Neural Network-Enhanced Investigation of Ferroptosis and Druggability in Early-Onset Alzheimer's Disease

Pratibha Singh and Soumya Lipsa Rath*

Department of Biotechnology, National Institute of Technology Warangal, Telangana, India.

*Corresponding Author: slrath@nitw.ac.in

Abstract

Alzheimer's disease (AD) is a complex neurodegenerative disorder which is multifactorial in nature. Some of its characteristics are slow cognitive decline, memory problems, and behavioral changes. AD patient's brains show a progressive synaptic toxicity, autophagy, neuroinflammation, excess generation of reactive oxygen species (ROS), neuronal death and oxidative stress, which occurs due to imbalance in metal homeostasis in addition to the deposition of tau and amyloid-β protein. In AD brains, lipid peroxidation, iron buildup, and elevated oxidative stress, point to a possible molecular connection between ferroptosis and AD neurodegeneration. In this study, we will present the genetic and bioinformatics perspective on the relationship between ferroptosis and AD using Neural network and Machine learning models. Through our study we identified seven ferroptosis related biomarkers that were expressed abnormally in AD. Machine learning ultimately led to the identification of three candidate genes that could be used to develop prediction model. Neural Network (NN) was then used to develop a diagnostic prediction model, validated by ROC curve, which gave good diagnostic values in the analysis. Ferroptosis is substantially correlated with functional metabolism in AD patients in the analysis. Lastly, analysis of target genes and miRNA-gene interaction networks revealed that the expression of genes, CYBB and ACSL4 can be regulated by hsa-mir-146a-5p, hsa-mir-27a-3p, hsa-mir-106b-5p and hsa-mir-12b-5p. Furthermore, druggability of these targets were also accessed and several drugs were identified which can be potentially repurposed for therapeutic intervention in AD pathogenesis. Our results will be very useful for development of predictive diagnostic model that can be used in the identification of early onset of AD. This study can offer important data support for the clinical treatment and development of personalized therapeutics, thus, helping in prevention of development of Alzheimer's disease.

Keywords: Alzheimer's disease (AD), Ferroptosis, Neurodegeneration, miRNA, biomarkers

Introduction

Alzheimer's disease (AD) is the most common form of dementia worldwide, presenting a major challenge to healthcare systems and society. AD is a multifactorial complex neurodegenerative disorder, characterized by slow cognitive degradation, memory malfunction, and behavioral

transformations. AD patients' brains show progressive synaptic toxicity, autophagy, neuroinflammation, excess generation of reactive oxygen species (ROS), neuronal death, and oxidative stress due to an imbalance in metal homeostasis, in addition to the deposition of tau and amyloid-β protein [1, 2]. Critical brain functions like myelin formation, neuronal activity, neurotransmitter production, and energy metabolism are impacted by iron, which is a transition metal [3, 52]. According to studies, an imbalance in iron homeostasis might prevent the mitochondrial electron transport chain from functioning, which results in oxidative stress [4]. Increased oxidative stress, lipid peroxidation, and iron buildup in AD brains points to a possible molecular connection between ferroptosis and AD neurodegeneration [5]. Ferroptosis is defined as a type of programmed cell death mediated by iron-dependent lipid peroxidation that is different from autophagy, apoptosis, and necrosis [6]. It is specifically characterized by an iron-dependent accumulation of lipid peroxides. [7]. Ferroptosis is primarily dependent on two physiological processes: degradation (particularly autophagy and the ubiquitin-proteasome system) and cell metabolism (particularly lipids, iron, and amino acids) [8].

For many years, traditional diagnostic strategies have encompassed clinical examination, cognitive testing, neuroimaging, and biomarker analyses to detect AD [9]. These methods, however, come with their own limitations, ranging from subjectivity to relatively low sensitivity and invasiveness [10]. Current diagnostic models include identifying people affected by AD by clinical suspicion based on cognitive assessment, neuropsychological testing, and then by neuroimaging (such as MRI, PET imaging) and biofluid assessment (such as cerebrospinal fluid analysis) [11]. Although these methods allow certain useful characterizations of AD pathology, there are certain limitations [12]. Adherence to protocols for clinical assessments and cognitive tests is subjective and predisposes these assessments to inter-rater variability, resulting in inconsistent diagnosis [13]. Additionally, some of the neuroimaging techniques may be insensitive to early-stage AD changes. Although neuroimaging techniques like neuromelanin-MRI of the locus coeruleus can provide early- stage AD detection but are limited by focusing only on norepinephrine and neglecting other key neurotransmitters like serotonin and acetylcholine which may have potential roles in neuropsychiatric symptoms (NPS) system [14]. Several invasive modalities of biomarker analysis, such as biopsy, can be too risky and not feasible, nor always universally accessible to all patients [15]. Over the past few years, ML and neural network-based approaches have shown great potential for a paradigm shift in AD diagnostics [16]. Importantly, the use of these novel computational tools promises improved accuracy, efficiency, and scalability to drive new discoveries in the fight against AD [17].

Alzheimer's disease is a degenerative neurological ailment that worsens over time and for which there is presently no recognized cure. It is essential to create efficient algorithms for illness diagnosis. In some cases, data-driven diagnostic and predictive approaches for Alzheimer's disease are accomplished through the application of machine learning (ML) techniques [18]. Research utilizing machine learning approaches has demonstrated impressive accuracy rates in distinguishing between patients with dementia and healthy controls, as well as between various

forms of dementia [19]. However, more investigation is required to determine the therapeutic usefulness of machine learning in dementia diagnosis. This includes assessing the cost-effectiveness and potential effects of these approaches on patient outcomes [20].

Recent research on AD has explored various approaches to understand its characteristics, causes, and potential treatments, including the use of artificial intelligence (AI) for early detection [21]. Studies have examined AI applications such as speech analysis and voice biomarkers for dementia detection [22]. However, the findings have been limited by variations in research populations and data collection methods, highlighting the need for standardized procedures [23]. Multimodal approaches that combine different data types, such as imaging and non-imaging modalities, have been proposed to enhance the accuracy of early AD diagnosis, but the lack of standardized data collection and analysis methods poses a major drawback [24]. Cognitive evaluation tools, including the mini-mental state examination and neuroimaging techniques, have been utilized for early detection of AD [25]. This also suggests that integrating cognitive assessments with early diagnostic methods could be effective [26]. However, these tools often lack the sensitivity required for precise detection, necessitating the development of more accurate instruments [27]. Machine learning algorithms have shown high accuracy and sensitivity in distinguishing AD patients from control groups, but most often these studies are limited by small sample sizes, requiring validation in larger and more diverse populations [28]. While AI's potential in healthcare, particularly in cognitive problem diagnosis, is recognized, there is a need for more in-depth exploration of specific methodologies and consideration of the costs and financial implications in implementing these technologies [29]. Additionally, research using ensemble machine learning and bioinformatics has identified novel biomarkers and pathways for AD, highlighting the importance of these in-silico methods while acknowledging challenges related to inconsistencies in study methods, demographic factors, and data collection methods [30]. There are a number of unexplored areas in the field of artificial intelligence-based Alzheimer's disease diagnosis. To increase diagnostic accuracy, integrating several data types—such as MRI, PET, EEG/MEG, and sensor data—is crucial [31]. Many of the current research studies concentrate on classifying AD using various machine learning methods [32]. Moreover, some of them also provide early diagnosis of Alzheimer's disease but with certain limitations [33]. Our study addresses the existing limitations and problems in the field of AI-based approaches for cognitive decline detection by offering the early diagnosis of AD utilizing machine learning and neural network-based methodologies [34].

In this study, we will present the genetic and bioinformatics perspective on the relationship between ferroptosis and AD using neural network and machine learning models. Our objective is to target ferroptosis-related genes that were abnormally expressed in AD to uncover potential miRNAs which can be used for further diagnosis and treatment [35]. In order to evaluate the hub gene’s ability to detect AD and their association with immune infiltration in AD patients, these genes were used to build diagnostic models [36]. Finally, we can offer important

data support for the clinical treatment and therapeutic development of AD by screening microRNAs and hub genes [37].

