

# The Online Datasets Used to Classify the Different Stages for the Early Diagnosis of Alzheimer's Disease (AD)

Sandeep C. S, Sukesh Kumar A, Susanth M. J.

**Abstract:** Alzheimer Disease (AD) is one of the common forms of dementia which is an irreversible neurodegenerative progressive disorder of the brain which affects the elderly population above the age of 65. Alzheimer is a brain disease that causes problems with memory, thinking and behavior. It is severe enough to interfere with daily activities. Alzheimer symptoms are characterized by memory loss that affects day-to-day function, difficulty performing familiar tasks, problems with language, disorientation of time and place, poor or decreased judgment, problems with abstract thinking, misplacing things, changes in mood and behavior, changes in personality and loss of initiative. There are different types of tests associated with AD such as neuropsychological tests, laboratory tests and various imaging modalities for the early diagnosis of AD. Although these tests are available, they are inadequate for the definite diagnosis of the disease. In this paper we focus on the databases related to AD such as ADNI (Alzheimer Disease Neuroimaging Initiative), OASIS (Open Access Series of Imaging studies), Alz Gene, AD&FTDMDB (The Alzheimer Disease & Frontotemporal Dementia Mutation Database), (CAMD) Alzheimer's disease Database and NAAC (National Alzheimer's Coordinating Center), TREAD (Trajectory-Related Early Alzheimer's Database), Coalition Against Major Diseases use of the soft computing techniques and image analysis from the different imaging modalities in an efficient way for making a definite diagnosis and early prediction of AD. Our aim is to predict the early diagnosis in a reliable manner such that to combine the values of different tests with the help of soft computing techniques to develop software tool for a definite diagnosis.

**Index Terms:** Alzheimer Disease, Dementia, ADNI, OASIS, AlzGene, AD&FTDMDB, TREAD, NAAC. Soft Computing techniques, image analysis.

## I. INTRODUCTION

Alzheimer's disease (AD) is an age related neuronal deterioration and disability of the brain which leads to cognitive decline and impairs the ability to perform routine familiar functions [1,2,4]. Alzheimer's disease was discovered in 1906 by Alois Alzheimer, a German neurologist and psychiatrist [2,3]. In 2001, eleven million people suffered from Alzheimer's disease worldwide. At present nearly 36 (35.6) million people are believed to be

living with Alzheimer's disease or other dementias, increasing to nearly 66 (65.7) million by 2030 and more than 115 (115.4) million by 2050[5,6]. The number of people with dementia will double by 2030, and more than triple by 2050 [7]. AD is often confused with normal aging and dementia. Dementia is defined as the significant loss of cognitive impairment severe enough to interfere with social functioning [8]. It can result from different diseases that cause damage to human brain cells. There are various types of dementia; each has its own cause and symptoms [9]. For an example, vascular dementia is caused by lower blood flow to a part of the brain, as caused by a disease stroke. Dementia may also present in Parkinson's disease patients and hydrocephalus patients. AD is the most common senile form of dementia, caused by the deposition of beta amyloid plaques in the brain.

There are three main stages of the disease, namely mild, moderate and severe. By identifying the current stage of the disease, physicians can predict what symptoms can be expected in the future and possible courses of treatment. The mild stage, which usually lasts 2 to 4 years, is often when the disease is first diagnosed. In this stage, family and friends may begin to realize that there has been degradation in the patient's mental ability. The moderate stage lasts up to 10 years and is the longest stage of the disease. Patients often experience increased difficulty with memory and may need help with activities of daily living [10]. In the severe stage of the disease, cognitive capacity continues to decline and physical ability is severely impacted. This stage can last up to 3 years. Due to the family's decreasing ability to care for the patient, this stage often results in nursing home or other long term care facility placement. Common symptoms appearing in this stage include [10]

The most well-known neuropathological hallmarks of AD are extra neuronal senile plaques and intraneuronal neurofibrillary tangles (NFTs). Neurofibrillary tangles are filamentous bundles in cytoplasm of the neurons displacing or encompassing nucleus. In the pyramidal cells, they appear as 'flame' while in rounder cells they appear as 'globos tangles' [11]. Senile (neuritic) plaques present outside the neuron, appear as spherical bodies bearing dilated and tortuous neuritic processes around an amyloid beta core which contains some abnormal proteins like amyloid beta plaques which are derived through the processing of Amyloid Precursor Protein (APP) [12,13]. The aggregates of amyloid beta obtained from processing of APP are difficult to degrade which consequently activate inflammatory cascade that lead to oxidative injury and alterations in phosphorylation [14].

Manuscript published on 30 April 2017.

\* Correspondence Author (s)

**Sandeep C. S.**, Research Scholar, Faculty of Engineering, University of Kerala, Trivandrum, India, E-mail: [sandeepcs07nta@gmail.com](mailto:sandeepcs07nta@gmail.com)

**Sukesh Kumar A.**, Research Guide, Faculty of Engineering, University of Kerala, Trivandrum, India, E-mail: [drsukeshkumar@yahoo.in](mailto:drsukeshkumar@yahoo.in)

