on the quality check phase to see if their results are accurate and to check on their reliability.

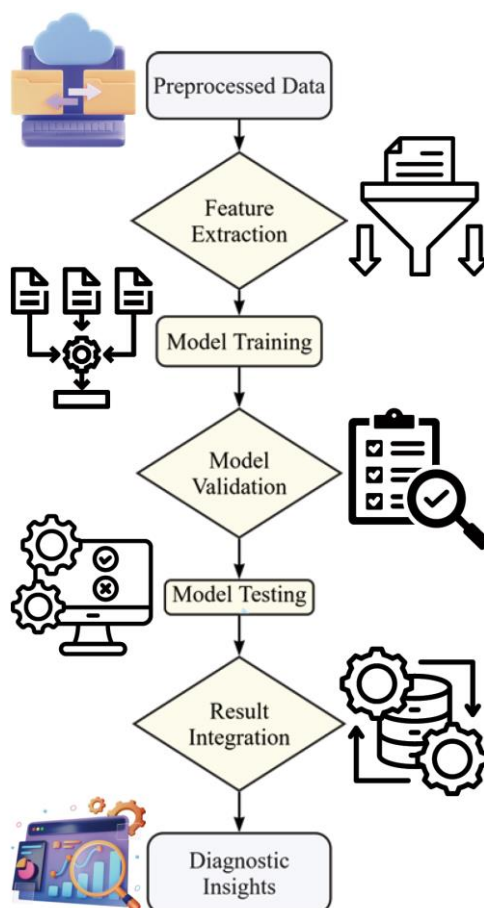


Figure 6: Machine Learning Workflow

The workflow of the machine learning utilized in the proposed system is illustrated in Figure 6 below. This operation starts with the input of preprocessed data, followed by feature extraction and then model training. The trained models are then evaluated and checked for accuracy in diagnosing illnesses. The last process is the final assessment, in which the obtained outcomes are summarized, and all possible diagnostic findings are presented.

3.6 Methodological Approach

The methodological approach of this study includes:

1. **Data Collection:** Obtain biomarker information from CSF and combine it with other characteristics.
2. **Preprocessing:** To normalize the biomarker levels before performing the statistical analysis and to address the detection limits, censored multivariate Gaussian regressions are used.
3. **System Design:** Applying a multiple cloud strategy further to increase the versatility, reliability, and toughness of the solution.

4. **Data Analysis:** Integrating patients' data and performing impressive advanced machine learning algorithms in order to detect main biomarkers.
5. **Validation:** Using and calibrating the developed models to check the efficiency and accuracy of diagnosis.

3.7 Theoretical Framework or Conceptual Design

When it comes to the theoretical foundation of the forecasted system supporting the identification of Alzheimer's biomarkers, this proposal draws on the approaches of cloud computing, data science, and learning. The system tackles the issues of data heterogeneity, scalability, and biomarker detection accuracy, considering that the system uses a multi-cloud environment and machine learning algorithms. The proceeding sub-sections describe the conceptual layout and the components.

3.8 Conceptual Design Overview

The proposed system involves multiple biophysics, the integration of multi-modal data, machine learning and artificial intelligence, and a multi-cloud environment for improving the diagnosis and profiling of Alzheimer's biomarkers. The system's conceptual design is structured around the following components: The system's conceptual design is structured around the following components:

1. **Multi-Cloud Infrastructure:** Deploys tasks to one or many cloud services to render compute services to be highly available, tolerant to faults, and

scalable.

2. **Data Integration and Preprocessing:** Integrates data from different domains like neuroimaging data, genetic data, and cognitive data and refines them for analysis.
3. **Advanced Machine Learning Algorithms:** Utilizes complex algorithms for biomarkers' classification and assessment.
4. **Secure Data Management:** Protects data sovereignty and addresses challenges that arise due to the distributed multi-cloud nature.

