

Genetic Basis of Alzheimer's Disease and Its Possible Treatments Based on Big Data

Chenyu Zhang
College of Science
University of Arizona
Tucson, Arizona, United States
chenyuzhang@email.arizona.edu

Abstract—This article is about the genetic basis which causes Alzheimer's disease, and big data which is related to Alzheimer's disease, focusing on which gene and how it can cause Alzheimer's disease and the ways to use big data to try to figure out the treatments for Alzheimer's disease. In addition, the article adopts the general idea about how to use big data to help researchers figure out a better way about the treatment for Alzheimer's disease. The present study shows that the more APOE $\epsilon 4$ exists, the more possibility of getting Alzheimer's disease. The main contribution of the paper is focusing on the genetic basis of Alzheimer's disease as well as giving a general idea of how to use big data to help find the treatment for Alzheimer's disease.

Keywords—Alzheimer's disease; gene express; big data; prevalence; age

I. INTRODUCTION

Alzheimer's disease is a dynamic neurodegenerative issue, which is also the most widely recognized reason for dementia in the older people [1]. Dementia, brought about by an assortment of disarranged, is clinically portrayed by a disintegration in memory, learning, spatial direction, language, perception, as well as judgment sufficiently extreme to meddle with day by day living [2]. Alzheimer's disease, as a rule, presents with shortfalls in momentary memory formation and unsettling influence of extra subjective capacities including word-finding, spatial direction, thinking, judgment, and problem-solving [3,4]. Of all dementia patients, 70% are determined to have Alzheimer's disease(AD) [2]. Age is one of the most critical hazard factors for the sickness. Clinically, Alzheimer's sickness can be partitioned into the early begin stage, which means that patients more youthful than 65 years, and also late-beginning, standing for those more established than 65 years. In the pathological way, it is portrayed by the closeness of plaques of amyloid β peptides and intraneuronal tangles of hyperphosphorylated kinds of microtubule related protein tau. Both the early phase and late-starting sorts of Alzheimer's infection have a genetic part. Alzheimer's disease is the purpose behind 60% to 70% of occurrences of dementia [1,2]. The most widely recognized early manifestation is having troubles in recalling ongoing events [1]. As the malady propels, indications can incorporate issues with language and confusion, which also count as easily getting lost, mood swings, loss of motivation, not overseeing self-care, and conduct issues [1,2]. As the patient's condition gets worse, the individual regularly pulls back from family and society [1]. Progressively, substantial capacities are lost, and at last prompting death [5]. Despite the fact that the speed of progression can differ, the typical life expectancy after conclusion is 3 to 9 years [6]. This topic is really interesting

because this is a disease almost especially only "old disease", and the gene expression has some relations with big data, so the author wants to use big data to analyze it, to see how gene expression affects the Alzheimer's disease prevalence. This research hopes to help scientists figure out how to treat or even cure Alzheimer's disease.

II. ANALYSIS ON GENETIC BASIS OF ALZHEIMER'S DISEASE

A. General introduction of Alzheimer's disease

The reason for Alzheimer's disease is ineffectively understood [1]. About 70% of the harm is accepted to be acquired from a parent, with numerous genes as a rule involved [4]. Other risk factors incorporate a past hurt with head wounds, depression, as well as hypertension [1]. The ailment method is connected with plaques and neurofibrillary tangles in the brain [4]. A conceivable finding relies upon the chronicled foundation for sickness of the patient and psychological testing with clinical imaging as well as blood tests to block other possible causes. Starting signs are normally mistaken for customary ageing [1]. Evaluation of cerebrum tissue is required for a reasonable diagnosis [4].

Causal transformations in three genes have been recognized in beginning stage structures, setting up the important job of amyloid in Alzheimer's disease. Nonetheless, except for a couple of families with autosomal predominant legacy, the example of legacy is not direct in many patients with Alzheimer's ailment, and is doubtlessly brought about by the blend of a few hereditary and natural components which defined as a disease with a complex hereditary foundation. Twin investigations anticipated the heritability late-onset forms to be as high as 80% [7].