**Susanth M J.**, Consultant Neurologist, SUT Hospital, Trivandrum, India, E-mail: [susanth2005@gmail.com](mailto:susanth2005@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Familial causes or genetic mutations involved in disease pathology include mutations on chromosomes 21, 14 and 1. Risk factors for AD are advanced age, lower intelligence, small head size, history of head trauma and female gender [15]. No single test can determine whether a person has Alzheimer's disease. A diagnosis is made by determining the Presence of certain symptoms and ruling out other causes of dementia. This involves a careful medical evaluation, including a thorough medical history, mental status testing, a physical and neurological exam, blood tests and brain imaging exams, including: CT imaging of the head: Computed tomography (CT) scanning combines special x-ray equipment with sophisticated computers to produce multiple images or pictures of the inside of the body. Physicians use a CT of the brain to look for and rule out other causes of dementia, such as a brain tumor, subdural hematoma or stroke. MRI of the head: Magnetic resonance imaging (MRI) uses a powerful magnetic field, radio frequency pulses and a computer to produce detailed pictures of organs, soft tissues, bone and virtually all other internal body structures. MRI can detect brain abnormalities associated with mild cognitive impairment (MCI) and can be used to predict which patients with MCI may eventually develop Alzheimer's disease. In the early stages of Alzheimer's disease, an MRI scan of the brain may be normal. In later stages, MRI may show a decrease in the size of different areas of the brain (mainly affecting the temporal and parietal lobes). PET and PET/CT of the head: A positron emission tomography (PET) scan is a diagnostic examination that uses small amounts of radioactive material (called a radiotracer) to diagnose and determine the severity of a variety of diseases. A combined PET/CT exam fuses images from a PET and CT scan together to provide detail on both the anatomy (from the CT scan) and function (from the PET scan) of organs and tissues. A PET/CT scan can help differentiate Alzheimer's disease from other types of dementia. Another nuclear medicine test called a single-photon emission computed tomography (SPECT) scan is also used for this purpose [16]. In addition to the above imaging modalities there are biomarkers for the detection of AD. The biomarkers include amyloid beta, csf-tau, plasma biomarkers, inflammatory markers, transport binding proteins, neuronal markers, genetic markers etc [17]. All of these tests help to show the memory recall of a patient and the possible areas where the patient may be deficient. Using these tests can be helpful to determine the types of treatment plans that can be used, however neuropsychological tests alone are not helpful in detecting early AD. Trials were then conducted combining neuropsychological tests with clinical tests and various imaging modalities. For an effective and early diagnosis of AD, a population based study is necessary and required, which gives an idea about the various tests involved in determining AD. In this paper we use different databases such as ADNI, OASIS, AlzGene, AD&FTDMDB, TREAD, (CAMD) Alzheimer's disease Database and NAAC for determining the sensitive disease progression changes in the affecting areas related to Alzheimer's Disease with the help of clinical trials and implementing the sensitive changes with the help of image analysis and soft computing techniques. This research review is the extension of the previous work cited in the reference section [1, 2, 3, 4]

## II. ALZHEIMER DISEASE NEUROIMAGING INITIATIVE (ADNI)

The Alzheimer's disease Neuroimaging Initiative (ADNI) unites researchers with study data as they work to define the progression of Alzheimer's disease. ADNI researchers aim is to collect, validate and utilize data such as MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors for the disease. Data from the North American ADNI's participants include Alzheimer's disease patients, mild cognitive impairment subjects and elderly controls. Alzheimer's disease (AD) affects almost 50% of those over the age of 85 and is the sixth leading cause of death in the US. Since 2005, the longitudinal Alzheimer's Disease Neuroimaging Initiative has been validating the use of biomarkers including blood tests, tests of cerebrospinal fluid, and MRI/ PET imaging for Alzheimer's disease (AD) clinical trials and diagnosis. ADNI also maintains an unprecedented data access policy intended to encourage new investigation and to increase the pace of discovery in the race to prevent, treat, and one day cure AD. All data is made available without embargo. Armed with better knowledge of the first indications of AD from ADNI and other studies, researchers are beginning to test potential therapies at the earliest stages feasible when there may be the greatest promise for slowing down progression of this devastating disease. ADNI is a global research effort that actively supports the investigation and development of treatments that slow or stop the progression of AD. This multisite, longitudinal study assesses clinical, imaging, genetic and biospecimen biomarkers through the process of normal aging to early mild cognitive impairment (EMCI), to late mild cognitive impairment (LMCI), to dementia or AD. With established, standardized methods for imaging and biomarker collection and analysis, ADNI facilitates a way for scientists to conduct cohesive research and share compatible data with other researchers around the world. The ADNI study has three phases: ADNI1, ADNI GO and ADNI2. New participants were recruited across North America during each phase of the study and agreed to complete a variety of imaging and clinical assessments. Participants are followed and reassessed over time to track the pathology of the disease as it progresses.

### A. ADNI or Adni1

This is a non-randomized natural history non-treatment study in which a total of 800 subjects including 200 normal controls, 400 individuals with MCI, and 200 subjects with mild AD will be recruited at approximately 50 sites in the United States and Canada for longitudinal follow up. The major goals of the ADNI are to develop improved methods which will lead to uniform standards for acquiring longitudinal, multi-site MRI and PET data on patients with Alzheimer's disease (AD), mild cognitive impairment (MCI), and elderly controls, acquire a generally accessible data repository which describes longitudinal changes in brain structure and metabolism.

In parallel, acquire clinical cognitive and biomarker data for validation of imaging surrogates; develop methods which will provide maximum power to determine treatment effects in trials involving these patients, test a series of hypotheses based on the clinical and biomarker data as outlined in the statistical analysis section. The enrolled subjects will be between 55-90 (inclusive) years of age, have a study partner able to provide an independent evaluation of functioning, and will speak either English or Spanish. All subjects must be willing and able to undergo all test procedures including neuroimaging and agree to longitudinal follow up. Between twenty and fifty percent must be willing to undergo two lumbar punctures. Specific psychoactive medications will be excluded [18].