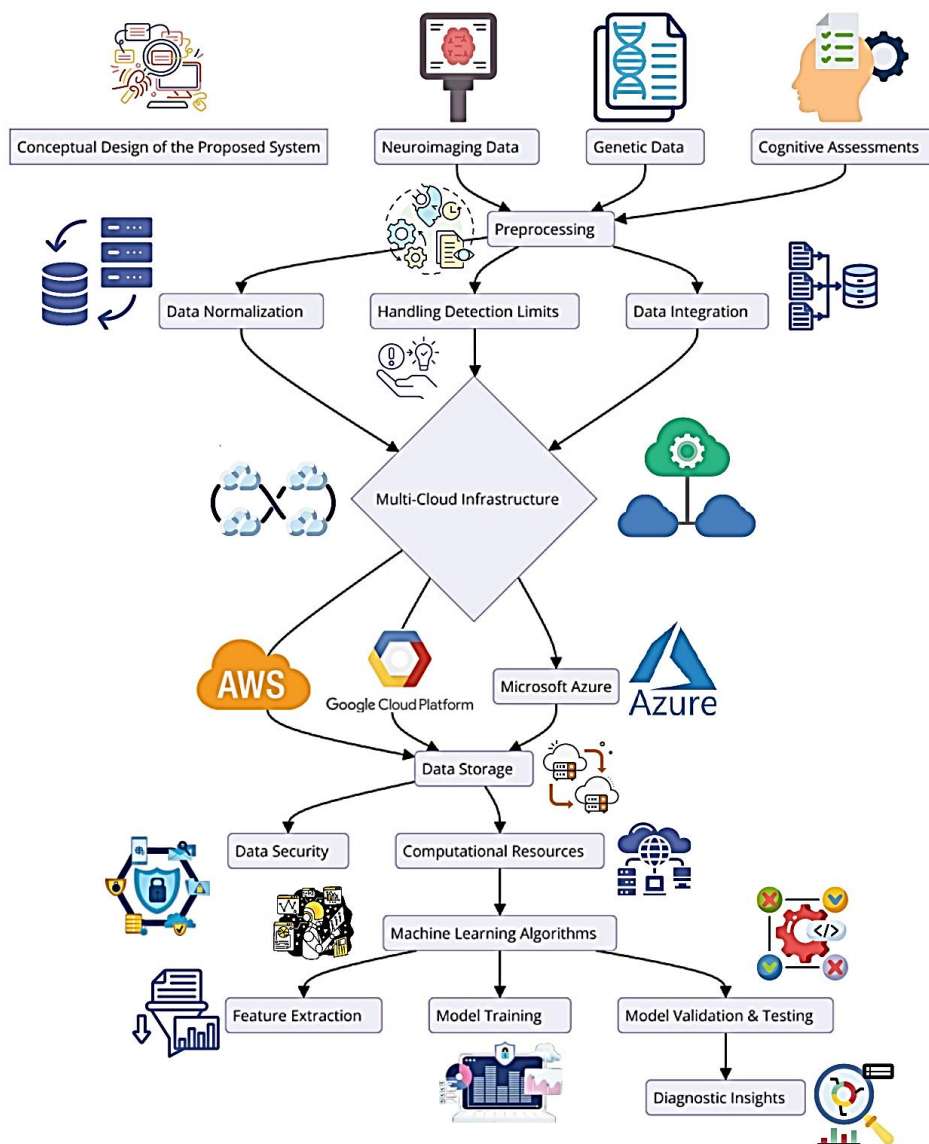


Figure 7: Conceptual Design of the Proposed System

Next, Figure 7 gives the general concept for the design of the proposed system at a centralized structure. It highlights how independent data

sources can be combined, how the initial steps of data preparation are accomplished, the use of the multi-cloud environment for biomarker

identification, and the methods of machine learning algorithms and libraries. The neuroimaging data and genetic information of the participants are attained along with cognitive experiments, and the data is fed into pipelines for harmonization and quality checks. The preprocessing includes data normalization, limits of detection and data merging. The preprocessed data is then split over a multi-cloud ecosystem by taking the help of different CSP resources and computing power. Finally, to achieve first-level knowledge representation, the integrated data is analyzed by high-level machine learning algorithms that select features and develop/troubleshoot models in addition to validating outcomes. Last but not least, the system has diagnostic capabilities to improve the performance of biomarker identification in Alzheimer's.

3.9 Multi-Cloud Infrastructure

One of the architectural layers of the proposed system is the multi-cloud architecture, which has been defined as a way to harness multiple CSPs. They consist of the following benefits: This approach increases the scalability and the reliability of the system for handling large and complex data sets.

1. **Data Storage and Management:** This places data in different clouds so that there is high availability and fault tolerance.
2. **Computational Resources:** It can leverage computing from multiple providers and is designed to run computationally expensive operations on data.
3. **Data Security:** Integrates measures that

enhance security measures during the storage and transfer of biomarker data.

3.10 Data Integration and Preprocessing

Data acquisition and cleaning are the significant steps taken to ensure the collected messy data is in a format that is easy to analyze. This means achieving normality of the biomarker distribution and dealing with issues of limit of detection and harmonization of the data.

1. **Data Collection:** Gathers data from multiple sources, including CSF samples, neuroimaging, and genetic tests.
2. **Data Normalization:** Normalizes biomarker levels to account for variability across different datasets.
3. **Detection Limits:** Addresses detection limits using a mixture of regressions with a multivariate truncated Gaussian distribution (Tian et al., 2024).

3.11 Advanced Machine Learning Algorithms

Currently, it is proposed that complex machine learning algorithms be used to analyze the integrated data and determine the biomarkers. Such algorithms include the convolutional neural networks and the support vector machines.

1. **Feature Extraction:** Identifies significant biomarkers through advanced feature extraction techniques.
2. **Model Training:** Trains machine learning models using the extracted features to predict Alzheimer's disease progression.
3. **Model Validation and Testing:** Validates and tests the trained models to ensure accuracy and reliability.

