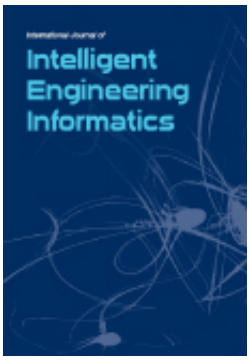



**International Journal of Intelligent Engineering Informatics**

ISSN online: 1758-8723 - ISSN print: 1758-8715
https://www.inderscience.com/ijiei

# Autoencoder with salp optimisation technique for exploring SNP-SNP interactions in Alzheimer's disease

S. Priya, R. Manavalan

**DOI:** 10.1504/IJIEI.2025.10065934

**Article History:**

| | |
|---|---|
| Received: | 11 August 2023 |
| Last revised: | 07 June 2024 |
| Accepted: | 08 June 2024 |
| Published online: | 04 February 2025 |

# Autoencoder with salp optimisation technique for exploring SNP-SNP interactions in Alzheimer's disease

S. Priya*

Department of Information Technology,
Arignar Anna Government Arts College,
Villupuram – 605 602, India
and
Affiliated to Annamalai University,
Chidambaram, Tamilnadu, India
Email: priyasri.ash@gmail.com
*Corresponding author

R. Manavalan

Department of Computer Science,
Arignar Anna Government Arts College,
Villupuram – 605 602, India
and
Affiliated to Annamalai University,
Chidambaram, Tamilnadu, India
Email: manavalan_r@rediffmail.com

**Abstract:** Genetic association research aims to identify genetic variations linked to specific disease states or conditions. Genetic interactions, also known as epistasis, typically involve interactions among numerous single nucleotide polymorphisms (SNPs). Detecting genetic interactions among millions of SNPs in GWAS is challenging. This study presents a two-stage epistasis model called autoencoder based feature selection with Salp optimiser for Epistasis identification (AE-SalpEpi). The autoencoder (AE) approach is designed in the screen phase to pick a subset of features having a high association with Alzheimer's disease. During the selection stage, disease-correlated SNP combinations are chosen using SalpEpi-SO and SalpEpi-MO. The simulated and real Alzheimer's disease dataset are used to quantify the performance of AE-SalpEpi-SO and AE-SalpEpi-MO and also contrasted to state-of-art techniques such as MCASO-Epi and MACOED. Experimental results showed that the suggested methods successfully identified interactions with the APOE gene variation, a known risk factor for Alzheimer's disease.

**Keywords:** epistasis; feature selection; diseases; single nucleotide polymorphism; SNPs; optimisation; multi-objective; Alzheimer's disease; complex diseases.

**Reference** to this paper should be made as follows: Priya, S. and Manavalan, R. (2025) 'Autoencoder with salp optimisation technique for exploring SNP-SNP interactions in Alzheimer's disease', *Int. J. Intelligent Engineering Informatics*, Vol. 13, No. 1, pp.55–77.

**Biographical notes:** S. Priya obtained her PhD in Computer Science from Thiruvalluvar University, Vellore. She works as a Guest Lecturer in Department of Computer Applications and Information Technology, Arignar Anna Government Arts College, Villupuram, Tamilnadu, India. Her areas of interest include bio data mining, Machine learning techniques.

R. Manavalan received his PhD in Computer Science from Periyar University, Salem. Currently, he is working as an Assistant Professor in Arignar Anna Government Arts College, Villupuram. His areas of interest are medical image processing and analysis, soft computing, bio data mining, pattern recognition and theory of computation.

## 1 Introduction

Computational intelligence techniques, such as machine learning algorithms and data mining approaches, enable efficient analysis of large-scale genomic datasets. These methods can handle vast amounts of genetic data and identify patterns or associations that may not be immediately apparent through manual inspection. computational intelligence plays a crucial role in genetic interaction analysis by enabling efficient data analysis, feature selection, predictive modelling, network analysis, and integration of multi-omics data (Ganesan et al., 2014). Machine learning algorithms, such as random forests, support vector machines, and neural networks, can learn complex patterns from genetic data and make accurate predictions about genetic interactions and their effects on phenotypic outcomes (Elshazly et al., 2013). These approaches facilitate the exploration of complex genetic architectures and contribute to the elucidation of underlying biological mechanisms in health and disease (Najm et al., 2020).

Genetic interactions can contribute to identify the disease ratio of particular disease among the individual. By understanding these interactions, researchers can better predict an individual's susceptibility to diseases like Alzheimer's and Parkinson's. This knowledge enables primordial preventive strategies, such as lifestyle interventions or early screening, for high-risk individuals (Kumar et al., 2015). Genetic interactions can influence disease progression and treatment response on an individual level (Humaidi et al., 2021). By identifying these interactions, healthcare providers can tailor treatments more precisely to each patient's genetic makeup.