### B. ADNI-GO

This is a non-randomized natural history non-treatment study in which 200 newly enrolled subjects from approximately 50 sites from the United States and Canada and approximately 450-500 subjects will be followed from the original ADNI study. The major goals of ADNI-GO are to, define and characterize the stage of the AD spectrum that precedes MCI as currently enrolled in ADNI1 (late MCI or LMCI) by enrolling 200 subjects in the mildest symptomatic phase of AD, early amnesic MCI (EMCI, defined more specifically below); perform F18 amyloid imaging (18F-AV-45) on the CN and LMCI subjects from ADNI1 (including those who had C-11 PIB) and the newly enrolled EMCI subjects. FDG PET will also be performed in association with F18 amyloid imaging. This establishes a national network for F18 amyloid imaging, and will test hypotheses concerning the prevalence and severity of brain amyloid accumulation and its relationship to current and previous changes of clinical state, MRI, FDG-PET, CSF and plasma biomarkers from ADNI1; define and characterize the stage of the AD spectrum that precedes MCI as currently enrolled in ADNI1 (late MCI or LMCI) by enrolling 200 subjects in the mildest symptomatic phase of AD, early amnesic MCI (EMCI, defined more specifically below); perform F18 amyloid imaging (18F-AV-45) on the CN and LMCI subjects from ADNI1 (including those who had C-11 PIB) and the newly enrolled EMCI subjects. FDG PET will also be performed in association with F18 amyloid imaging. This establishes a national network for F18 amyloid imaging, and will test hypotheses concerning the prevalence and severity of brain amyloid accumulation and its relationship to current and previous changes of clinical state, MRI, FDG-PET, CSF and plasma biomarkers from ADNI1. The newly enrolled subjects will be between 55-90 (inclusive) years of age, have a study partner able to provide an independent evaluation of functioning, and will speak either English or Spanish. All subjects must be willing and able to undergo all test procedures including neuroimaging and agree to longitudinal follow up. 100% of the newly recruited EMCI subjects must be willing to undergo at least one lumbar puncture at baseline. Specific psychoactive medications will be excluded [18].

### C. ADNI-GO

This is a non-randomized natural history non-treatment study where 550 newly enrolled subjects (150 CN, 100 EMCI, 150 LMCI, 150 mild AD) from approximately 55 sites from the United States and Canada, approximately 450-500 CN and LMCI subjects will be followed from the original ADNI

study and approximately 200 EMCI subjects will be followed from the ADNI-GO study. The major goals of ADNI2 are to determine the relationships among clinical, imaging, genetic, and biochemical biomarker characteristics of the entire spectrum of Alzheimer's Disease (AD), as the pathology evolves from normal aging through very mild symptoms, to mild cognitive impairment (MCI), to dementia; inform the neuroscience of AD, identify diagnostic and prognostic markers, identify outcome measures that can be used in clinical trials, and help develop the most effective clinical trial scenarios; develop improved methods which will lead to uniform standards for acquiring longitudinal multi-site MRI and PET data on patients with AD, MCI, and elderly controls; perform longitudinal clinical, cognitive, MRI, PET (18F-AV-45 and FDG), and blood and CSF biomarker studies on 550 newly enrolled subjects in four diagnostic categories – cognitively normal (CN), early MCI (EMCI), late MCI (LMCI), and mild AD. Continue these longitudinal studies for approximately 500 LMCI and Cognitively Normal subjects from ADNI1 and approximately 200 EMCI subjects from ADNI-GO for an additional 5 years. The newly enrolled subjects will be between 55-90 (inclusive) years of age, have a reliable study partner able to provide an independent evaluation of functioning, and will speak either English or Spanish. All subjects will have clinical/cognitive assessments, biomarker and genetic sample collection, and imaging. A reduced battery of tests is allowable if the subject is not able/willing to complete the full battery after the participant's original Baseline Visit. All MRI and PET scans will be rapidly assessed for quality so that subjects may be rescanned if necessary. All clinical data will be collected, monitored, and stored by the Coordinating Center at University California San Diego. University of Pennsylvania will collect biomarker samples and NCRAD will collect genetic samples [18].

## III. OPEN ACCESS SERIES OF IMAGING STUDIES (OASIS)

The Open Access Series of Imaging Studies (OASIS) is a project aimed at making MRI data sets of the brain freely available to the scientific community. By compiling and freely distributing MRI data sets, we hope to facilitate future discoveries in basic and clinical neuroscience. OASIS is made available by the Washington University Alzheimer's Disease Research Center, Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN). The dataset can be classified into two: First one is the Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults which consists of a cross-sectional collection of 416 subjects aged 18 to 96. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. 100 of the included subjects over the age of 60 have been clinically diagnosed with very mild to moderate Alzheimer's disease (AD).