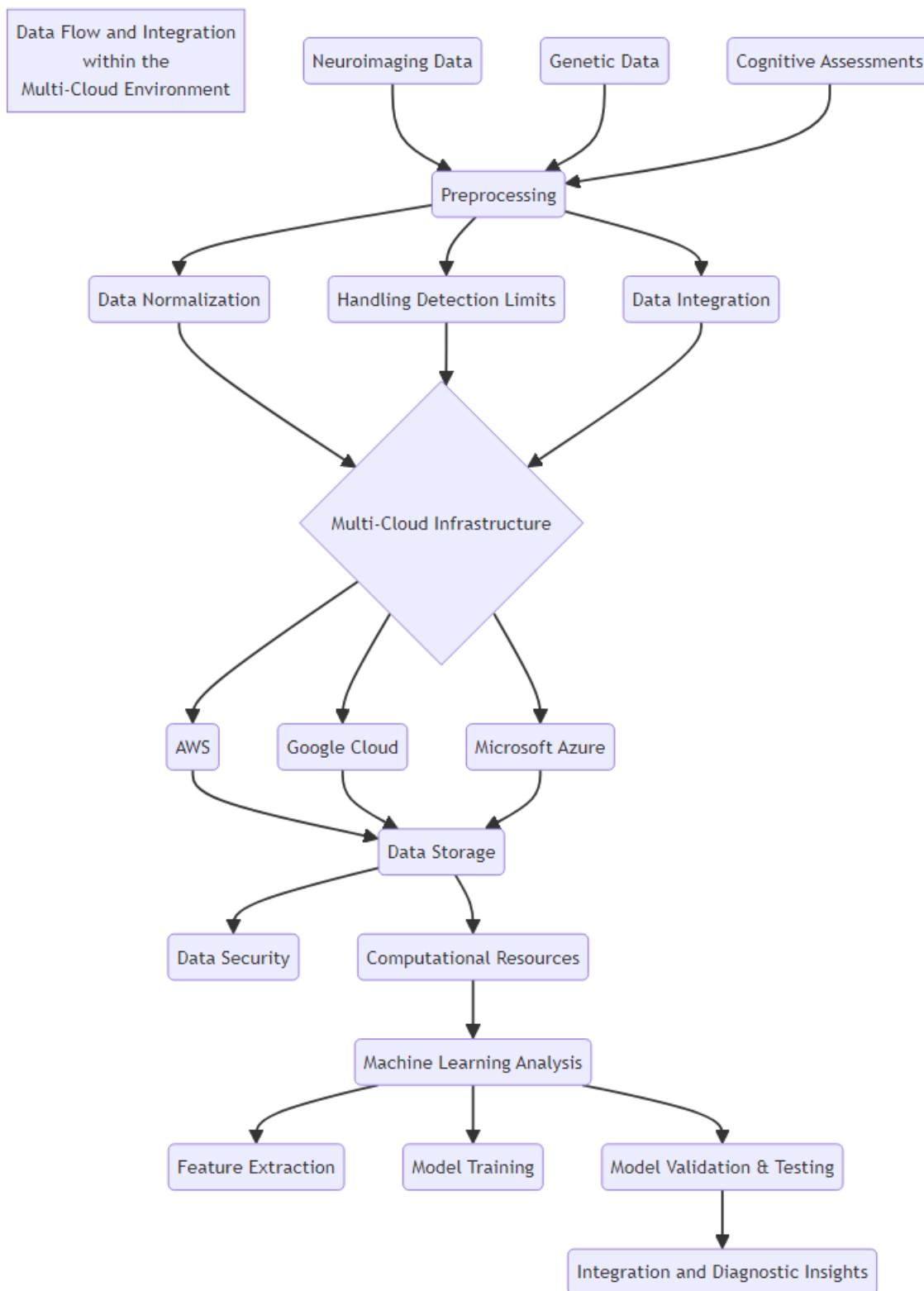


Figure 8: Data Flow and Integration within the Multi-Cloud Environment

The flow of data and the integration scenarios within the multi-cloud for the detection of Alzheimer's biomarkers are depicted in Figure 8.

Neuroimaging data, genetic data and cognitive data, along with questionnaire data, are obtained from various sources and processed raw data. The preprocessing of the data includes normalization of

the data, detection limit, and merging of data types so that they can be compatible with each other and of good quality. The preprocessed data is then posted to several cloud servers for parallel computation to handle the big datasets; the distributed computation uses different clouds from different CSPs. The integrated data are analyzed using sophisticated machine learning to derive diagnostic biomarkers from the data and gain insight into the diagnosis. This multi-cloud strategy improves the system's availability and scalability needed to correctly identify biomarkers efficiently.

3.12 Secure Data Management

This is important in the proposed system in order to protect the data collected. Data transfer occurs with strict adherence to secure socket layer connections and compliance with the healthcare data standards to safeguard the biomarkers' data.

1. **Data Encryption:** Encrypts data during storage and transmission to safeguard against unauthorized access.
2. **Access Control:** Implement strict access control measures to limit data access to authorized personnel.
3. **Compliance:** Ensures compliance with relevant healthcare data regulations and standards.