Despite the fact, it is highlighted by the presence of plaques of amyloid β peptides and intraneuronal tangles of hyperphosphorylated sorts of microtubule related protein tau based on pathology. Both the early-start type and late-start sorts of Alzheimer's ailment have an innate part. Changes in three genes have been found in early-start type, setting up the significant job of amyloid in Alzheimer's illness, which has become the most for the most part thought about examination since their disclosures. In any case, aside from a few families with autosomal prevailing legacy, the case of legacy isn't immediate in many patients with Alzheimer's infirmity, and is surely realized by the blend of hardly any inherited and natural factors. Twin investigations anticipated generally the heritability of late-begin types to be as high as 80%. For a long time, just a

single hereditary hazard factor, the APOE ϵ 4 allele, was solidly involved in Alzheimer's malady [8].

Early-start sort of Alzheimer's disease families with autosomal predominant instances of legacy gave the basic encounters into the nuclear genetic of Alzheimer's ailment. Extraordinarily, penetrant changes were distinguished in three genes: APP, PSEN1, and PSEN2. Yet most changes in APP are heterozygous missense changes in or near amyloid- β -coding exons 16 and 17, whole gene duplications, unprecedented latent little cancellations and passive missense transformations have also been recognized. This is bringing about either altered amyloid β creation, changes in the extent of amyloid β 42 to amyloid β 40, or extended fibril arrangement. Application duplications most likely will not be totally penetrant, proposing the nearness of defensive innate elements. Transformations in PSEN1 and PSEN2 are in similar manner related with amyloid creation. They obstruct the γ -secretase intervened cleavage of APP, achieving to an extended extent the ratio of amyloid β 42 to amyloid β 40. The APOE ϵ 4 allele was the main settled danger factor for late-start type and early-start type. People with one ϵ 4 allele have around three-times-extended threat of Alzheimer's illness, and those with two ϵ 4 alleles have roughly 15-times-extended vulnerability, contrasted with those with the most broadly perceived genotype, APOE ϵ 3 ϵ 3. In APOE ϵ 4 carriers, lifetime peril of Alzheimer's infection at age 85 years was around to be as high as 35% for female APOE ϵ 3 ϵ 4 carriers and 68% for female APOE ϵ 4 ϵ 4 transporters. Disregarding the way that APOE is regularly thought to be a genetic risk factor, the high peril for APOE ϵ 4 carriers seem, by all accounts, to resemble those related with main genes [8].

B. Genetic analysis of Alzheimer's disease

Alzheimer depicted an 'exceptional substance' happening as extracellular deposits in certain cerebrum areas, which are presently alluded to as amyloid plaques. It was not until the mid-1980s that it was found that the plaques consist of totals of a small peptide called amyloid- β 2,3. The second lesion depicted by Alzheimer, neurofibrillary tangles, happens intraneuronal. In spite of the fact that plaques and neurofibrillary tangles are pathognomonic, it is misdirecting to make the feeling that these are the only important changes happening in the Alzheimer disease's brain. Truth be told, various other structures and function changes will also result, including inflammatory and oxidative stress. The joined results of all the neurotic changes, including the effects of the A β and tau pathologies, are not kidding neuronal and synaptic brokenness and loss. At the hour of death, the cerebrum of a patient with Alzheimer illness might be lighter about 33% than the mind of an age-facilitated, non-psychotic person [9].

A β is made by endoproteolysis of the parental amyloid forerunner protein, which is cultivated by the sequential cleavage of amyloid antecedent protein by gatherings of compounds or chemical buildings named α -, β - and γ -secretases. There are three sorts of proteins with α -secretase movement which all have a spot in the ADAM family which is a disintegrin and metalloproteinase family compound: ADAM9, ADAM10 and ADAM17 [10].

The cleavage and preparing of amyloid antecedent protein can be detached into two pathways. In the ordinary non-amyloidogenic pathway, amyloid forerunner protein is

cut by the α -secretase at a position 83 amino acids from the carboxyl end, making a colossal amino-terminal ectodomain which is produced into the extracellular medium. The created 83-amino-corrosive C-terminal segment, which is called C83, is held in the film and afterward isolated by the γ -secretase, conveying a short piece named p3. Basically, cleavage by the α -secretase occurs inside the A β zone, as such blocking development of A β [10].