Additionally, a reliability data set is included containing 20 nondemented subjects imaged on a subsequent visit within 90 days of their initial session. Second one is based on the Longitudinal MRI Data in Nondemented and Demented Older Adults which consists of a collection of 150 subjects aged 60 to 96. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. 72 of the subjects were characterized as nondemented throughout the study. 64 of the included subjects were characterized as demented at the time of their initial visits and remained so for subsequent scans, including 51 individuals with mild to moderate Alzheimer's disease. Another 14 subjects were characterized as no demented at the time of their initial visit and were subsequently characterized as demented at a later visit. For each subject, a number of images are provided. The Demographics data includes Gender (M/F), Handedness (Hand), Age, Education (Educ), socioeconomic status (SES)[18]. Education codes correspond to the following levels of education: 1: less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. The Clinical data includes Mini-Mental State Examination (MMSE)[18], Clinical Dementia Rating (CDR; 0= nondemented; 0.5 – very mild dementia; 1 = mild dementia; 2 = moderate dementia) [19]. All participants with dementia (CDR >0) were diagnosed with probable AD. The derived anatomic volumes data includes estimated total intracranial volume (eTIV) (mm<sup>3</sup>), Atlas scaling factor (ASF) [16], Normalized whole brain volume (nWBV)[19].

#### IV. MINIMAL INTERVAL RESONANCE IMAGING IN ALZHEIMER'S DISEASE (MIRIAD)

The MIRIAD dataset is a database of volumetric MRI brain-scans of Alzheimer's sufferers and healthy elderly people. Many scans were collected of each participant at intervals from 2 weeks to 2 years, the study was designed to investigate the feasibility of using MRI as an outcome measure for clinical trials of Alzheimer's treatments. Includes finding the minimal interval over which trials would need to be conducted, assess whether combining more than two scanning time points would increase statistical power and, if so, the optimal combination and timing of scans for trials of varying lengths, provide a means of assessing the reproducibility of techniques within a single day and over short intervals where changes in individual's hydration and scanner fluctuations, but not pathological atrophy, might be expected. These scans, together with demographic and psychological data, are now publicly available in anonymised form to aid researchers in developing new techniques for the analysis of serially acquired MRI. All subjects were requested to attend seven imaging visits at 0, 2, 6, 14, 26, 38 and 52 weeks from baseline. 39 subjects who completed all these visits during the study attended a further scan at 18 months, 22 of these had a further scan at 24 months. At 0, 6 and 38 weeks two back-to-back scans were conducted. All scans were conducted on the same 1.5 T Signa MRI scanner (GE Medical systems, Milwaukee, WI) and acquired by the same radiographer. Three-dimensional T1-weighted images were acquired with an IR-FSPGR (inversion recovery prepared fast spoiled gradient recalled) sequence, field of view 24 cm,

256 × 256 matrix, 124 1.5 mm coronal partitions, TR 15 ms, TE 5.4 ms, flip angle 15°, TI 650 ms [20].

#### V. NATIONAL ALZHEIMER'S COORDINATING CENTER (NACC)

NACC serves as a repository for data collected at approximately 29 Alzheimer's Disease Centers (ADCs) throughout the United States. The ADCs conduct clinical and biomedical research on Alzheimer's disease and related disorders. Centers enroll their study subjects in various ways, including referral from clinicians, self-referral by patients themselves or concerned family members, active recruitment through community organizations, and volunteers who wish to contribute to research on various types of dementia. Most centers also enrol volunteer control subjects. Study subjects at each center are best regarded as a case series, not necessarily representing all cases of disease in a defined population [21].

#### VI. TRAJECTORY-RELATED EARLY ALZHEIMER'S DATABASE (TREAD)

The overall aim of this research is to help detect the subtle declines in memory that occur in people with very early Alzheimer's disease (AD) in order to find suitable volunteers for trials of promising treatments. We will establish a database for recruitment into new and promising treatment trials. Alzheimer's disease is the most common cause of dementia. AD is increasingly common as our population ages, and a hallmark is early decline in memory, which can be detected 5-10 years prior to severe memory problems using computerized memory testing. This research will use computer tests of memory and thinking developed in Melbourne that can be done over the internet and repeated periodically (e.g. monthly for 6 months, and then 3 monthly) to see if a person's memory is getting worse compared to their own baseline performance. This is a very sensitive way of detecting decline in memory and to identify people who should be offered further evaluation for possible causes, one of which is Alzheimer's disease. If evaluation suggests AD as the cause, we aim to offer participation in separate trials of promising new treatments. Participants in this study should have ready access to a computer that is connected to the internet, and be sufficiently skilful with that computer to be able to open a web browser and then follow the simple instructions of the test themselves. They should also live or be able to visit clinics in the Greater Melbourne area since they will be offered evaluation at a Melbourne clinic if decline in their memory is detected. The study will use computer tests including one developed by the Melbourne-based company Cog State Ltd. The CogState test uses playing-cards and simple keyboard responses to assess speed and accuracy, which are direct measures of a person's ability to think clearly and quickly. It is designed to be easy to use and brief, taking about 15-20 minutes at most to complete all tasks. There is a practice test to familiarize you with what's required. The practice test does not keep score. Then there is a scored test which would be used as the baseline for future comparisons.

The test is started by clicking a web page link, runs itself interactively and when finished the results are sent to the study's secure database and analysed automatically. If you agree to participate in this research, you will be asked to do the computer testing every 1-3 months over at least 1-3 years. Initial testing would be at about monthly intervals for 6 months and then 3 monthly intervals. Most participants are likely to show no significant decline in test results, but about 10% (based on our prior studies) may show decline in memory. In addition, it's possible that some participants with symptoms or even early dementia may not decline on the tasks used in this study. However, it is important to be aware that it's therefore possible that you will show decline in memory. We ask you to think carefully about whether you would like to know this information since you should not participate if you are unwilling to face this possibility at this time. In addition, a decline on the computer tests does not necessarily mean a person will get dementia or Alzheimer's disease, because there are many other possible reasons. If decline is found, it is important to consider undergoing a doctor's evaluation to seek identifiable causes of the decline. It would be your decision whether to undergo such evaluation and if so, by whom, whether by your own general practitioner (GP) or a Memory Clinic (which would liaise with your GP). Thus, at the start of the study, we would like you to nominate a preferred general practitioner (GP) to contact if necessary. The study is being conducted by the Florey Institute of Neuroscience and Mental Health (FINMH), with Principal Investigator, A/Prof David Darby, assisted by Associate Investigator Dr Amy Brodtmann. Funding is through the FINMH. However, there is a potential conflict of interest of which you should be aware. The Principal Investigator, A/Prof Darby, is a Founder, former director, and currently consultant and substantial shareholder of Cog State Ltd, which is an ASX listed public company which supervises the commercial development of some of the cognitive testing software [22].