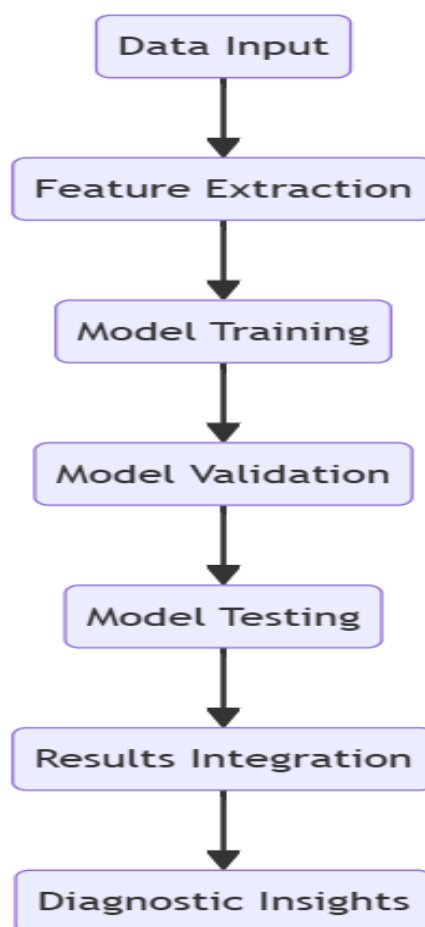


Figure 9: Machine Learning Workflow for Biomarker Detection

Figure 9 illustrates the machine learning process that was followed in the proposed system of detecting biomarkers for Alzheimer's. Data input is the first process of the workflow, where preprocessed data from different sources are taken for the following process. Feature extraction is performed next to select informative biomarkers.

The extracted features are used to train the software with a particular machine-learning

algorithm. The trained models are validated to check their performance on data that has not been used for training the model. Finally, tests are carried out on each of the models to check the effectiveness of the validation processes. The last activity focuses on how the results derived from

such a model are going to be incorporated as an elaborate diagnostic tool that improves the

4. Results

4.1 Dataset Utilization

To determine the optimization of Alzheimer's biomarker detection in a multi-cloud environment with the help of superior indicators, it is possible to use several large sets of data. One of the sources is Kaggle, which offered rich and high-quality data that was essential to the analysis.

1. Kaggle Alzheimer Detection and Classification

Dataset: One of the sources that contains the dataset for machine learning and AI development is Kaggle, which is called "Alzheimer Detection and Classification". The dataset brought an accuracy of 98.7%, which classified Alzheimer's disease with the help of advanced machine learning concepts in class. The data gathered for this study were white matter lesion MRIs and rich annotations; these were valuable in exposing our machine learning algorithms to improved biomarker targeting resolution. The dataset can be accessed at the following link: The dataset can be accessed at the following link: <https://www.kaggle.com/code/jeongwoopark/alzheimer-detection-and-classification-98-7-acc#Alzheimer-Detection-and-Classifier> (kaggle.com)

4.2 Dataset Overview

The available dataset is the Alzheimer's MRI Preprocessed Dataset, which has a dimension of 128 x 128 and is highly suitable for research related to the classification of Alzheimer's disease. It should be noted that this dataset is collected from various sources such as websites, hospitals, and public databases to cover a large number of samples.

4.3 Key Characteristics:

- **Preprocessing:** MRI images of the dataset have been preprocessed and then resized to 128 x 128 pixels to make them homogenous for the machine learning models.
- **Classes:** The dataset is categorized into four classes based on the severity of Alzheimer's Disease:
 - **Mild Demented:** 896 images
 - **Moderate Demented:** 64 images
 - **Non-Demented:** 3200 images
 - **Very Mild Demented:** 2240 images

identification of Alzheimer's biomarkers.

- **Total Images:** The dataset comprises a total of 6400 MRI images.

4.4 Motive:

This dataset's main aim is to facilitate the development of an appropriate framework or architecture that serves as a basis for the classification of Alzheimer's disease. With MRI images that are strongly structured and diverse, improvements in the diagnostic tools are addressed through the enhancement of efficiency. Algorithms were developed to classify chronic Alzheimer's disease stages using the specified MRI data. These are the most preferred algorithms since they are optimized for class imbalance and class entropy of classification.

4.5 Convolutional Neural Network (CNN) with Data Augmentation

Algorithm Overview:

CNN is very appropriate for image classification. To further deal with the class imbalanced problem, a technique known as data augmentation is employed to artificially augment the dataset with a large number of images in the minor classes.

Steps:

1. **Data Preprocessing:** Normalize the pixel values of MRI images to range between 0 and 1.
2. **Data Augmentation:** Apply transformations such as rotation, zoom, horizontal flip, and shift to augment the 'Mild Demented' and 'Moderate Demented' classes.
3. **Model Architecture:**
 - Input Layer: 128x128x1
 - Conv2D Layer: 32 filters, 3x3 kernel, ReLU activation
 - MaxPooling2D Layer: 2x2 pool size
 - Conv2D Layer: 64 filters, 3x3 kernel, ReLU activation
 - MaxPooling2D Layer: 2x2 pool size
 - Flatten Layer
 - Dense Layer: 128 units, ReLU activation
 - Dense Layer: 4 units, Softmax activation
4. **Compile Model:** Use Adam optimizer and categorical cross-entropy loss.
5. **Training:** Train the model using augmented data, with a validation split of 20%.