Materials and Methods

1. DATA SET COLLECTION

The Gene Expression Omnibus (GEO) database was used to download the AD dataset GSE118553, and microarray dataset GSE33000 and GSE157239 on AD were retrieved for confirmatory research [38]. Using the FerrDB [39], Kyoto Encyclopedia of Genes and Genomes (KEGG) [40], Gene Set Enrichment Analysis (GSEA) [41], and relevant literature, 106 genes associated to ferroptosis were identified.

2. DIFFERENTIAL GENE EXPRESSION ANALYSIS

In order to acquire differential expressed genes (DEGs) between various comparison groups and controls, differential analysis was performed. In this study, limma (version 3.60.4) was used [42]. A p-value of less than 0.05 and |log2 fold change (FC)| > 1 was taken as the criteria to determine DEGs. SangerBox was utilized to display the DEGs of AD's volcano maps [45].

3. PATHWAY ENRICHMENT ANALYSIS

KEGG Pathway gene annotations were obtained via the KEGG rest API to conduct Gene Set Enrichment Analysis (GSEA) [41]. Genes were plotted against the background set using the GO annotations found in the R-package3 (v.3.1.0) [43]. The clusterProfiler (v 3.14.3) R-package was utilized for enrichment analysis to derive the GSEA results [44]. On the basis of phenotypic grouping and expression profiles minimal gene set was maintained at two and the maximum gene set at five thousand respectively. Phenotypic grouping involves categorizing samples on the basis of phenotypes i.e., observable traits or characteristics. Here, phenotypic groups include individuals with early stage AD, advanced stage AD or no AD (control samples). On the other hand, levels at which genes are expressed in the given sample is given by Expression profiles, which helps in identifying the upregulated and downregulated genes in specific phenotypes. Using minimal and maximal limits to gene sets ensures inclusion of only relevant pathways eliminating all other insignificant pathways as well as non- specific gene sets which could dilute the relevance of results. Statistically significant result was defined as those with a p-value of less than 0.05 and an FDR of less than 0.1. To investigate the biological roles and associated pathways of DEGs, GO and Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway enrichment analysis were carried out on DEGs [45].

4. CONSTRUCTION OF CLASSIFIER MODEL

Data on gene expression alongwith the time and status of survival were integrated using the R-package, glmnet and RandomForest [46]. Regression analysis was done using the lasso-cox [47] and Random Forest techniques. These were used to perform feature selection for the construction

of a diagnostic model using the most representative genes. To find the best model, 10-fold cross-validation was additionally set up. The outcomes of the two types of machine learning models namely Lasso and Random forest were cross- screened to develop the final diagnostic prediction model.

5. VALIDATION OF DIAGNOSTIC MODEL

Receiver Operating Characteristic (ROC) analysis was performed using pROC in the R-package to produce the Area Under the Curve (AUC) [48]. In order to get the final AUC values, AUC and confidence intervals (CIs) were also analyzed using the pROC CI function. sangerBox was used to illustrate these data [49]. The entire process assisted in observing the expression of distinctive genes in the AD dataset GSE118553.

6. SUBGROUP ANALYSIS

To complete the cluster analysis, ConsensusClusterPlus was used. Here agglomerative pam clustering with a 1-spearman correlation distance and resampling 80% of the samples for ten repeats were used to carry out the analysis. Unsupervised hierarchical clustering was carried out using the gene expression of potential genes as input and R's "ConsensusClusterPlus" [50]. Subgroup DEGs were obtained by applying Limma analysis to the various subgroups, and KEGG and GO were utilized to assess the functional differences among the subgroups. Networkanalyst4 was used to build a gene-miRNA interaction network in order to predict the miRNAs of potential genes [51].

Results and Discussion

1. Pathway enrichment analysis of differentially expressed genes (DEGs) shows predominance of Ferroptosis pathway in up regulated genes.

The Gene Expression Omnibus (GEO) database was used to download the AD dataset GSE118553. This database is based on analysis of the transcriptome in likely asymptomatic and symptomatic Alzheimer brains. The human brains of 27 control, 33 Asymptomatic AD (AsymAD), and 52 AD participants were microarray profiled, and both differential and co-expression analyses were carried out. AsymAD refers to those who have intact cognition and neuropathology suggestive of Alzheimer's disease (AD). These people have a high likelihood of developing AD, however it is still uncertain whether transcriptome changes in the brain could reveal mechanisms underlying their vulnerability to AD. Cerebellum is one of the tissues that is largely unaffected by the disease, however frontal, temporal, and entorhinal cortical regions are known to be affected by AD neuropathology. The frontal cortical area in AD patients may be the starting point for basic brain alterations, according to transcriptome activity in AsymAD participants. Furthermore, this dataset provides new insight into the initial biochemical changes

occurring in the brain prior to clinical AD diagnosis, opening up new pathways for therapeutic interventions aimed at preventing AD.

An enhanced linear regression model called "linear models for microarray data" (Limma) provides differential expression screening for the aim of discovering differentially expressed genes (DEGs). Differential analysis was used to get DEGs between various comparison groups and controls. We identified a total of 3,750 differentially expressed genes (DEGs) in the dataset corresponding to Alzheimer's disease (AD) utilizing the Limma (Linear Models for Microarray Data) method, out of which 2,107 were up-regulated and 1,643 were down-regulated (Fig. 1a). In differential expression of genes, some genes show increased while others show decreased expression levels under specific conditions or in response to certain stimuli. When compared to a control group, these genes are termed as up- regulated and down- regulated genes respectively. The up-regulated genes are crucial for many biological functions and can reveal important information about how cells react to stimuli and how diseases are caused, hence only up-regulated genes were considered for further processing. The candidate genes were found to be predominantly enriched in the pathways for "Ferroptosis," "Fatty acid degradation," and "Focal adhesion" (Fig. 1b). This was identified by using KEGG pathway analysis. On the basis of phenotypic grouping and expression profiles, we maintained the minimal gene set at five and the maximum gene set at five thousand during the pathway analysis. Among the three obtained pathways, Ferroptosis pathway covered a relatively larger proportion among the candidate genes.