The GWAS method has recently been applied to examine complicated diseases etiology (Tam et al., 2019). Most prevalent diseases, such as neurological disorders (such as Alzheimer's and Parkinson's), cardiovascular diseases, various malignancies, diabetes, and osteoporosis, are characterised by the accumulation of numerous genes and their interactions. Single nucleotide polymorphisms (SNPs) are the common and prevalent genetic variations in human populations (SNPs). Individuals can bear different nucleotides at each SNP location. The epistasis originally representing gene-level interactions which also examined at the SNP level because the SNPs are the fundamental units in evaluating population variations (Manavalan and Priya, 2021).

Analysing high-dimensional SNP data for GWAS is a difficult task. The huge amount of variables, SNPs, creates a significant computational burden for epistasis computational models (Chattopadhyay and Lu, 2019). In addition, studies of basic human diseases have shown an essential role in explaining the risk of disease in the case of the non-additive

impact of multiple interacting genetic variables (Consortium, 2007). Enumerating all possible potential combinations of millions of SNPs in a GWAS dataset is prohibitive for machine learning approaches for discovering gene-gene interactions (GGIs). Additionally, those variables are often redundant or unnecessary for the disease being examined. Feature selection is frequently required to identify a subset of relevant and prospective variables for the epistasis investigation (Azar et al., 2013). It is common practice in machine learning to perform feature selection as a pre-processing step when the original data contains noisy or irrelevant features that could impair the prediction power of learning algorithms (Assareh et al., 2012). The feature selection approach picks only a subcategory of the essential features, reducing the number of features and boosting learning speed and prediction accuracy (Anter et al., 2013). Feature selection includes two key targets such as maximising the accuracy of prediction and reducing the number of characteristics (Chen et al., 2020).

Medically, neurological conditions impact the spinal cord as well as the brain and other bodily nerves. Approximately 600 disorders are classified as neurological disorders (Ucsfhealth.org, n.d.). Alzheimer's disease, schizophrenia (SCZ), bipolar illness (BD), autism, Parkinson's disease, and sclerosis disease are some of the examples of neurological disorders. Alzheimer's disease is a degenerative neurological disease that results in atrophy of the brain and the death of brain cells. Research has proved that the abnormal protein deposition in brain tissue and surrounding nerve tissue causes Alzheimer's disease (Mayo Clinic, 2020).

The early signs of Alzheimer's disease differ from person to person. For many patients, non-memory cognitive traits like word finding issues, vision/spatial issues, and poor decision-making could be early signs of the disease. Biomarkers (biological indicators of disease detected in brain imaging, cerebrospinal fluid, and blood) are being studied by researchers to detect early changes in the brains of people with MCI and cognitively normal people who may be at higher risk for Alzheimer's disease. Every three seconds, someone in the world has the chance of getting dementia. Worldwide, 55 million individuals will have dementia by the year 2020. This number will roughly double every 20 years, rising to 78 million in 2030 and 139 million in 2050. The majority of the growth of dementia will occur in developing countries. Dementia affects 60% of adults in poor and middle-income nations, but this number will climb to 71% by 2050. China, India, South Asia, and western Pacific neighbouring nations are experiencing the fastest rising number of older people ('2020 Alzheimer's disease facts and figures', 2020).

Most Alzheimer's patients have the late-onset version of the disease, which manifests symptoms in their mid-60s or later. Early-onset Alzheimer's disease strikes adults between the age group of 30 and 60, accounting for less than 10% of Alzheimer's patients. Scientists have made enormous progress in understanding Alzheimer's disease in recent years, and the pace is increasing. Scientists are still unsure about the causes of Alzheimer's disease in most people. Acetylcholine is a neurotransmitter that plays a significant role in brain function. Typically, people with Alzheimer's disease or Parkinson's disease tend to have low levels of acetylcholine. A genetic mutation may cause early-onset Alzheimer's disease in some people. Alzheimer's disease develops late in life as a result of a complex series of brain changes that can take decades to manifest. Genetic, environmental, and behavioural variables are likely to be involved in the development of the diseases (*BrainFacts*, n.d.).