## VII. ALZHEIMER DISEASE & FRONTOTEMPORAL DEMENTIA MUTATION DATABASE (AD&FTDMDB)

The Alzheimer Disease & Frontotemporal Dementia Mutation Database (AD&FTDMDB) aims at collecting all known mutations and non-pathogenic coding variations in the genes related to Alzheimer disease (AD) and frontotemporal dementia (FTD). The database is updated continuously and contains mutations reported in the literature and at scientific meetings, and unpublished mutations directly submitted to the database. To date, AD&FTDMDB contains mutations in the genes encoding the Amyloid Beta Precursor Protein (APP), Presenilin 1 (PSEN1), Presenilin 2 (PSEN2), chromosome 9 open reading frame 72 (C9orf72), Chromatin Modifying Protein 2B (CHMP2B), fusion (involved in t(12;16) in malignant liposarcoma) (FUS), Granulin (GRN), Microtubule Associated Protein Tau (MAPT), TAR DNA binding protein (TARDBP) and Valosin-containing Protein (VCP) and holds 424 different mutations observed in 1451 patients or families.

### A. AD & FTD and PD Mutation Database

The Alzheimer disease and frontotemporal dementia (AD&FTLD) and Parkinson disease (PD) Mutation Databases make available curated information of sequence variations in genes causing Mendelian forms of the most common neurodegenerative brain disease AD,

frontotemporal lobar degeneration (FTLD), and PD. They are established resources for clinical geneticists, neurologists, and researchers in need of comprehensive, referenced genetic, epidemiologic, clinical, neuropathological, and/or cell biological information of specific gene mutations in these diseases. In addition, the aggregate analysis of all information available in the databases provides unique opportunities to extract mutation characteristics and genotype-phenotype correlations, which would be otherwise unnoticed and unexplored. Such analyses revealed that 61.4% of mutations are private to one single family, while only 5.7% of mutations occur in 10 or more families. The five mutations with most frequent independent observations occur in 21% of AD, 43% of FTLD, and 48% of PD families recorded in the Mutation Databases, respectively. From the start, the AD&FTLD Mutation Database stores curated genetic, clinical, and biological information of DNA variations in the Mendelian AD genes APP, PSEN1, and PSEN2 (Table 6) [27]. Because of observed genetic overlaps between the etiology of both, AD and FTLD, all known Mendelian FTLD genes (Table 6) were added to the AD&FTLD Mutation Database from 2004 onward [29,30]. The PD Mutation Database was set up in 2010, essentially in response to the lack of comprehensive LSDBs of PD genes. Today, it contains extended genetic and clinical information of variations in the five most common Mendelian PD genes [23].

## VIII. CRITICAL PATH INSTITUTE (C-PATH) ONLINE DATA REPOSITORY (CODR): COALITION AGAINST MAJOR DISEASES (CAMD) ALZHEIMER'S DISEASE DATABASE

The Critical Path Institute (C-Path), in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, has formed the Coalition Against Major Diseases (CAMD). Members include 6 non profit groups representing patients' interests, 15 leading pharmaceutical companies, the US Food and Drug Administration (FDA), the European Medicines Agency (EMA), 2 institutes of the National Institutes of Health (NIH)—the National Institute on Aging (NIA) and the National Institute of Neurological Disorders and Stroke (NINDS)—and representatives from academia. The coalition's purpose is to transform the drug development paradigm for neurodegenerative diseases and serve as a model for other major diseases. In 2004, the FDA's Critical Path Initiative identified neuropsychiatric diseases and disease models as priority areas of active research opportunities.<sup>2</sup> The work of the coalition will focus on member companies' sharing of precompetitive data, which may include data from placebo groups from clinical trials not submitted as part of a New Drug Application, disease models, and/or protocol elements. In addition, industry will contribute scientific expertise that will lead to improved knowledge across disciplines; an important component in the development of treatments for Parkinson's and Alzheimer's diseases. Improved management of existing knowledge will be aimed at qualifying for use in drug development, novel imaging or biochemical markers (here, both are referred to as biomarkers), and quantitative disease progression models.

The CAMD will intentionally avoid using terms such as “valid” or “surrogate” to describe biomarkers. Instead, the coalition will seek to develop methods that are “qualified for use” based on a rigorous review of scientific data by scientists from industry, academia, and regulatory agencies. These “qualified” methods are expected to lead to an increased efficiency in decision making during the drug development process and to a reduction in drug failures during late phase testing. The database regarding is as follows: The database contains 6,500 patients across 24 clinical trials of AD and MCI, but is not limited to, demographic information, APOE4 genotype, concomitant medications and cognitive scales (MMSE and ADAS-Cog); all data has been remapped to a common data standard (CDISC SDTM v3.1.2) such that all the data can be analyzed across all studies; it is openly available to CAMD members, as well as to external qualified researchers who submit, and are approved for, a request for access; biomarker data of AD biomarkers (imaging, bio fluids, expanded genetics beyond ApoE status), Exact names of test drug candidates from sponsor companies and background therapies per individual case are not included. The Primary applications for the Alzheimer's disease C-Path Online Data Repository are characterizes the dynamics of the placebo-arm within clinical trials of AD and MCI, Serves as a tool for the development of modelling and simulation tools for AD clinical trials; and item level data of clinical scales is present allowing investigators to analyze sub-items for specific analyses.