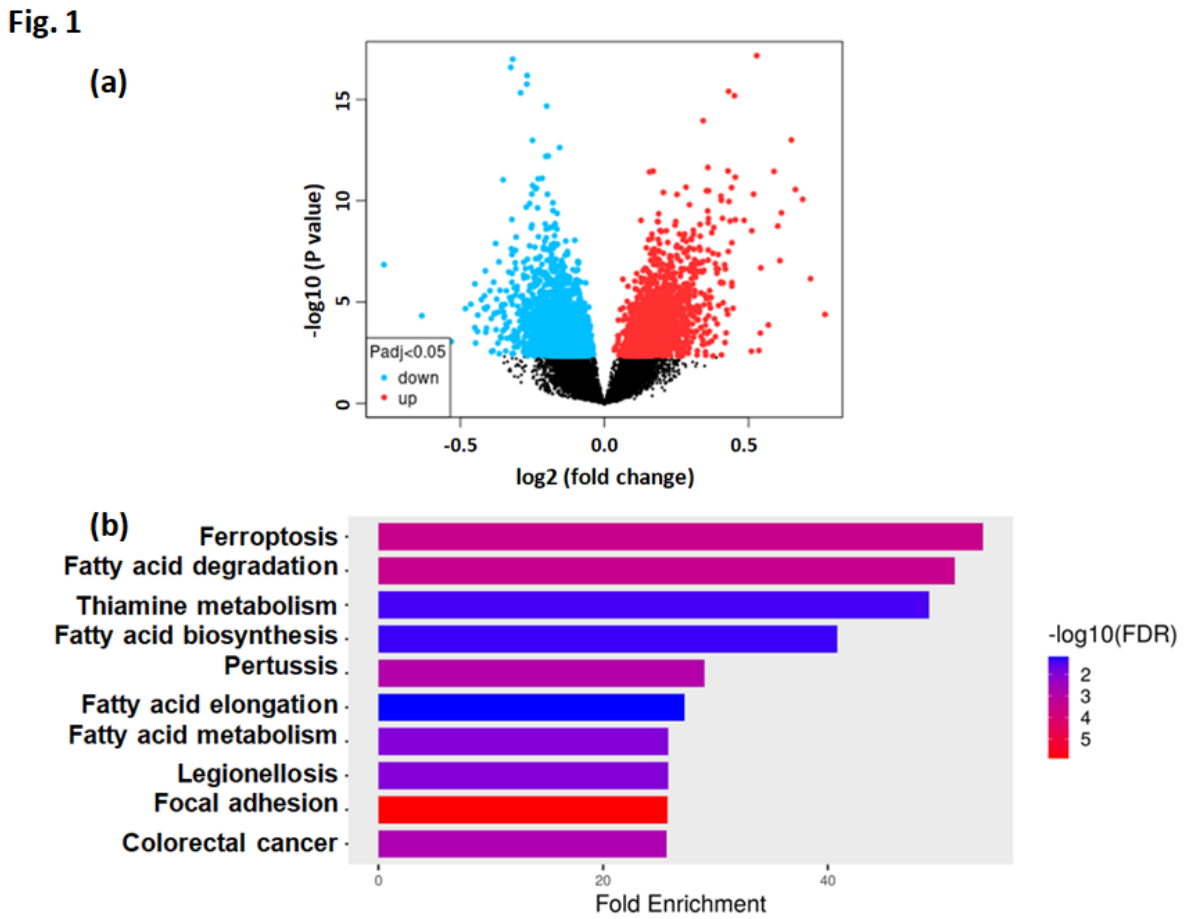

Fig. 1 a Volcano map of differentially expressed genes (DEGs) in Alzheimer's disease dataset (GSE118553), b KEGG analysis pathways associated with DEGs obtained from Alzheimer's dataset.

According to Gene Ontology (GO) analysis, candidate genes were primarily found in "mitochondria", "organelle envelope" and "envelope" in terms of cellular components (Fig. 2a). The "positive regulation of multicellular organismal process", "cell motility", "localization of cells" and other transport-related processes make up the main biological functions of the candidate genes (Fig. 2b). The three most significant roles of the candidate genes, as determined by molecular function (MF), were "oxidoreductase activity", "heme binding" and "iron ion binding" all of which are closely associated with ferroptosis (Fig. 2c). From the GO and KEGG pathway analyses, the insights on different biological roles and other pathways associated with the DEGs were obtained from AD dataset. Results indicate that there is a strong molecular connection between Alzheimer's disease and ferroptosis process.

Fig. 2

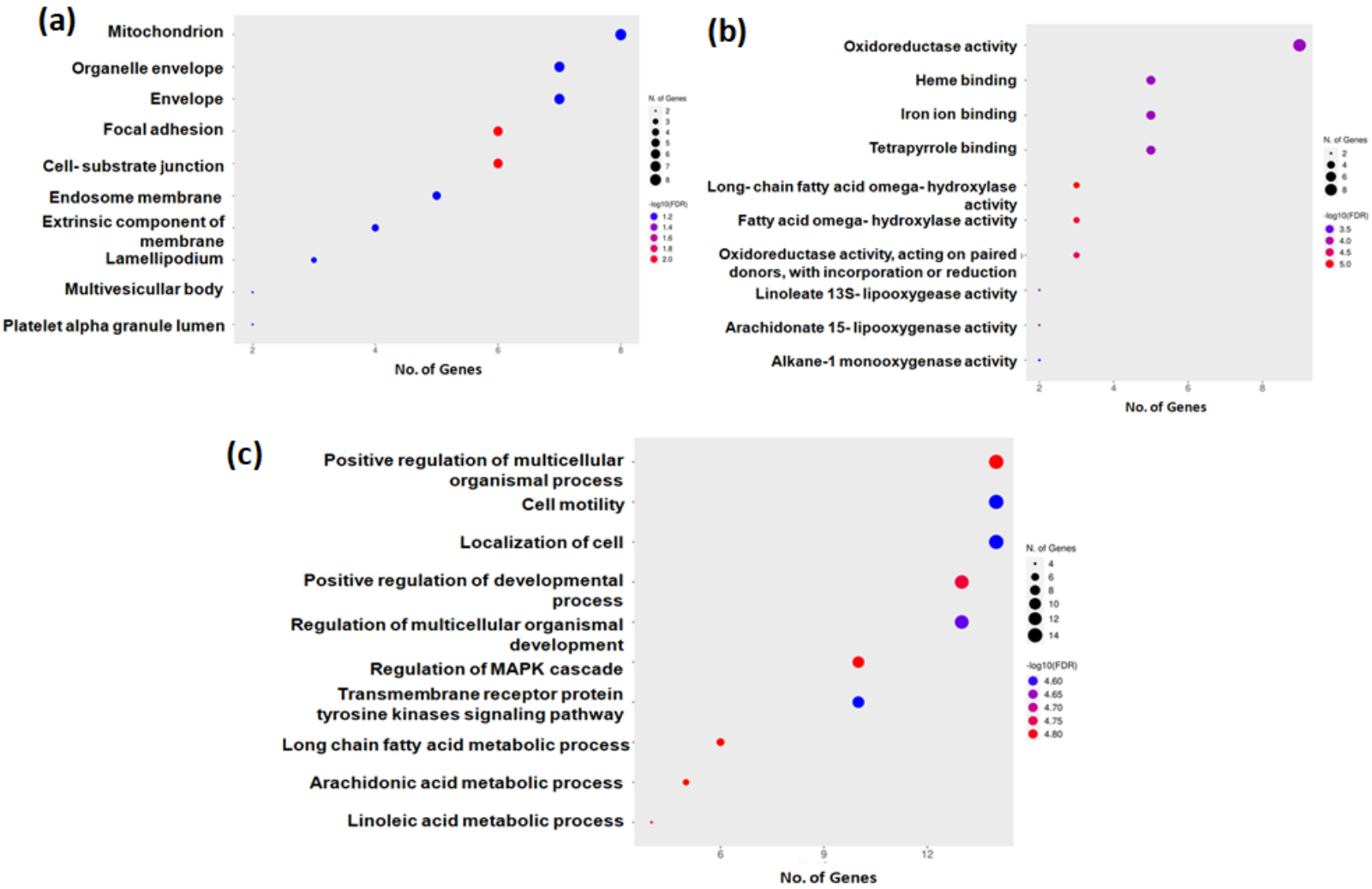

Fig. 2 a GO analysis of potential genes for biological processes, **b** GO analysis of potential genes for cellular distribution, and **c** GO analysis of potential genes for molecular function

2. Identification of candidate genes using Machine learning models.

2.1. Potential candidate genes were identified utilizing supervised ML models on AD dataset and ferroptosis related genes

In the previous section, up-regulated differentially expressed genes alongwith their associated pathways and functions were identified. In this section, we have also included genes that are linked to ferroptosis. 106 genes associated to ferroptosis were identified using the FerrDB, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Set Enrichment Analysis (GSEA), and relevant literature. Subsequently, a set of common genes were obtained after performing a crossover on both the gene sets of AD DEG and ferroptosis related genes (SI. 1). According to previous comparative studies done on performance and accuracy of various Machine leaning

(ML) models (like Lasso (Least Absolute Shrinkage and Selection Operator), Random Forrest, Support Vector Machine (SVM) and Decision Tree), Lasso and Random Forest have been found to give exceptionally good models with high accuracy [46, 47]. Hence, two machine leaning models viz., Lasso and Random Forest (RF) were developed using the obtained common genes. Data on gene expression alongwith the time and status of survival were integrated using the R-package, glmnet and RandomForest. In order to identify potential genes, LASSO regression was first utilized. LASSO is a supervised regularization technique. The AD dataset yielded the identification and determination of seven putative candidate genes namely CYBB, AKR1C1, ACSL4, STEAP3, CISD1 and NFS1 (Fig. 3a, b). Additionally, we employed RF (Random Forest) regression which is also a supervised ML technique to find candidate genes, from which three possible candidate biomarkers were identified namely CYBB, AKR1C1 and ACSL4 (Fig. 3c). Both the models were then used to perform feature selection for the construction of a diagnostic model using the most representative genes. To find the best model, 10-fold cross-validation was additionally set up. After cross-analyzing the genes found by these two-machine learning (ML) algorithms, three potential genes (CYBB, AKR1C1, and ACSL4) were identified which can be used to develop the final diagnostic prediction model. We verified the diagnostic value of these three candidate genes using ROC curves at the point in the AD dataset. All candidate genes were employed as joint indicators and the obtained values were (AUC 0. 75, CI 0.91–0.60) thus, indicating a very high diagnostic significance (Fig. 3d). Furthermore, an investigation of these three potential gene's expression profiles was also carried out. Except for AKR1C1, which did not change significantly, the data show that there were statistically significant differences ($p < 0.05$) among the candidate genes (Fig. 3e). The entire process assisted in observing the expression of distinctive genes in the AD dataset GSE118553.