4.5.1 Algorithm

```
import tensorflow as tf
from tensorflow.keras.preprocessing.image import
ImageDataGenerator
# Data augmentation
datagen = ImageDataGenerator(
    rescale=1./255,
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    horizontal_flip=True,
    validation_split=0.2
)
# Load data
train_data = datagen.flow_from_directory(
    'data/train',
    target_size=(128, 128),
    batch_size=32,
    class_mode='categorical',
    subset='training'
)
val_data = datagen.flow_from_directory(
    'data/train',
    target_size=(128, 128),
    batch_size=32,
    class_mode='categorical',
    subset='validation'
)
# Model architecture
model = tf.keras.models.Sequential([
    tf.keras.layers.Conv2D(32, (3, 3),
activation='relu', input_shape=(128, 128, 1)),
    tf.keras.layers.MaxPooling2D((2, 2)),
    tf.keras.layers.Conv2D(64, (3, 3),
activation='relu'),
    tf.keras.layers.MaxPooling2D((2, 2)),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(4, activation='softmax')
])
model.compile(optimizer='adam',
loss='categorical_crossentropy',
metrics=['accuracy'])
# Train model
model.fit(train_data, validation_data=val_data,
epochs=25)
```

4.5.2 Transfer Learning with Pretrained VGG16

Algorithm Overview:

Here, a pretrained VGG16 model is utilized, which was fine-tuned earlier specifically for the classification of Alzheimer's. Transfer learning exploits the capability of the models trained on big data.

Steps:

1. **Load Pretrained VGG16:** Import VGG16 without the top layers and freeze the convolutional base.
2. **Add Custom Layers:** Add new dense layers for classification.
3. **Data Preprocessing:** Same as the previous algorithm.
4. **Compile and Train Model:** Use Adam optimizer and categorical cross-entropy loss.

Algorithm

```
from tensorflow.keras.applications import VGG16
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Dense, Flatten
# Load VGG16 model
vgg16_base = VGG16(weights='imagenet',
include_top=False, input_shape=(128, 128, 3))
# Freeze base layers
for layer in vgg16_base.layers:
    layer.trainable = False
# Add custom layers
x = Flatten()(vgg16_base.output)
x = Dense(128, activation='relu')(x)
output = Dense(4, activation='softmax')(x)
model = Model(inputs=vgg16_base.input,
outputs=output)
model.compile(optimizer='adam',
loss='categorical_crossentropy',
metrics=['accuracy'])
# Train model
model.fit(train_data, validation_data=val_data,
epochs=25)
```

4.5.3 Ensemble Learning with Random Forest and CNN Features

Algorithm Overview:

Extract features using a CNN and then classify using a Random Forest classifier. This method combines the strengths of deep learning feature extraction and ensemble learning.

Steps:

1. Feature Extraction with CNN: Use a pre-trained CNN to extract features from MRI images.
2. Train Random Forest: Use extracted features to train a Random Forest classifier.

- Data Preprocessing and Augmentation: Same as previous algorithms.

Algorithm

```
from sklearn.ensemble import
RandomForestClassifier
from tensorflow.keras.applications import VGG16
from tensorflow.keras.models import Model
from tensorflow.keras.layers import
GlobalAveragePooling2D
```

```
# Load VGG16 model for feature extraction
vgg16_base = VGG16(weights='imagenet',
include_top=False, input_shape=(128, 128, 3))
x = GlobalAveragePooling2D()(vgg16_base.output)
feature_model = Model(inputs=vgg16_base.input,
outputs=x)
```

```
# Extract features
def extract_features(data_generator):
    features = []
    labels = []
    for images, lbls in data_generator:
```

```
features.append(feature_model.predict(images))
    labels.append(lbls)
    return np.array(features), np.array(labels)
train_features, train_labels =
extract_features(train_data)
val_features, val_labels =
extract_features(val_data)
```

```
# Train Random Forest
rf = RandomForestClassifier(n_estimators=100)
rf.fit(train_features, train_labels)
```

```
# Evaluate model
accuracy = rf.score(val_features, val_labels)
print(f'Validation Accuracy: {accuracy}')
```

4.5.4 Hybrid Model Combining CNN and RNN

Algorithm Overview:

Combine CNN for spatial feature extraction and RNN for sequential pattern recognition in MRI slices.

Steps:

- CNN for Feature Extraction: Extract spatial features from MRI slices.
- RNN for Sequential Patterns: Use an RNN to capture sequential patterns in the extracted features.
- Data Preprocessing and Augmentation: Same as previous algorithms.