The data is mapped to the CDISC foundational and AD-specific Study Data Tabulation Model (SDTM). Knowledge of SDTM is required for effective use of the data. Information and training on SDTM training is available from CDISC: no SDTM training is provided within CODR. The data consists of 11 SDTM domains: Disposition (DS), Demographics (DM), Adverse Events (AE), Concomitant Medications (CM), Medical History (MH), Subject Characteristics (SC), Subject Visits (SV), Questionnaires (QS), Laboratory Tests (LB), Vital Signs (VS), and Supplemental Qualifier (SUPQUAL) [24].

### IX. ALZGENE

The goal of the AlzGene database is to serve as a comprehensive, unbiased, publicly available and regularly updated field-synopsis of published genetic association studies performed on AD phenotypes. Eligible publications are identified following systematic searches of scientific literature databases, as well as the table of contents of journals in genetics, neurology, and psychiatry. Data selected for display summarize key characteristics of the investigated study cohorts (e.g., gene overview), as well as genotype distributions in cases and controls (e.g., polymorphism details). For eligible polymorphisms with genotype data in at least four case-control samples, continuously updated random-effects meta-analyses are presented (see meta-analysis methods). Note that data obtained from family-based studies are not included in the meta-analyses, as crude odds ratios cannot be readily calculated from overall genotype distributions. In an effort to facilitate the identification of the most promising meta-analysis results available in AlzGene, a continuously updated list displaying the most strongly associated genes. The list includes

genes/loci which contain at least one variant showing a nominally significant summary or in the analysis of all studies (“All”), or those limited to samples of a specific ethnicity (e.g. “Caucasian”). The nominally significant meta-analyses are then graded based on interim guidelines “Venice criteria” for the grading of the epidemiological credibility of genetic association studies recently developed by the Human Genome Epidemiology Network [25].

#### A. Hu GE Net “Venice Criteria”

We rate overall epidemiological credibility as ‘strong’ if associations received three A grades, ‘moderate’ if they received at least one B grade but no C grades, and ‘weak’ if they received a C grade in any of the three assessment criteria. While we believe that this list represents an up-to-date summary of particularly promising AD candidate genes that warrant follow-up with high priority, we note that many of these may still represent false-positive findings. Briefly, each meta-analyzed association in AlzGene is graded on the basis of the amount of evidence, consistency of replication, and protection from bias. For amount of evidence, we assign the grade ‘A’ when the total number of minor alleles of cases and controls combined in the meta-analyses exceeds 1,000, ‘B’ when it is between 100 and 1,000, and ‘C’ when it is less than 100. For consistency of replication, we assign the grade ‘A’ for I2 point estimates <25%, ‘B’ for I2 values of 25–50%, and ‘C’ for I2 values >50%. Note that this criterion does not apply to meta-analyses with a P-value <1x10<sup>-7</sup> after exclusion of the initial studies [22]. For protection from bias, the guidelines propose consideration of various potential sources of bias, including errors in phenotypes, genotypes, confounding (population stratification) and errors or biases at the meta-analysis level (publication and other selection biases). A grade A implies that there is probably no bias that can affect the presence of the association, grade B that there is no demonstrable bias but important information is missing for its appraisal, and grade C that there is evidence for potential or clear bias that can invalidate the association. Errors and biases are also considered in the framework of the observed summary OR. Whenever the summary OR deviates less than 1.15-fold from the null in meta-analyses based on published data, we acknowledge that occult publication and selective reporting biases alone may invalidate the association, regardless of the presence or absence of other biases, and therefore assign a grade of C. When the summary OR deviates more than 1.15-fold from the null, we assign a grade of C when the modified regression test or excess test suggest the possibility of publication-bias or significance-chasing bias or when the association is no longer nominally statistically significant upon exclusion of the initial study or studies violating HWE. Genetic variants for which a significant meta-analysis result is likely due to linkage-disequilibrium with the APOE-ε4 allele (e.g. in APOC1), are not listed separately as “Top Results”. However, meta-analyses of these genes and polymorphisms are still available via the specific gene-summary pages. View also information below on how APOE-ε4 itself is handled in AlzGene.

## B. Database Organization and Methods Meta-Analysis Methods

For all polymorphisms with minor allele frequencies in healthy controls >1%, and for which case-control genotype data are available in four or more independent samples, crude odds ratios (ORs) and 95 percent confidence intervals (CIs) are calculated from the reported allele distributions for each study. Summary ORs and 95 percent CIs are calculated using the random-effects model (using the 'rmeta' package in R)[24]. This procedure is done including all studies irrespective of ethnicity (denoted by "All Studies" on the meta-analysis figures), and for all ethnic groups with independent genotype data in at least three populations. Whenever applicable, the results of a number of sensitivity analyses are also displayed, e.g. after exclusion of the initial study, after exclusion of studies in which a violation of Hardy-Weinberg Equilibrium (HWE; calculated using the 'HardyWeinberg' package in R) was detected. Overlapping samples (of which usually only the largest is included), studies with missing data, or control samples deviating from HWE are indicated on the meta-analysis graphs. Please note, that when only few studies are included in the meta-analyses (i.e. less than ~10), the random effects model may yield summary ORs and confidence bounds that are slightly anti-conservative.