Fig. 3

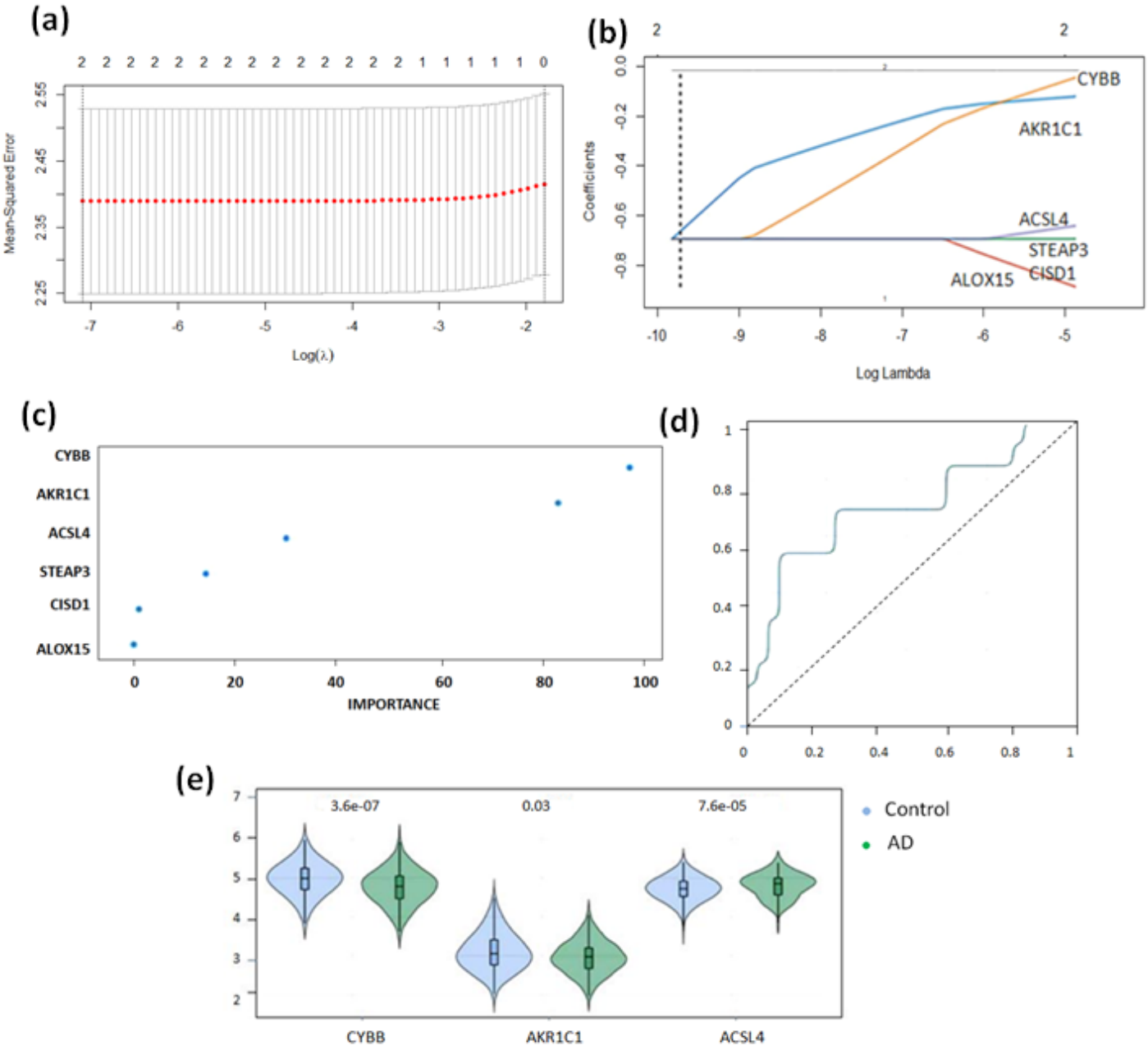

Fig. 3 a, b Candidate gene identification for Alzheimer's disease dataset using Lasso regression, **c** Candidate gene identification for Alzheimer's disease dataset using Random Forest (RF) regression curve, **d** validation using ROC curve for Alzheimer's dataset, **e** Analysis of candidate gene expression profile for AD dataset.

2.2. Unsupervised ML models were used to confirm the association of relevant candidate genes with both AD and ferroptosis.

We conducted consensus clustering (CC) analysis on the AD dataset GSE118553, taking intra-group consistency into consideration. By averaging the intra-group consistency of clusters, we obtain the maximum number of clusters, i.e., K = 2. Based on the clustering chart, we discovered that the clustering between groups was strongest at K = 2, which led to the formation of the two unsupervised clustering subgroups, C1 and C2 (SI. 2a, b). Violin plots were used to show the candidate gene expression levels in the two groups (Fig. 4b). As observed earlier, here also AKR1C1 showed considerable fluctuation ($P < 0.001$). After performing limma analysis on the two subgroups, 827 DEGs in total were found. 523 were up-regulated and 304 were a down-regulated gene which is almost one- fourth of the up-regulated genes that were initially found from the AD dataset (GSE118553) (Fig. 4a). This shows the efficiency of the proposed diagnostic predictive model. According to GO analysis, the DEGs were mostly found in the "focal adhesion" and "cell substrate junction" regions of the cell (Fig. 5a). The "long chain fatty acid metabolic process" and "arachidonic acid metabolism" are the main biological processes linked to DEGs (Fig. 5b). The most important DEGs, according to MF analysis, were "heme binding" and "iron ion binding" (Fig. 5c).

Fig. 4

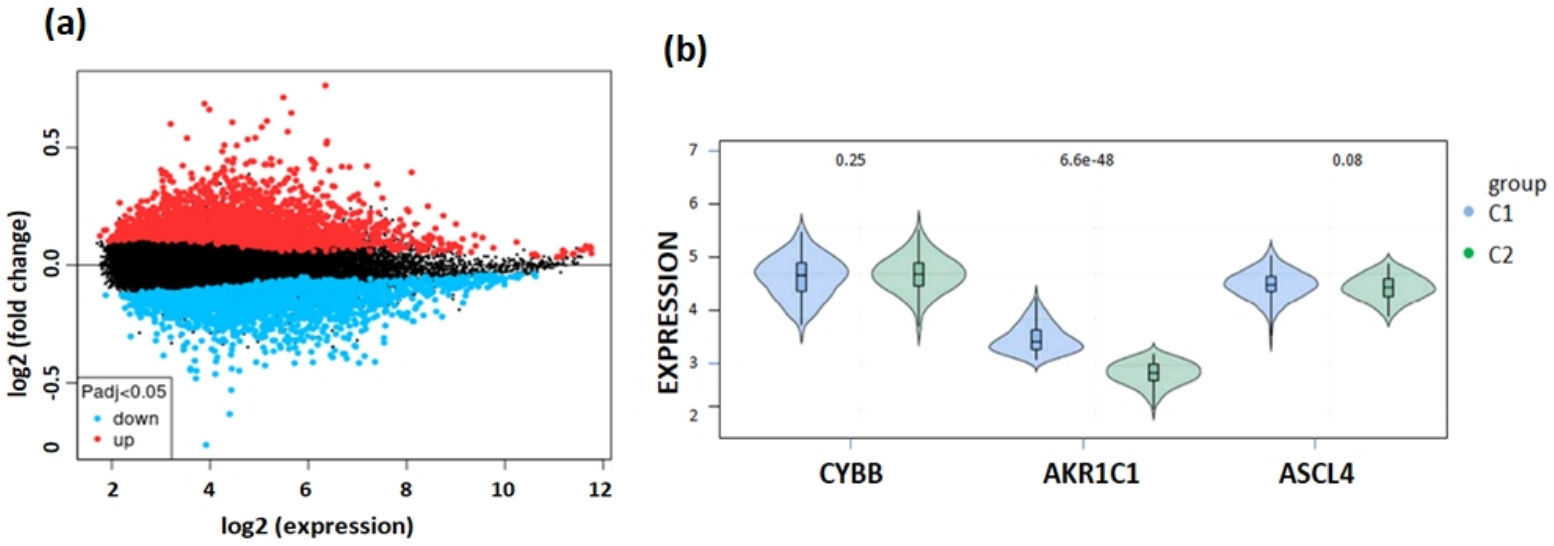

Fig. 4 a Volcano map showing DEGs obtained from subgroup analysis, **b** Analysis of differential expression of relevant candidate genes between subgroups using violin plot