The genetic interaction involves understanding how variations in multiple genes (genetic variants) interact with each other and with environmental factors to influence an organism's phenotype, such as disease susceptibility, drug response, or trait expression. Genetic interactions can manifest in various forms, including epistasis, where the effect of one gene depends on the presence of another, and genetic buffering, where the effect of a mutation is mitigated by another gene. The goal of studying genetic interactions is to unravel the complex networks of molecular pathways underlying biological processes and diseases, ultimately leading to insights into disease mechanisms, personalised medicine approaches, and drug discovery.

Genetic variations are a substantial cause of Alzheimer's and Parkinson's diseases. Identifying the possible genetic causes of neurological disorders is essential to recognise the molecular mechanisms responsible for forming these neurological diseases and to delineate a genotype-phenotype association that can better monitor the disorder's progress and predict future complications (Parenti et al., 2020). Genetic risk factors for neurological diseases are multifactorial and are influenced by gene and environmental interactions. This work focuses on identifying 2-locus genetic interactions responsible for Alzheimer's diseases through AESalpEpi-SO and AESalpEpi-Mo.

This research introduced a novel epistasis identification strategy with a two-phase model called autoencoder-based feature selection and Salp optimiser (AESalp-Epi) to identify SNP interactions. During the screening step, autoencoder techniques are used to pick the featured SNPs that have a high association with disease status. Salp optimisation with G-test is used over the selected SNPs during the search phase to find disease-relevant SNP correlation. The primary goal of this study is to develop an effective epistasis identification technique, AE-SalpEpi, to identify disease-related GGIs from numerous SNPs. Here, the following key contributions to accomplish the research goal.

- Disease-related SNPs in genotyping data can be identified with the help of a new autoencoder-based Salp optimiser for epistasis discovery (AE-SalpEpi). The effectiveness of the suggested strategy is compared to multi-objective chaotic atom search optimisation (MCASO-Epi) technique and MACOED.
- The effectiveness of AESalp-Epi is evaluated based on disease models that include marginal effects (DMEs) as well as disease models that do not include marginal effects (DNMEs).
- The AESalp-Epi model was applied to a real Alzheimer's disease dataset obtained from the ADNI and discovered substantial disease-related genetic associations.

The following is the structure of the research work. Section 2 presents the related work in epistasis detection through computational approaches. In Section 3 outlines the detailed description about the proposed method. Section 4 investigates and discusses experimental analysis. Finally, Section 5 concludes the article by discussing potential futures.

## 2 Related work

In medical research, finding the relation between genes and diseases is a key process for disease prevention, diagnosis and treatment (Emary et al., 2014). Both genetic and environmental risk factors increase pathogenicity. Genome-wide association studies

(GWAS) researchers aim to find genotype variants of interest for several categories of diseases such as hypertension, rheumatoid arthritis, cancer, chronic illness, cardiovascular disease, diabetes, psoriasis, and so on (Wu et al., 2009).

Currently, various methods are available for identifying epistasis, including stochastic search, exhaustive search, and optimisation-based approaches (Sayed and Hassanien, 2024). Evolutionary optimisation techniques are utilised to discover SNP combinations with high scores (Chelliah et al., 2023). These strategies aim to decrease search time complexity, while scoring functions determine the most favourable SNP combinations (Kumar and Sharma, 2023). A multi-objective ant colony optimisation (ACO) technique (MACOED) was proposed for genetic interactions. ACO is applied in the screening stage to filter the SNPs, and the outcome was tested in clean stage using chi-squared test to detect significant SNP combinations (Jing and Shen, 2014).

Bio-based computing has recently been applied to solve demands in the e-health domain, such as patient risk stratification issues and individualised healthy meals for diabetic patients (Chifu et al., 2019). The ACO is a path-finding algorithm inspired by the behaviour of ants hunting for food [190]. Tabu search is a mathematical optimisation paradigm that searches for optimal solutions nearer to the current solution space (Marzouki et al., 2018).

A tabu list is used to prevent the ACO algorithm from selecting SNPs that can be rearranged with primary effects. For GGIs, ACO with tabu list (ACO-Tabu) was implemented and tested on WTCCC datasets including T2D, T1D, RA, and inflammatory bowel disease (IBD) of 5,000 individuals with 500,000 SNPs. The tabu list includes SNPs that are linked to the primary impacts IBD. ACO-Tabu identified that the gene variant TCF7L2 is commonly found in every 5000 generations with 100 runs. The result indicated that there was a strong association between the TCF7L2 gene's SNP rs4506565 and the IGR gene's SNP rs2578050 in T2D data. With varied data sizes, the ACO-Tabu outperforms PLink and INTERSNP, whereas the performance of ACO-Tabu was inferior to BOOST (Sapin et al., 2015).