## C. Inclusion of Genome-wide Association Studies (GWAS)

For the systematic inclusion of data from GWAS and other large-scale studies, GWAS devised the following step-wise protocol, which we believe allows to capture the most relevant genetic information without the need to include every data-point from these studies.

Stage I: Represents the inclusion of genes and polymorphisms "featured" or highlighted by the authors of the GWAS or other large-scale study, usually because they show some degree of genetic association after completion of all analyses, e.g. testing multiple independent samples. These genes and polymorphisms probably represent the most important findings of each GWAS and are therefore included in AlzGene with highest priority. Genomic loci that do not map within any known gene are represented by a surrogate name specifying the cytogenetic location (e.g. "GWA\_1q25.12"). Markers in linkage-disequilibrium with APOE  $\epsilon$ 2/3/4, (i.e.. variants located in APOC1) are summarized as "APOE" as featured gene.

Stage II: Will add GWAS and large-scale study genotype data for polymorphisms already available in AlzGene, i.e. usually derived from candidate gene studies. GWAS data for such overlapping polymorphisms will be added to the gene-specific entries and, if applicable, included and displayed in the meta-analyses. This stage adds valuable information to the existing AlzGene meta-analyses as it is derived from assessments that are largely unbiased with respect to gene function, in contrast to most conventional candidate gene studies. Note that genotype data from large-scale association studies using a "pooled" genotyping approach will not be considered for these analyses due to the sometimes substantial variability of genotype and allele frequencies when compared to subject-level genotyping.

Stage III: Focuses on published meta-analyses of the existing GWAS datasets. Genes and loci resulting from these analyses are treated equivalently to the "featured genes" of Stage I (above). Genotype data on the gene-specific pages

will be extracted from the primary GWAS studies (if their data is publicly available) and displayed alongside a "GWAS meta-analysis" entry. This stage also entails the inclusion of more complex genetic analyses, e.g. those jointly analyzing large numbers of polymorphisms at different loci based on assumptions regarding the functional interconnection of these loci, e.g. in forms of "pathways". To the degree that it can be achieved in this context, these pathway-based results are labelled as such and stored in separate "unmapped" section of the database [25].

## D. Association studies on Mitochondrial Genes

Studies assessing a potential association between AD and genetic variants in the mitochondrial (mt) genome are subject to the same inclusion criteria as studies investigating markers from the nuclear genome, and are displayed on a separate "chromosome graph" (which is adapted from imagery on the "Mito Map" website [http://www.mitomap.org/]). Owing to the specific characteristics of human mt-inheritance (e.g. its multicopy nature and the high frequency of somatic mutation events) and the innate heterogeneity of mt-association studies, however, genotype data from these studies are not included on AlzGene and therefore not subject to meta-analysis.

## E. Association of APOE Polymorphisms with AD

In contrast to essentially all other association findings in AD, the risk effect of APOE- $\epsilon$ 4 has been consistently replicated in a large number of studies across many ethnic groups [25,26]. Many studies have also observed a more modest protective effect for the minor allele,  $\epsilon$ 2. Because the established role of the  $\epsilon$ 4- allele, we did not seek to catalog every APOE- $\epsilon$ 2/3/4-result in the published literature. Instead, as a proof of concept, we only considered the 43 samples included in the previous meta-analysis [26]. Please note that the restrictions in inclusion criteria for the  $\epsilon$ 2/ $\epsilon$ 3/ $\epsilon$ 4 variants in APOE do not apply to any of the other published APOE polymorphisms tested for association with AD (e.g., those in the promoter region). For those, we have attempted to sample and analyze every study fulfilling general AlzGene inclusion and exclusion criteria. AlzGene provides summaries of studies that use genetic association methods on common polymorphisms (minor allele frequency in controls >1%) by either case-control or family-based designs. For a summary of rare mutations in early-onset familial AD genes (APP, PSEN1, and PSEN2) please visit the Alzheimer Disease & Front temporal Dementia Mutation Database of the Department of Molecular Genetics, University of Antwerp, Belgium. See also the Alzforum Mutations Directory [25].

## X. IMPLEMENTATION USING SOFT COMPUTING TECHNIQUES

Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth and approximation. In effect, the role model for soft computing is the human mind. The guiding principle of soft computing is: "Exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost".

The clinical data may consists of missing , incorrect and sometimes incomplete values set so using soft computing is the better alternative to handle such data. The principal constituents of soft computing are fuzzy logic, neural computing, evolutionary computation and probabilistic reasoning. They can also be used in combination with each other for better performance [26].The principal constituent methodologies in soft computing are complementary rather than competitive. Fuzzy logic handles imprecision, neural computing deals with learning, evolutionary computation is for optimization and probabilistic reasoning handles uncertainty. The main aim of the research work is to use the Biomedical Engineering technologies with the datasets given here for an early diagnosis of AD.

### XI. CONCLUSION

There are a lot of clinical tests, drug therapies and diagnostic tools such as biomarkers and neuroimaging techniques are available for the diagnosis of Alzheimer's disease. But the fact is that these techniques are inadequate for the definite diagnosis at the earlier stages. The different database of Alzheimer disease patients discussed here gives a complete idea of how the disease can be predicted before the disease progresses and very much useful for creating computer operated software with the help of various soft computing techniques. From the above understandings it is clear that a newly reliable and efficient method should be developed in order to diagnose the disease with the advanced biomedical engineering technology techniques which can be useful to a great extent for the early and definite diagnosis of the disease.