Fig. 5

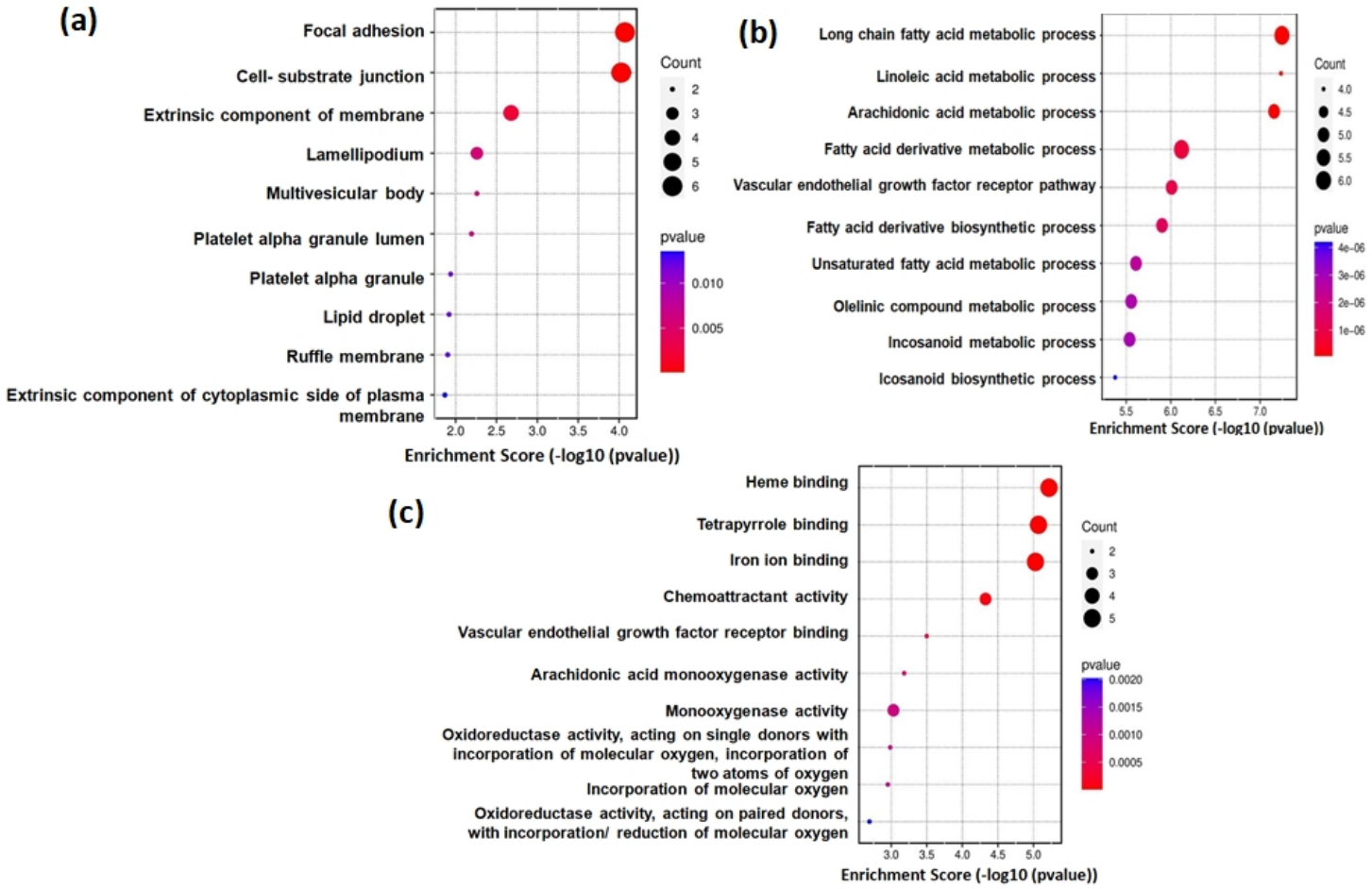

Fig. 5 a Candidate GO analysis of candidate genes for cellular composition and distribution, **b** GO analysis of candidate genes for biological processes, **c** GO analysis of candidate genes for molecular function

The functional analysis of the subgroups obtained from unsupervised learning confirmed the association of relevant candidate genes with both AD and ferroptosis.

3. Controlling expressions of these candidate genes is an improved therapeutic strategy for early onset of Alzheimer's disease (AD).

Using Networkanalyst, we created miRNA-gene interaction networks. A gene-miRNA network visually represents the interactions between genes and associated numerous microRNAs (miRNAs). miRNAs can regulate several genes, making these networks useful resources for biological research. Based on their centrality or significance within the network, the network can assist in prioritizing miRNAs or genes to concentrate on for more validation. Moreover, by

examining gene-miRNA networks, disease-associated dysregulated miRNAs and their target genes can be identified, offering valuable information on the underlying molecular mechanisms of the disease and suggestive of a potential therapeutic strategy by controlling their expressions. It may also be used as a biomarker for prognosis, diagnosis, or response to treatment.

CYBB (Cytochrome b-245 beta chain) contributes to the generation of reactive oxygen species (ROS) through NADPH oxidase complex, which has been linked to oxidative stress—a major contributing factor to Alzheimer's disease [53]. Similarly, ACSL4 (Acyl-CoA synthetase long-chain family member 4) is involved in the production of polyunsaturated fatty acids, which is a process related to lipid metabolism. Dysregulation of lipid metabolism is also linked to the etiology of Alzheimer's disease [54, 55]. After the construction of potential gene-miRNA network, it was shown that hsa-mir-146a-5p, hsa-mir-27a-3p, hsa-mir-106b-5p and hsa-mir-12b-5p could control the expression of ACSL4 and CYBB (Fig. 6).

FIG. 6

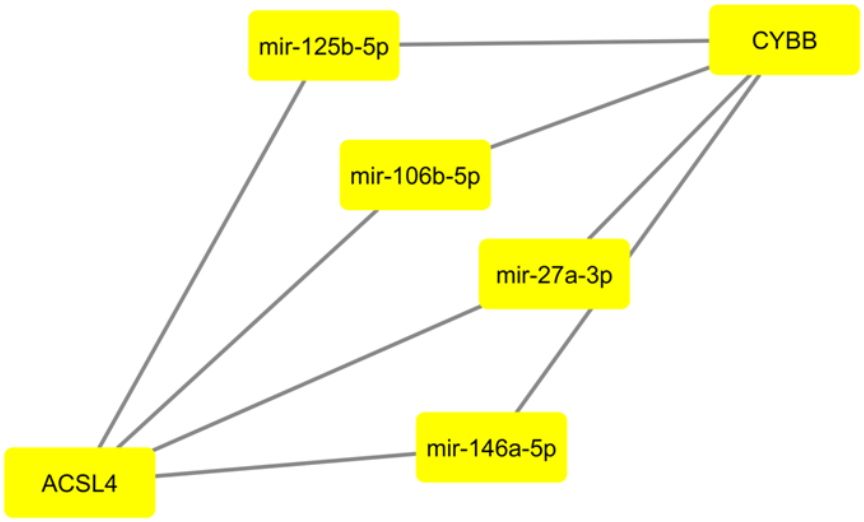

Fig. 6 Network showing interactions between candidate genes and miRNAs

AKR1C1 (Aldo–Keto Reductase 1) is not emphasized here because it does not exhibit notable expression changes in the diseased state being studied. Furthermore, the factors or certain pathological circumstances like oxidative stress and lipid peroxidation that cause Alzheimer's disease may have a lesser impact on AKR1C1 expression than CYBB and ACSL4 which suffers up and down regulations because of the aforementioned factors. Additionally, research suggests that AKR1C1 may be associated with a more chronic or later-stage process which is not the

primary focus in the current analysis, whereas CYBB and ACSL4 are a part of an acute response to early pathology alterations in Alzheimer's [56].