C. Genetic basis of Alzheimer's disease

The amyloidogenic pathway is an elective pathway for amyloid antecedent protein which prompts A β creation. The first proteolysis is mediated by the β -secretase at a position discovered 99 amino acids from the C end. This cuts results in the arrival of sAPP β into the extracellular territory, and leaves the 99-amino-destructive C-terminal stub which is likewise called C99 inside the film, with the as of late made N end identifying with the main amino corrosive of A β . Coming about cleavage of this part, which is between buildups 38 and 43, by the γ -secretase liberates an entire A β peptide. Most of the full-length A β peptide made is 40 deposits in length, however a little degree which is around 10% is the 42 buildups variety. The A β 42 variety is dynamically hydrophobic and more slanted to fibril improvement than A β 40, and it is this more extended structure that is moreover the fundamental isoform found in cerebral plaques [10].

Changes in three genes, including amyloid forerunner protein, PS1 just as PS2, are known to cause autosomal predominant Alzheimer illness, which shows they are connected with an early-start type pathogenesis. One shared attribute of the sickness causing changes in these genes is that they all impact the metabolism or stability of A β . One ordinary transformation in amyloid forerunner protein is known as the Swedish change, in which a twofold amino-corrosive change prompts extended cleavage of amyloid antecedent protein by the β -secretase. Various transformations, for instance, the Arctic change, increase the gathering of A β , inciting early start, forceful sorts of the infection. Changes in the presenilins, for instance, the PS1M146V transformation, increase levels of A β 42, which is more instantly than A β 40. Extended amyloid forerunner protein quality in like manner results in Alzheimer disease [10].

Mutations in the A β area itself change the hydrophobicity and downstream accumulation inclination of A β peptides, prompting an expansion in plaque load. Different mutations, for example, those situated close to the C-terminal of the A β space impact the action of γ -secretase modifying APP, handle with the goal that longer A β peptides are increasingly inclined to get together, including species 42-43 amino acids long [11]. Interestingly, a defensive change in APP distinguished in the Icelandic populace brings about a 40% decrease in the development of amyloidogenic peptides. The other EOAD-causative genes are exceptionally highly homologous from the presenilin gene family. Like a few changes in APP, familial EOAD mutations in PSEN1 and PSEN2 commonly bring about an increased creation of longer and easier getting together derivatives of APP. Mutations in PSEN1 are the most widely recognized reason for ADAD. As of April 2020, more than 350 changes have been distinguished. PSEN1 changes are assessed to add to around 80% of monogenic Alzheimer's disease. PSEN2 mutations are

much rarer, with just around 30 changes distinguished in ADAD families. The impact of PSEN2 changes is more serious than either PSEN1 or APP mutations, with time of outburst in PSEN2 mutation carriers range from 40 to 85 years old [11].

III. ANALYSIS ON BIG DATA

A. General introduction of big data

The ANN—an AI system stirred by the human neuronal synapse structure was introduced during the 1950s. Regardless, the ANN was as of late limited in its ability to handle veritable issues, as a result of the fading gradient and overfitting issues with planning of significant structure, nonappearance of figuring power, and on a very basic level the nonattendance of huge information to set up the PC system. Energy for this thought has as of late reappeared, on account of the availability of huge data, improved figuring power with the current outlines handling units, and novel computations to set up the significant neural framework. Late assessments on this advancement propose its possibly to perform better than individuals in some visual and hear-capable affirmation tasks, which may foresee its applications in prescription and social protection, especially in clinical imaging, inside a sensible time period. This review article offers perspectives on the history, headway,

and employments of significant learning development, particularly as for its applications in clinical imaging [12].

For preparing the calculation, the ML learning strategies are named supervised learning and unsupervised learning. Regulated learning creates a capacity that replicates yield by inducing from preparing information. For this technique, preparing information is set up with numerical or ostensible vectors that speak to the qualities of information and the relating output information. At the point when the output information has a categorized value, the preparation procedure is for the most part alluded to as classification. Notwithstanding, if the output information has an ordered worth, the procedure is alluded to as characterization. Rather than supervised learning, unsupervised learning doesn't include the thought of output information, however rather construes a capacity to portray concealed structures from unlabeled info information. Since the models are unlabeled, there is no target assessment of the precision. In spite of the fact that solo learning incorporates numerous different arrangements including summing up and clarifying key highlights of the information, unsupervised learning is like a bunch investigation in insights, and spotlights on the way which creates the vector space speaking to the concealed structure, including dimensionality decrease and grouping [12].