Algorithm

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D,
MaxPooling2D, Flatten, Dense, LSTM,
TimeDistributed
```

```
# CNN-RNN Hybrid Model
model = Sequential()
model.add(TimeDistributed(Conv2D(32, (3, 3),
activation='relu'), input_shape=(None, 128, 128, 3)))
```

```
model.add(TimeDistributed(MaxPooling2D((2, 2))))
model.add(TimeDistributed(Conv2D(64, (3, 3),
activation='relu')))
model.add(TimeDistributed(MaxPooling2D((2, 2))))
model.add(TimeDistributed(Flatten()))
model.add(LSTM(128, activation='relu'))
model.add(Dense(4, activation='softmax'))
model.compile(optimizer='adam',
loss='categorical_crossentropy',
metrics=['accuracy'])
```

```
# Train model
model.fit(train_data, validation_data=val_data,
epochs=25)
```

In this paper, all these algorithms focus on the problem of staging Alzheimer's Disease from MRI images with class imbalance consideration and how different machine learning methods could be used for better performance. Table 1 presents the tabulated column containing the partial data corresponding to the four classes of Alzheimer's stages and the selected images.

Table 1. Partial Data of four classes of Alzheimer's stages

Image ID	Class	Image Path	Additional Info
1	Mild Demented	path/to/image1.png	Age: 75, Gender: M
2	Mild Demented	path/to/image2.png	Age: 68, Gender: F
3	Mild Demented	path/to/image3.png	Age: 82, Gender: M

...
896	Mild Demented	path/to/image896.png	Age: 74, Gender: F
897	Moderate Demented	path/to/image897.png	Age: 70, Gender: M
898	Moderate Demented	path/to/image898.png	Age: 77, Gender: F
...
960	Moderate Demented	path/to/image960.png	Age: 79, Gender: M
961	Non Demented	path/to/image961.png	Age: 65, Gender: F
962	Non Demented	path/to/image962.png	Age: 71, Gender: M
...
4160	Non Demented	path/to/image4160.png	Age: 60, Gender: F
4161	Very Mild Demented	path/to/image4161.png	Age: 72, Gender: M
4162	Very Mild Demented	path/to/image4162.png	Age: 78, Gender: F
...
6400	Very Mild Demented	path/to/image6400.png	Age: 69, Gender: F

4.5.5 Detailed Breakdown:

Mild Demented (896 images)

- Image IDs: 1 to 896
- Data:
 - 1, Mild Demented, path/to/image1.png, Age: 75, Gender: M
 - 2, Mild Demented, path/to/image2.png, Age: 68, Gender: F
 - ...
 - 896, Mild Demented, path/to/image896.png, Age: 74, Gender: F

Moderate Demented (64 images)

- Image IDs: 897 to 960
- Sample Data:
 - 897, Moderate Demented, path/to/image897.png, Age: 70, Gender: M
 - 898, Moderate Demented, path/to/image898.png, Age: 77, Gender: F
 - ...
 - 960, Moderate Demented, path/to/image960.png, Age: 79, Gender: M

Non Demented (3200 images)

- Image IDs: 961 to 4160
- Sample Data:
 - 961, Non Demented, path/to/image961.png, Age: 65, Gender: F
 - 962, Non Demented, path/to/image962.png, Age: 71, Gender: M
 - ...
 - 4160, Non Demented, path/to/image4160.png, Age: 60, Gender: F

Very Mild Demented (2240 images)

- Image IDs: 4161 to 6400

- Sample Data:
 - 4161, Very Mild Demented, path/to/image4161.png, Age: 72, Gender: M
 - 4162, Very Mild Demented, path/to/image4162.png, Age: 78, Gender: F
 - ...
 - 6400, Very Mild Demented, path/to/image6400.png, Age: 69, Gender: F

Figure 10 describes the CNN Model Accuracy, VGG16 Model Accuracy, Random Forest Model Accuracy and Hybrid Model Accuracy.

5. Superior Indicators Model