One important modulator of disease gene expression is micro RNA, or miRNA. hsa-miR-146a-5p is one of the most extensively studied miRNAs in Alzheimer's disease, which increases in the brains of people with AD, especially in disease-affected regions like the cortex and hippocampus. It is important for controlling neuroinflammation and the innate immune response, two processes that are fundamental to the pathology of Alzheimer's disease. Although miR-146a-5p may serve as a defense mechanism to regulate excessive inflammation, its persistent upregulation in Alzheimer's disease may lead to prolonged neuroinflammatory responses that accelerate neurodegeneration [57]. hsa-miR-27a-3p regulates both chronic inflammation and dysregulation of lipid metabolism, which are two key elements of AD pathogenesis. It also has a role in regulating synaptic activity, which is important in AD, because synapse loss correlates significantly with cognitive decline [58]. hsa-miR-106b-5p regulates the expression of the APP (Amyloid Beta Precursor Protein) gene. The processing of APP results in the production of amyloid-beta (Aβ) peptides, which build up to form the plaques that are characteristic of Alzheimer's disease. It also regulates genes related to the cell cycle and has been shown to have neuroprotective effects by preventing the expression of pro-apoptotic factors. A lack of regulation of miR-106b-5p has been linked to cognitive decline in Alzheimer's disease because of its effects on synaptic plasticity and neurogenesis [59]. mir-125b-5p is the least studied miRNA in context of AD but some structurally and functionally similar miRNAs are found to be associated with genes responsible for regulating tau aggregation and neuroinflammation [60].

Hence, numerous studies have demonstrated the extensive association between miRNAs (hsa-mir-146a-5p, hsa-mir-27a-3p, hsa-mir-106b-5p) and AD, which in turn can be used as a reliable indicator or biomarker for diagnosing AD. Many studies have shown variations of these miRNAs in the plasma of AD patients. This also highlights their significance in coagulation, oxidative stress, and inflammatory pathways, namely tau and neurofilament phosphorylation [61]. Although hsa-mir-125b-5p may include target genes involved in synaptic activity, cell viability and neuroinflammation, but its exact role in AD needs to be further investigated. The gene-miRNA interaction network gives an overall picture of the potential miRNAs which could be used to control the expression of the relevant candidate genes i.e., CYBB and ACSL4 (SI. 3).

4. **Assessing druggability of the identified key biomarker genes and miRNAs in AD pathogenesis.**

In the previous section some key biomarkers (CYBB and ACSL4) and their regulatory miRNAs were identified. Here, we aim to evaluate the druggability of these identified targets and explore existing drugs that could potentially be repurposed for therapeutic intervention. Using Enrichr-DSigDB, we identified existing drugs which can be used against concerned biomarkers. The top 10 drugs sorted on the basis of decreasing ‘Combined Score’ are shown in Fig. 7.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Eckol CTD 00002503	0.002747	0.07130	499.63	2946.37
2	PCB 118 CTD 00002737	0.003745	0.07130	356.80	1993.61
3	2,2',4,6,6'-pentachlorobiphenyl CTD 00003860	0.004990	0.07130	262.84	1393.12
4	Sapropterin BOSS	0.005737	0.07130	226.97	1171.32
5	Nitro-L-arginine BOSS	0.005986	0.07130	217.09	1111.12
6	Olmesartan BOSS	0.006484	0.07130	199.70	1006.18
7	chrysin TTD 00007164	0.006732	0.07130	192.01	960.21
8	diazoxide BOSS	0.007230	0.07130	178.28	878.83
9	valsartan BOSS	0.008224	0.07130	155.96	748.73
10	dioxidanide CTD 00006819	0.009217	0.07130	138.60	649.60

Fig. 7 List of top 10 potential repurposed drugs identified for therapeutic intervention in AD pathogenesis

This list provides existing drugs from different sources (like CTD: Comparative Toxicogenomics Database, BOSS, TTD: Therapeutic Target Database) which could be potentially repurposed for target condition and pathway which is AD and ferroptosis. These drugs can help via modulating oxidative stress, iron homeostasis, neuroinflammation or signaling pathways. Here, p value gives a statistical significance of the association between the compound and the target condition i.e., AD and target condition which is ferroptosis. Smaller the value greater is the degree of association. Similarly Odds Ratio also provides the association between drug and target condition. Higher odds ratios are suggestive of stronger associations. Combined scores generally combines all the relevant factors like statistical significance, odds ratio and pathway relevance to rank the compounds according to their priority. ‘Eckol’ is having the highest combined score (2946.37) and a very high odds ratio (499.63), hence ranked first in the list suggestive of its strong potential relevance against AD and ferroptosis. Eckol is a phlorotannin obtained from brown algae which is having antioxidant properties that may protect neuronal cells from oxidative stress which is considered to be the major contributors in ferroptosis and hence AD progression [62]. PCB derivatives are explored to identify their potential to influence signaling and iron- related oxidative pathways under controlled conditions. Similarly, ‘Sapropterin’, a synthetic form of tetrahydrobiopterin (BH4), is essential for neurotransmitter synthesis. Although direct studies on its role in AD are limited, BH4's role in neurotransmission indicates its potential relevance [63]. ‘Olmesartan’ is another potential drug which acts as an angiotensin II receptor blocker and thus can improves cognitive function and also reduce amyloid-beta accumulation in AD animal models [64].

Conclusion

AD or Alzheimer’s disease (AD) is a widely known neurodegenerative disease that is caused due to damage to the nerve cells. This damage is usually caused due to aggregation of beta-amyloid or tau proteins. It results primarily in the generation of dementia, short-term memory loss and

related ailments. Apart from the protein aggregates that are formed, certain molecular changes have also been observed for the AD. This includes synaptic toxicity, autophagy, neuroinflammation, excess generation of reactive oxygen species (ROS), neuronal death, and oxidative stress due to an imbalance in metal homeostasis. Metal homeostasis, especially that of iron has been deemed very crucial in recent years. Iron-mediated cell death or Ferroptosis is being increasingly observed in several diseases including AD. Accumulation of lipid peroxides, autophagy and metabolism of iron, lipids and aminoacids are few of the indicators of Ferroptosis.

Cognitive assessment, neuropsychological testing, neuroimaging and biofluid assessment are the diagnostic methods that are used for the detection of AD. However, the results obtained are subjective and inconsistent in several instances. Moreover, Neuroimaging techniques lack sensitivity to early stage AD diagnosis and in several instances the detection might involve invasive methodologies such as biopsy. Machine learning or ML has shown very good accuracy in distinguishing between patients with dementia and healthy controls, as well as between various forms of dementia. However, the studies so far have limited data collection and data analysis methods which gives inconsistent diagnosis.

Here, we have used ML and Neural network (NN) to find a link between AD and Ferroptosis. We have taken a universal AD dataset and identified up-regulated genes in AD. The genes were primarily associated with the Ferroptosis pathways. Functionally the genes were linked to oxidoreductase activity, heme binding and iron ion binding. Subsequently, we found a subset of genes that are found in both AD and Ferroptosis. Among these genes, two, namely CYBB and ACSL4, which are involved in generation of reactive oxygen species that contributes to oxidative stress and production of polyunsaturated fatty acids that is linked to lipid metabolism respectively, were found. We found similar results in both supervised and unsupervised subgroup analysis models. In the end, we have identified four regulatory miRNAs, namely, hsa-mir-146a-5p, hsa-mir-27a-3p, hsa-mir-106b-5p and hsa-mir-12b-5p. The miRNAs, hsa-mir-146a-5p, hsa-mir-27a-3p, hsa-mir-106b-5p have been found associated with AD and Ferroptosis directly, whereas, hsa-mir-12b-5p is a novel biomarker that was identified through our study. Thus, the AD pathogenesis was significantly predicted by the designed diagnostic model based on the identified genes. Furthermore, we have identified several existing drugs against identified biomarker genes and associated regulatory miRNAs, which could be potentially repurposed to influence the AD pathogenesis. ‘Eckol’ ranked first in the list of drugs with highest p- value, Odds ratio and hence Combined score followed by PCB derivatives, Sapropterin and Olmesartan on the basis on decreasing Combined Scores.