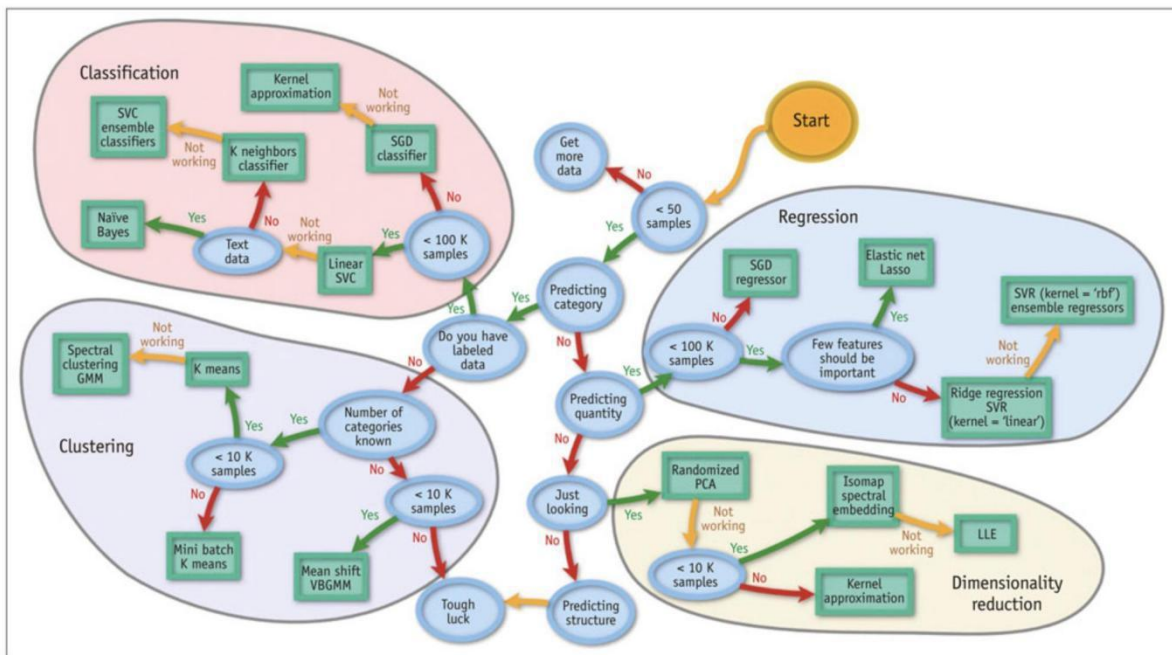


Figure 1. Use big data in general ways [12]

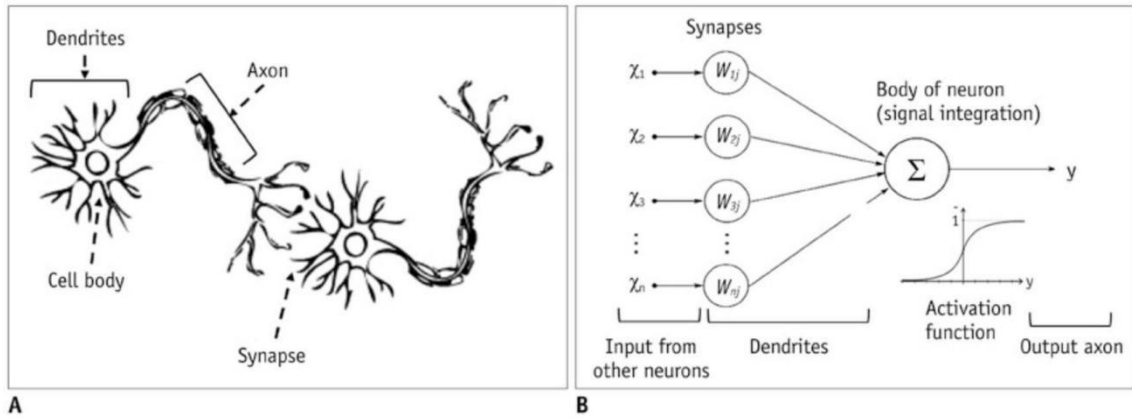


Figure 2. Use big data from eal neurons (A) to artificial neurons (B) [12]