### REFERENCES

1. Sandeep C S, Sukesh Kumar A, Susanth M J "The Early Diagnosis Of Alzheimer Disease (Ad) Using CAMD, TREAD And NAAC Databases" International Journal for Science and Advance Research In Technology, ISSN ONLINE 2359-1052, IJSART - Volume 3 Issue 3 –MARCH 2017:366-371
2. A Review on the Early Diagnosis of Alzheimer's Disease (AD) through Different Tests, Techniques and Databases AMSE JOURNALS –2015-Series: Modelling C; Vol. 76; N° 1; pp 1-22 Submitted June 2014; Revised Dec. 26, 2014; Accepted Feb. 20, 2015
3. Sandeep C.S, Sukesh Kumar.A, "A Review Paper on the Databases Used for the Early Diagnosis of Alzheimer's Disease(AD)", in proceedings of the International Conference on Intelligent Engineering Systems , EASA College of Engineering, Coimbatore, India, ISBN 978-93-83459-33-9, Apr 2014, p.145
4. Sandeep C.S, Sukesh Kumar.A, "A Review Paper on the Early Diagnosis of Alzheimer's Disease(AD) through Profiling of Human Body Parameters", Scientistlink, Coimbatore, India, 2013, International Journal of Computer Science and Engineering Communications (IJCSEC), Vol.1 Issue.1, pp. 21-29, December 2013.
5. 2010 Alzheimer's Disease Facts and Figures. Rep. Vol. 6. Chicago: Alzheimer's Association, 2010. Print. Alzheimer's and Dementia.
6. Alzheimer's Association 2010. Web. 01 Oct. 2010. <<http://alz.org>>.
7. Alzheimer's Disease International, World Alzheimer Report 2011, "The benefits of early diagnosis and intervention," by Prof Martin Prince, Dr Renata Bryce and Dr Cleusa Ferri, Institute of Psychiatry, King's College London, Summary, p.4
8. ADI press release (<http://www.alz.co.uk/media/nr100921.html>) for "Alzheimer's Disease International World Alzheimer Report 2010: The Global Economic Impact of Dementia," Prof Anders Wimo, Karolinska Institutet, Stockholm, Sweden Prof Martin Prince, Institute of Psychiatry, King's College London, UK. Published by Alzheimer's Disease International (ADI) 21 September 2010
9. Healthy Aging-Normal Aging." WebMD - Better Information. Better Health. Web. 12 Oct. 2010. <<http://www.webmd.com/healthy-aging/tc/healthy-aging-normal-aging?page=2>>.
10. Dementia Definition - Alzheimer's Disease Information on MedicineNet.com." Medicine Net. 21 June 2002. Web. 10 Oct. 2010.

11. Mayo Clinic Medical Information and Tools for Healthy Living. 2010 Mayo Foundation for Medical Education and Research. Web. 21 Nov. 2010.
12. Frosch, M.P., D.C. Anthony and U.D. Girolami, 2010. The Central Nervous System. In: Robbins and Cotran Pathologic Basis of Disease, Robbins, S.L., V. Kumar, A.K. Abbas, R.S. Cotran and N. Fausto (Eds.), Elsevier srl, Philadelphia, ISBN-10: 1416031219, pp: 1313-1317.
13. Harvey, R.A., P.C. Champe, B.D. Fisher, 2006. Lippincott's Illustrated Reviews: Microbiology. 2nd Edn., Lippincott Williams and Wilkins, ISBN-10: 0781782155, pp: 432.
14. Cummings, J.L., H.V. Vinters, G.M. Cole and Z.S. Khachaturian, 1998. Alzheimer's disease: etiologies, pathophysiology, cognitive reserve and treatment opportunities. Neurology. 51: 2-17. PMID: 9674758
15. Yaari, R. and J. Corey-Bloom, 2007. Alzheimer's disease: Pathology and pathophysiology. Semin Neurol. 27: 32-41.
16. <https://www.radiologyinfo.org/>
17. Meha Fatima Aftab and Rizwana S. Waraich /, a review of biochemical markers for early diagnosis of alzheimer's disease ,American Journal of Neuroscience 3 (2) (2012) 54-62
18. <http://adni.loni.usc.edu/>
19. <http://www.oasis-brains.org/>
20. <https://www.ucl.ac.uk/>
21. <https://www.alz.washington.edu/>
22. David Darby1 , Amy Brodtmann1 , Michael Woodward1 , Colin Masters2 , 1 Florey Institute of Neuroscience and Mental Health, Parkville VIC, Australia; 2 University of Melbourne, Melbourne, VIC, Australia THE TRAJECTORY-RELATED EARLY ALZHEIMER'S DATABASE (TREAD) STUDY,P4-188
23. <https://www.molgen.ua.ac.be/ADMutations/>
24. <https://c-path.org/programs/camd/>
25. <http://www.alzgene.org/>
26. A good introduction to evolutionary computing: <http://www.cs.bham.ac.uk/Mirrors/ftp.de.uu.net/EC/clife/www/top.htm>

**First Author** Sandeep C S, B.E, M.E, MBA, PhD Doing

**Second Author** Prof (Dr) Sukesh Kumar A, B.Tech, M.Tech, PhD

**Third Author** Susanth M J, MBBS, MD, DM