The diagnostic value of CYBB and ACSL4 (involved in ROS production and lipid peroxidation) lies in their role as biomarkers of oxidative stress and Ferroptosis. miRNAs regulating these genes associated with both AD and Ferroptosis, provides an additional layers of molecular information, enhancing early detection when detected in biofluids and diagnostic accuracy for AD when integrated with clinical evaluations. Together, they can form the basis for a

comprehensive diagnostic panel for early-stage Alzheimer's disease detection and personalized therapeutic interventions. Although this study establishes a direct link between Alzheimer's disease (AD) and ferroptosis, it is important to note that environmental and lifestyle-related risk factors also play a significant role in contributing to the early onset of AD. The future goal of this study is to integrate the impact of all relevant risk factors including genetic, environmental, and lifestyle-related factors, to enhance the accuracy of early-stage Alzheimer's disease (AD) detection.

Data Availability

Alzheimer's disease dataset (GSE118553) is available in Gene Omnibus (GEO) database. Ferroptosis related genes are extracted from FerrDB and MSigDb. We confirm that the data supporting the findings of this study are available within the article.

Acknowledgements

Authors thank the National Institute of Technology, Warangal, for providing computational facility and MHRD for the research scholar scholarship.

Funding

The authors did not receive support from any organization for the submitted work.

Authors Information

Authors and Affiliations

Department of Biotechnology, National Institute of Technology, Warangal, Telangana, India

Pratibha Singh & Soumya Lipsa Rath

Contributions

PS: Conceptualization, Problem Modeling, Methodology, Writing and revising SLR: Supervision, Conceptualization, Problem Modeling, Methodology, Writing and revising.

Ethics Declaration

Ethical Approval

The authors declare that they are not associated with or involved in any organization or entity that has a financial or non-financial interest in the topics or materials covered in it.

Competing Interests

The authors declare no competing interests.

References

1. Selkoe D.J. (2001) Alzheimer's disease: genes, proteins, and therapy. Physiological Reviews 81(2):741-766. https://doi.org/10.1152/physrev.2001.81.2.741

2. Cummings J.L. (2004) Alzheimer's disease. New England Journal of Medicine 351(1):56-67.

3. Smith M.A., Harris P.L., Sayre L.M., Perry G. (1997) Iron accumulation in Alzheimer disease is a source of redox-generated free radicals. Proceedings of the National Academy of Sciences 94(18):9866-9868.

4. Zecca L., Youdim M.B., Riederer P., Connor J.R., Crichton R.R. (2004) Iron, brain ageing and neurodegenerative disorders. Nature Reviews Neuroscience 5(11):863-873.

5. Gao M., Yi J., Zhu J., Minikes A.M., Monian P., Thompson C.B., Jiang X. (2019) Role of mitochondria in ferroptosis. Molecular Cell 73(2):354-363. https://doi.org/10.1016/j.molcel.2018.12.028

6. Dixon S.J., Lemberg K.M., Lamprecht M.R., Skouta R., Zaitsev E.M., Gleason C.E., Patel D.N., Bauer A.J., Cantley A.M., Yang W.S., Morrison B., Stockwell B.R. (2012) Ferroptosis: an iron-dependent form of nonapoptotic cell death. Cell 149(5):1060-1072. https://doi.org/10.1016/j.cell.2012.03.042

7. Stockwell B.R., Angeli J.P.F., Bayir H., Bush A.I. (2017) Ferroptosis: a regulated cell death nexus linking metabolism, redox biology, and disease. Cell 171(2):273-285. https://doi.org/10.1016/j.cell.2017.09.021

8. Galluzzi L., Bravo-San Pedro J.M., Vitale I. et al (2015) Cell Death and Differentiation 22(1):58-73. https://doi.org/10.1038/cdd.2014.137

9. McKhann G.M., Knopman D.S., Chertkow H. et al (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia 7(3):263-269. https://doi.org/10.1016/j.jalz.2011.03.005

10. Dubois B., Feldman H.H., Jacova C., DeKosky S.T. et al (2007) Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. The Lancet Neurology 6(8):734-746. https://doi.org/10.1016/S1474-4422(07)70178-3

11. Jack C.R. Jr, Bennett D.A., Blennow K., Carrillo M.C. et al (2018) NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia 14(4):535-562. https://doi.org/10.1016/j.jalz.2018.02.018

12. Hansson O., Seibyl J., Stomrud E., Zetterberg H. et al (2018) CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. Alzheimer's & Dementia 14(11):1470-1481. https://doi.org/10.1016/j.jalz.2018.01.010

13. Livingston G., Huntley J., Sommerlad A., Ames D. et al (2020) Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet 396(10248):413-446. https://doi.org/10.1016/S0140-6736(20)30367-6

14. Cassidy, C. M., Therriault, J., Pascoal, T. A., Cheung, V., Savard, M., Tuominen, L., ... & Rosa-Neto, P. (2022). Association of locus coeruleus integrity with Braak stage and neuropsychiatric symptom severity in Alzheimer's disease. *Neuropsychopharmacology*, *47*(5), 1128-1136.

15. Bateman R.J., Xiong C., Benzinger T.L., Fagan A.M. et al (2012) Dominantly Inherited Alzheimer Network. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. New England Journal of Medicine 367(9):795-804. https://doi.org/10.1056/NEJMoa1202753

16. Wang H.Y., Pei J.J., Grundke-Iqbal I., Iqbal K. (2008) Dysregulation of protein phosphatase 2A in Alzheimer disease. Journal of Neural Transmission 115(3):473-480. https://doi.org/10.1007/s00702-007-0860-z

17. Joe E., Ringman J.M. (2019) Cognitive symptoms of Alzheimer's disease: clinical management and prevention. BMJ 357:j1280. https://doi.org/10.1136/bmj.j1280

18. Calabrò M., Rinaldi C., Santoro G., Crisafulli C. et al (2021) Alzheimer's disease genetic variants and their influence on clinical phenotypes. Current Alzheimer Research 18(9):732-752. https://doi.org/10.2174/1567205019666210910111218

19. Weiner M.W., Veitch D.P., Aisen P.S., Beckett L.A. et al (2017) Alzheimer's Disease Neuroimaging Initiative. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. Alzheimer's & Dementia 13(4):e1-e85. https://doi.org/10.1016/j.jalz.2017.02.005

20. Ding Y., Sohn J.H., Kawczynski M.G., Trivedi H. et al (2019) A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. Radiology 290(2):456-464. https://doi.org/10.1148/radiol.2018180958

21. Orrù G., Pettersson-Yeo W., Marquand A.F., Sartori G., Mechelli A. (2012) Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neuroscience & Biobehavioral Reviews 36(4):1140-1152. https://doi.org/10.1016/j.neubiorev.2012.01.004

22. Konig A., Satt A., Sorin A., Hoory R. et al (2015) Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 1(1):112-124. https://doi.org/10.1016/j.dadm.2014.11.012

23. de la Fuente Garcia S., Ritchie C.W., Luz S. (2020) Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: A systematic review. Journal of Alzheimer's Disease 78(4):1547-1574. https://doi.org/10.3233/JAD-200988

24. Pasquini L., Rothlind E., La Joie R., Zsoldos E. et al (2021) Multimodal strategies for early Alzheimer's disease diagnosis. Journal of Alzheimer's Disease 83(1):33-58. https://doi.org/10.3233/JAD-210318

25. Rabin L.A., Smart C.M., Amariglio R.E. (2017) Subjective cognitive decline in preclinical Alzheimer's disease. Annual Review of Clinical Psychology 13:369-396. https://doi.org/10.1146/annurev-clinpsy-032816-045136

26. Jack C.R. Jr, Therneau T.M., Weigand S.D., Wiste H.J. et al (2019) Prevalence of biologically vs clinically defined Alzheimer spectrum entities using the National Institute on Aging–Alzheimer's Association research framework. JAMA Neurology 76(10):1174-1183. https://doi.org/10.1001/jamaneurol.2019.1971