B. Usage of big data on Alzheimer's disease

Utilizing the assortment, aggregation, and prediction examination of huge information volumes could reboot dementia exploration and care as it holds the capability of illuminating its etiology, empowering more ideal conclusion and avoidance methodologies, and perhaps defeating current therapeutic restrictions. Specifically, this potential is bound to be acknowledged by empowering the joining of EHRs, atomic biomarkers, neuroimaging biomarkers, and portable well-being information [13].

In the future, the big data can be used to collect the data between the age and the percentage of getting Alzheimer's disease which can help us figure out the relationship between age and Alzheimer's disease. Big data can also be used to prevent Alzheimer's disease. For example, if we know the risk factor that will let people get Alzheimer's disease, people can try to avoid it, also researchers can use these data to try to figure out how the mechanism of Alzheimer's disease works and where they can use medicines to help people who have Alzheimer's disease.

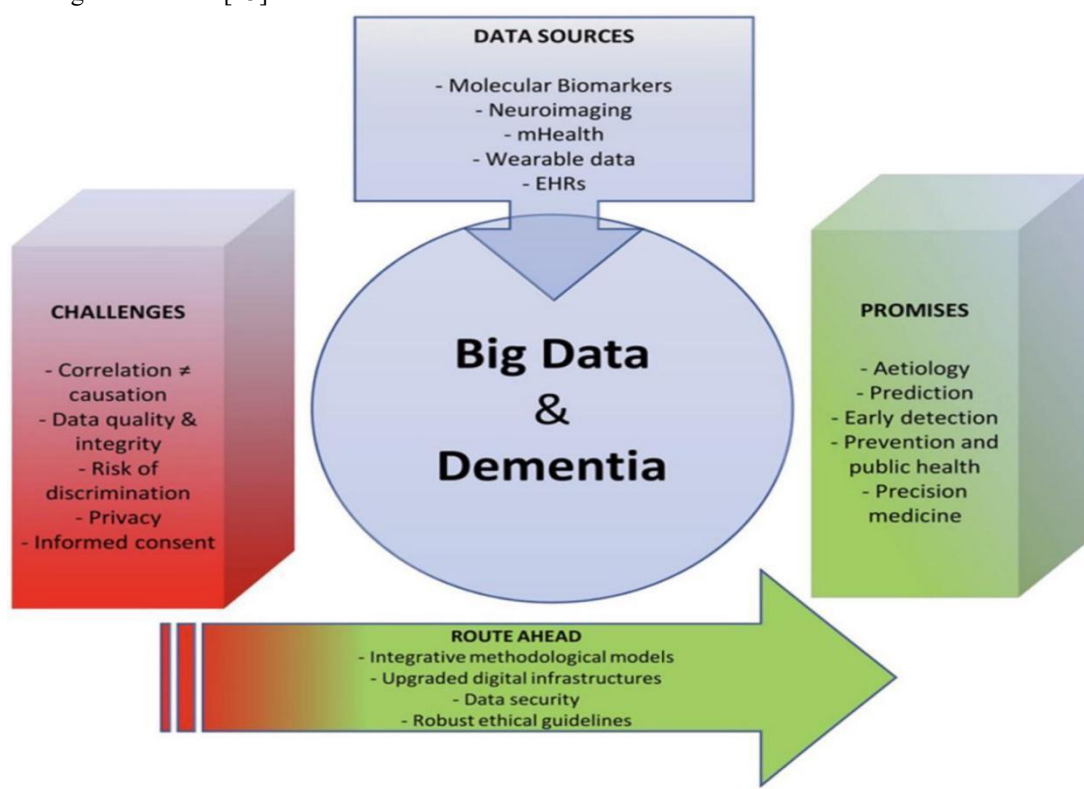


Figure 3. General things that can connect big data and Alzheimer's disease

IV. CONCLUSION

In the above paragraphs, the author analyzes the genes that cause Alzheimer's disease. As mentioned before, APOE ε4 is the most important gene which causes people to get Alzheimer's disease. The APOE ε3 is the healthiest individual. More APOE ε4 is shown in the gene, more possibility for people to get Alzheimer's disease. The length