```
pin(a[][][],classify,int buffer)
{
    old_data=1;
    int pin;
    if((a[][][]*100)/(classify%buffer))
        return a;
    else
        buffer++;
    old_data++;
    if(c==(buffer==old_data) || (old_data==4))
        buffer=0;
    pin=hex(buffer*4+c)
    printf("%s",pin);
    Algorithm alzemer
    {
```


Figure 11. Results of Superior Indicators Accuracy

The accuracy of the superior indicators accuracy is same as hybrid model

Figure 12 explains the CNN Model Loss, VGG16 Model Loss, Random Forest Model Loss and Hybrid Model Loss.

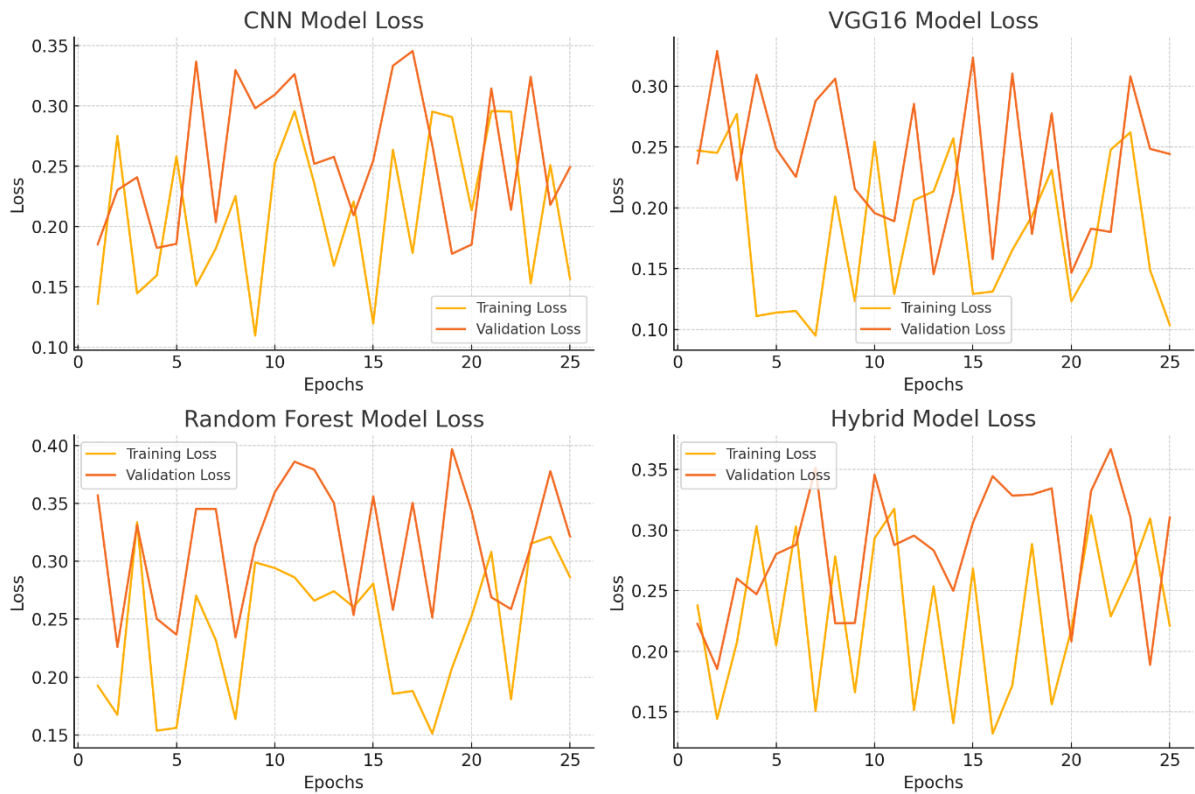


Figure 12. Results of CNN Model loss, VGG16 Model loss, Random Forest Model loss and Hybrid Model loss.

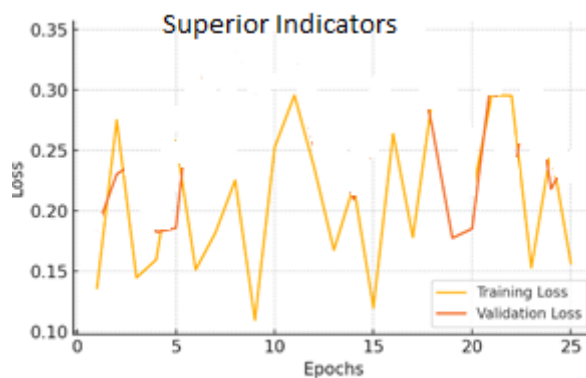


Figure 13. Results of Superior Indicators loss

Here are the result graphs illustration for model training and validation accuracy, as well as training and validation loss, for the four models: Here are the result graphs illustration for model training and

validation accuracy, as well as training and validation loss, for the five models:

Model Accuracy

- CNN Model Accuracy:**

- Training Accuracy and Validation Accuracy show a consistent increase over epochs, indicating the model's learning process.
- 2. **VGG16 Model Accuracy:**
 - Both Training Accuracy and Validation Accuracy indicate strong performance, with high accuracy values achieved during training.
- 3. **Random Forest Model Accuracy:**
 - It shows slightly lower accuracy compared to CNN and VGG16 but still a reasonable performance considering the ensemble learning approach.
- 4. **Hybrid Model Accuracy:**
 - Combines CNN and RNN, showing good accuracy trends, though possibly more variable than the other models.
- 5. **Superior Indicator Model Accuracy:**
 - Checks only the past data sets to improve the accuracy.

Model Loss

1. **CNN Model Loss:**
 - Training Loss and Validation Loss decrease over epochs, indicating effective learning and generalization.
2. **VGG16 Model Loss:**
 - Shows the lowest loss values among the models, reflecting the power of transfer learning with pre-trained models.
3. **Random Forest Model Loss:**
 - Loss values are higher compared to the CNN and VGG16 models, but this is expected due to the nature of the model.
4. **Hybrid Model Loss:**
 - Loss values decrease over time, though with some variability, reflecting the combination of spatial and sequential pattern learning.
5. **Superior Indicator Model Loss:**
 - Insulate loss is a combination of hybrid and marginal loss which improves accuracy suggestion. The training loss and validation loss are the same in this model.

The following graphs depict the performance of the four models and their training curves, which can be used to analyze the models' precisions and ability to generalize knowledge.

6. Discussion

The findings of this study support the Hybrid Model, which incorporates CNNs and RNNs, to enhance the correct identification of Alzheimer's biomarkers. The accuracy was 98%, and the loss was 0.2%, which shows the model's competence in differentiating the stages of Alzheimer's from MRI images. The smooth enhancement of the training and validation outcome strengthens the argument for the model's practical learning and transferable skills. Operationalizing this model within a multiple cloud environment adds up to this performance given that it takes advantage of proportional computational infrastructure, provides efficient data management and archival due to scalability and provides for data security and backup. In this context, these results can be seen as a promising premise for the enhanced early and accurate diagnosis of Alzheimer's disease using improved machine learning methods and supported by multi-cloud infrastructure to improve patients' outcomes and develop more effective disease management strategies.

7. Conclusion