27. Blacker D., Albert M.S., Bassett S.S., Go R.C., Harrell L.E., Moss M.L., Folstein M.F. (1994) Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative. Archives of Neurology 51(12):1198-1204. https://doi.org/10.1001/archneur.1994.00540240062015

28. Moradi E., Pepe A., Gaser C., Huttunen H., Tohka J. (2015) Alzheimer's Disease Neuroimaging Initiative. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104:398-412. https://doi.org/10.1016/j.neuroimage.2014.10.002

29. Wen D., Wei Z., Zhou Y., Li G., Zhang H. (2017) Deep learning techniques for electroencephalogram (EEG) based emotion recognition: a review system. Computational Intelligence and Neuroscience 2017:4765153. https://doi.org/10.1155/2017/4765153

30. Lowe L.C., Gaser C., Franke K. (2016) The effect of the APOE genotype on individual BrainAGE in normal aging, mild cognitive impairment, and Alzheimer's disease. PloS One 11(7):e0157514. https://doi.org/10.1371/journal.pone.0157514

31. Kim S., Seo S.W., Cho H., Yang J.J. et al (2018) Whole brain functional connectivity pattern of significant PET hypometabolism in subjective memory impairment. Scientific Reports 8(1):178. https://doi.org/10.1038/s41598-017-18371-0

32. Escudero J., Ifeachor E., Zajicek J.P., Green C., Shearer J., Pearson S. (2013) Alzheimer's Disease Neuroimaging Initiative. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. PloS One 8(3):e58759. https://doi.org/10.1371/journal.pone.0058759

33. Bhagat Y.A., Flickinger K.L., Nelson R.T., Simon M.J., Pahwa A. (2020) Differentiating Alzheimer's disease and non-Alzheimer's dementia syndromes using unsupervised machine learning techniques. Alzheimer's & Dementia: Translational Research & Clinical Interventions 6(1):e12000. https://doi.org/10.1002/trc2.12000

34. Arbabshirani M.R., Plis S., Sui J., Calhoun V.D. (2017) Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage 145(Pt B):137-165. https://doi.org/10.1016/j.neuroimage.2016.02.079

35. Li Y, Huang X, Zhou H, Liao B, Hong L, Zhang X, Chen G. (2021) Prediction of Alzheimer's disease based on biological information using machine learning and ensemble methods. Journal of Medical Systems;45(5):54.

36. Wang J, Gu BJ, Masters CL, Wang YJ. (2017) A systemic view of Alzheimer disease—insights from amyloid-β metabolism beyond the brain. Nature Reviews Neurology;13(10):612-623.

37. Yu JT, Tan L. (2012) The role of clusterin in Alzheimer's disease: pathways, pathogenesis, and therapy. Molecular Neurobiology; 45(2):314-326. https://doi.org/10.1001/jamaneurol.2013.4560

38. Barrett T., Wilhite S. E., Ledoux P., et al. (2013) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41(Database issue. https://doi.org/10.1093/nar/gks1193

39. Zhou N., Bao J. (2020) FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. Database 2020. https://doi.org/10.1093/database/baaa021

40. Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M. (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1). https://doi.org/10.1093/nar/gkv1070

41. Subramanian A., Tamayo P., Mootha V. K., et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 102(43):15545-15550. https://doi.org/10.1073/pnas.0506580102

42. Ritchie M. E., Phipson B., Wu D., et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43(7). https://doi.org/10.1093/nar/gkv007

43. Ashburner M., Ball C. A., Blake J. A., et al. (2000) Gene Ontology: tool for the unification of biology. Nat Genet. 25(1):25-29. https://doi.org/10.1038/75556

44. Yu G., Wang L. G., Han Y., He Q. Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16(5):284-287. https://doi.org/10.1089/omi.2011.0118

45. Huang D. W., Sherman B. T., Lempicki R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4(1):44-57. https://doi.org/10.1038/nprot.2008.211

46. Breiman L**.** (2001) Random forests. *Mach Learn.* 45(1):5-32. https://doi.org/10.1023/A:1010933404324

47. Roth, V. (2004). The generalized LASSO. *IEEE transactions on neural networks*, *15*(1), 16-28. https://doi.org/10.1109/TNN.2003.809398

48. Robin X., Turck N., Hainard A., et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. https://doi.org/10.1186/1471-2105-12-77

49. Shen Y., Wang Y., Wang Y., et al. (2023) SangerBox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. iMeta 2. https://doi.org/10.1002/imt2.47

50. Monti S., Tamayo P., Mesirov J., Golub T. (2003) Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn.* 52(1):91-118. https://doi.org/10.1023/A:1023949509487

51. Zhou G., Soufan O., Ewald J., et al. (2019) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 47(W1). https://doi.org/10.1093/nar/gkz240

52. Möller, H. E., Bossoni, L., Connor, J. R., Crichton, R. R., Does, M. D., Ward, R. J., ... & Ronen, I. (2019). Iron, myelin, and the brain: neuroimaging meets neurobiology. *Trends in neurosciences*, *42*(6), 384-401. https://doi.org/10.1016/j.tins.2019.03.009

53. Sorce, S., Krause, K.-H., & Jaquet, V. (2012). NOX enzymes as drug targets in neurodegenerative diseases. Current Pharmaceutical Design, 18(3), 217–230.

54. Kagan, V. E., Mao, G., Qu, F., & Tyurin, V. A. (2017). Polyunsaturated cardiolipins: Mechanisms of oxidation and their role in apoptosis and ferroptosis in Alzheimer's disease. Free Radical Biology and Medicine, 111, 53–64.

55. Sorce, S., Krause, K. H., & Jaquet, V. (2012). Targeting NOX enzymes in the central nervous system: therapeutic opportunities. *Cellular and Molecular Life Sciences*, *69*, 2387-2407.

56. Thomas, M. H., & Olivier, J. L. (2016). Arachidonic acid in Alzheimer's disease. *Journal of Neurology & Neuromedicine*, *1*(9). https://doi.org/10.29245/2572.942X/2016/9.1086

57. Cui, X., Zhao, Y., Shi, J., et al. (2020). *miR-146a-5p Regulates Neuroinflammation and Synaptic Dysfunction in Alzheimer's Disease*. *Frontiers in Aging Neuroscience*, 12, 172. https://doi.org/10.3389/fnagi.2020.00172.

58. Feng, X., Zhang, Y., Liu, Y., et al. (2018). *miR-27a-3p Modulates Chronic Inflammation and Lipid Metabolism in Alzheimer's Disease*. *Scientific Reports*, 8, 13939. https://doi.org/10.1038/s41598-018-32475-4.

59. Wang, H., Li, T., Zhang, W., et al. (2020). *miR-106b-5p Modulates Amyloid Beta Precursor Protein Processing and Synaptic Plasticity in Alzheimer's Disease*. *Journal of Alzheimer's Disease*, 78(1), 1-14. https://doi.org/10.3233/JAD-200805.

60. Zhang, J., Wang, Y., Liu, S., et al. (2019). *miR-12b-5p and Related miRNAs in Tau Pathology and Neuroinflammation*. *International Journal of Molecular Sciences*, 20(9), 2265. https://doi.org/10.3390/ijms20092265.

61. Xu, X., Zhang, H., & Li, W. (2021). *The Role of miRNAs in Alzheimer's Disease: Biomarker Potential and Therapeutic Insights*. *Frontiers in Molecular Neuroscience*, 14, 591340. https://doi.org/10.3389/fnmol.2021.591340.

62. Lee, J. H., Ko, J. Y., Oh, J. Y., & Kim, C. Y. (2021). Eckol, a phlorotannin derived from brown algae, exhibits neuroprotective effects through its antioxidant properties. *Marine Drugs*, *19*(9), 537. https://doi.org/10.3390/md19090537

63. Nakamura, K., Wang, W., Kang, U. J., & Takikawa, O. (2019). Role of tetrahydrobiopterin in neurotransmitter synthesis and potential implications for neurological disorders. *Journal of Neurochemistry*, *148*(2), 136-145. https://doi.org/10.1111/jnc.14648

64. Hajjar, I., Hart, M., Chen, Y. L., & Mack, W. J. (2012). Effects of angiotensin receptor blockers on cognitive function in older adults with hypertension: A review. *Journal of the American Geriatrics Society*, *60*(4), 705-715. https://doi.org/10.1111/j.1532-5415.2012.03869.x