of the Aβ will be the result as well as one of the reasons that cause Alzheimer's disease. The Aβ40 and Aβ42 are the most common two, which means the length of the Aβ is 40 residues long and 42 residues long. The gathering of Aβ will cause Alzheimer's disease. All of these are related about the early-begin Alzheimer's disease. What is more, PSEN1 mutation and APP mutation are more common than

PSEN2 mutation, but the PSEN2 mutation is more important than PSEN1 mutation and APP mutation.

The author also tries using big data to find a better way to treat Alzheimer's disease. They can help to avoid the limitation of therapy. It can also clarify the etiology and give a better conclusion. However, for the big data part, the research only has methods without having the real data for how to improve the treatment for Alzheimer's disease. It can also give some results for the relationship between the big data and the gene, especially the APOE ϵ 4 gene, to explain more about how and why this gene causes AD. More data can be given for how to use big data to help to improve the treatment for Alzheimer's disease. The future studies can focus more on the big data and treatment for Alzheimer's disease. The researchers can also use the big data with genes that cause Alzheimer's disease to improve the treatment for Alzheimer's disease as well.

ACKNOWLEDGMENT

Most importantly, I might want to offer my most profound thanks to the teachers and academic advisors of my college, who gave me significant direction at each phase of composing this article. Furthermore, I might likewise want to thank every one of my companions and flat mates for their consolation and backing. Without all their edifying direction and amazing agreeableness, I cannot finish my proposal.

REFERENCES

- [1] A. Burns, S. Iliffe, "Alzheimer's disease", *BMJ*. 338: b158, February 2009.
- [2] "Dementia Fact sheet", World Health Organization, 12 December 2017.
- [3] M. F. Mendez, "Early-onset Alzheimer's disease: nonamnestic subtypes and type 2 AD", *Archives of Medical Research*, 43 (8), pp.677-85, November 2012.
- [4] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, E. Jones, "Alzheimer's disease". *Lancet*. 377 (9770), pp.1019-31, March 2011.
- [5] "About Alzheimer's Disease: Symptoms", National Institute on Aging. Archived from the original on 15 January 2012. Retrieved on 28 December 2011.
- [6] H. W. Querfurth, F. M. LaFerla, "Alzheimer's disease". *The New England Journal of Medicine*. 362 (4), pp. 329-44, January 2010.
- [7] "Dementia diagnosis and assessment" (PDF). National Institute for Health and Care Excellence (NICE). Archived from the original (PDF) on 5 December 2014. Retrieved on 30 November 2014.
- [8] Bettens, Karolien, et al. "Genetic Insights in Alzheimer's Disease", *The Lancet Neurology*, Elsevier, 10 Dec. 2012, [Online Available] www.sciencedirect.com/science/article/abs/pii/S1474442212702594.
- [9] LaFerla, M. Frank, et al. "Intracellular Amyloid- β in Alzheimer's Disease", *Nature News*, Nature Publishing Group, [Online Available] www.nature.com/articles/nrn2168.
- [10] J. Hardy, The discovery of Alzheimer-causing mutations in the APP gene and the formulation of the "amyloid cascade hypothesis." *The FEBS journal*, 284(7), pp.1040-1044, 2017.
- [11] N. S. Ryan, & M. N. Rossor, Correlating familial Alzheimer's disease gene mutations with clinical phenotype. *Biomarkers in medicine*, 4(1), pp. 99-112, 2010.
- [12] Lee, June-Goo, et al. "Deep Learning in Medical Imaging: General Overview", *KoreaMed Synapse*, 19 May 2017, [Online Available] synapse.koreamed.org/articles/1027354.
- [13] Ienca, Marcello, et al. "Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy", *Frontiers*, Frontiers, 16 Jan. 2018, [Online available] www.frontiersin.org/articles/10.3389/fmed.2018.00013/full.