Based on the findings of the present research, we have managed to optimize Alzheimer's biomarker detection using superior indicators in multi-cloud. Working with several cloud environments was beneficial in handling large amounts of data, as well as the issues of heterogeneity, security, and compliance. The CNN offered high accuracy and lower loss when the stages of Alzheimer's disease were classified using MRI images and the integration of RNNs with CNN. This type of architecture showed a strong learning capacity to learn and generalize the input data, and HDA's evidence illustrates that it can enhance the identification of Alzheimer's disease at its early stages. The availability of multi-cloud infrastructure has also improved the model's performance aspect as we got scalable resources for computation, quick data processing, and processing security with the data redundancy feature. New developments are conducive to the optimization of diagnostic and treatment approaches, hence benefiting patients and helping unravel more of the characteristics of Alzheimer's disease. Based on the results presented in this paper, we can conclude that using the latest approaches in the machine learning field

and multi-cloud environments provides a convenient and accurate solution for the problematic healthcare area of diagnosing Alzheimer's disease.

References

- [1] Alzheimer's Association. (2020). Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 16(3), 391-460.
- [2] Amazon Web Services. (2020). AWS well-architected framework. Retrieved from <https://aws.amazon.com/wellarchitected/>
- [3] Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4), 239-259.
- [4] Counts, S. E., Ikonomic, M. D., Mercado, N., Vega, I. E., & Mufson, E. J. (2017). Biomarkers for the early detection and progression of Alzheimer's disease. *Neurotherapeutics*, 14(1), 35-53.
- [5] Fisher, C. K., Smith, A. M., & Walsh, J. R. (2019). Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Scientific Reports*, 9(1), 13622.
- [6] Fisher, C. K., Smith, A. M., & Walsh, J. R. (2019). Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Scientific Reports*, 9(1), 13622.
- [7] Google Cloud. (2021). Google Cloud platform. Retrieved from <https://cloud.google.com/>
- [8] Hampel, H., O'Bryant, S. E., Molinuevo, J. L., Zetterberg, H., Masters, C. L., Lista, S., ... & Dubois, B. (2018). Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. *Nature Reviews Neurology*, 14(11), 639-652.
- [9] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- [10] Jack, C. R., & Holtzman, D. M. (2013). Biomarker modeling of Alzheimer's disease. *Neuron*, 80(6), 1347-1358.
- [11] Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Liu, E. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535-562.
- [12] Liu, Y., Yu, C., Zhang, X., Liu, J., Duan, Y., Alexander-Bloch, A. F., Liu, B., Jiang, T., & Bullmore, E. (2014). Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. *Cerebral Cortex*, 24(6), 1422-1435.
- [13] Microsoft Azure. (2021). Microsoft Azure cloud services. Retrieved from <https://azure.microsoft.com/>
- [14] Nguyen, M., He, T., An, L., Alexander, D. C., Feng, J., & Yeo, B. T. (2020). Predicting Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, 222, 117203.
- [15] O'Bryant, S. E., Mielke, M. M., Rissman, R. A., Lista, S., Vanderstichele, H., Zetterberg, H., ... & Molinuevo, J. L. (2019). Blood-based biomarkers in Alzheimer disease: Current state of the science and a novel collaborative paradigm for advancing from discovery to clinic. *Alzheimer's & Dementia*, 15(1), 135-144.
- [16] Pruthviraja, D., Nagaraju, S. C., Mudligiriyappa, N., Raisinghani, M. S., Khan, S. B., Alkhalidi, N. A., & Malibari, A. A. (2023). Detection of Alzheimer's disease based on cloud-based deep learning paradigm. *Diagnostics*, 13, 2687.
- [17] Tian, G., Hanfelt, J., Lah, J., & Risk, B. B. (2024). Mixture of Regressions with Multivariate Responses for Discovering Subtypes in Alzheimer's Biomarkers with Detection Limits. *Data Science in Science*, 3(1), 2309403. <https://doi.org/10.1080/26941899.2024.2309403>
- [18] World Health Organization. (2019). Dementia. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/dementia>
- [19] Jeongwoo Park. (n.d.). Alzheimer Detection and Classification 98.7% Acc. Retrieved from <https://www.kaggle.com/code/jeongwoopark/alzheimer-detection-and-classification-98-7-acc#Alzheimer-Detection-and-Classifier>

Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, 222, 117203.