Wang and Nikouei Mehr (2019) used an optimisation strategy to find the required variables for detecting the epistasis impact effectively in 2019. The simulated soybean dataset included 42,509 cultivars with 20,087 genes. The proposed model can successfully detect 2-way to 5-way epistasis with a 95.5%.

Gonçalves et al. (2020) presented a multi-objective genetic algorithm (MOGA) for finding epistasis in simulated and BC datasets in 2020. MOGA earned 92% average power for the simulated dataset. The parallel MOGA technique took only 3 seconds to conduct the 10-order analysis on a BC dataset including 31,341 SNPs.

A two-stage approach, incorporating multi-objective atom search optimisation for epistasis detection (MASO-Epi) and multi-objective chaotic atom search optimisation for epistasis detection (MCASO-Epi), is introduced to identify two-locus epistasis associations across diverse simulated disease models (Priya and Manavalan, 2021).

The multiple objective genetic algorithms (EpiMOGA) was proposed for GGI identification by Chen et al. (2021) in 2021. EpiMOGA achieved 88% power for the simulated data analysis, whereas EpiMOGA discovered 48 significant 2-order SNP associations in the Alzheimer's disease dataset.

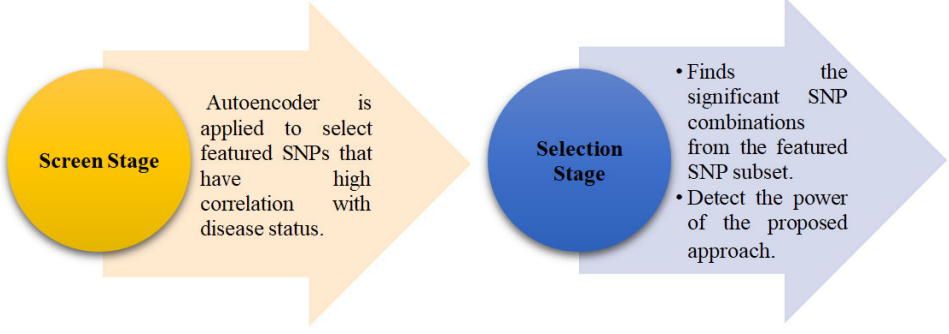
Priya and Manavalan (2022) introduced a two-stage approach named multi-objective salp optimisation for epistasis detection (MSalp-Epi) designed to uncover two-locus epistasis associations across various simulated disease models. In the initial stage, Salp optimisations utilise AIC and K2 score as objective functions to identify non-dominated

disease-related SNPs. Subsequently, in the second stage, the G-test statistic is applied to these non-dominated SNPs to identify significant SNP pairs.

## 3 Autoencoder with salp optimisation (AESalp-Epi) for SNP-SNP Interactions

Identifying genetic interactions is crucial for unravelling the complexity of diseases like Alzheimer's and Parkinson's. It offers insights into disease mechanisms, enables personalised approaches to treatment and prevention, informs drug development efforts, and enhances our understanding of systems biology. The design of novel computational methods can efficiently analyse large datasets and identify potential genetic interactions more quickly than traditional experimental approaches. The AESalpEpi-SO and AESalpEpi-MO was applied to find disease-relevant SNP interactions in Alzhemier's disease. The AE-SalpEpi is divided into two stages: screen and selection. Figure 1 exposes the goal of the screen and the clean phases. Figure 2 depicts the general functioning of the proposed system. In the screening phase, the AE method is utilised to find the notable SNPs in the dataset. To recognise the disease-causing SNP pairs, these prominent SNPs are moved on to the selection stage.

**Figure 1** Objectives of screen and selection stage (see online version for colours)




### *3.1 Screen stage – autoencoder for feature selection*

Feature selection is necessary in the context of GWAS to reduce dimensionality, focus on relevant genetic variants, improve interpretability, enhance statistical power, and promote the reproducibility of genetic association studies (Surya and Muthukumaravel, 2023). It is a critical step in identifying genetic factors underlying complex traits and diseases (Tripathi and Patheja, 2022). The autoencoder is applied for selecting the featured SNPs to find genetic interactions. The genotype dataset is passed into the autoencoder technique in the screening stage. The autoencoder selects SNPs that are highly linked with illness state. An auto encoder is a neural networking technique that can learn features by minimising reconstruction errors in an unsupervised manner. The basic structure of autoencoder adapted in this research is shown in Figure 3. A typical autoencoder with a hidden layer of h dimensions is composed of two elements: an encoder and a decoder (Sun et al., 2016).

**Figure 2** Architecture of proposed AE-SalpEpi (see online version for colours)

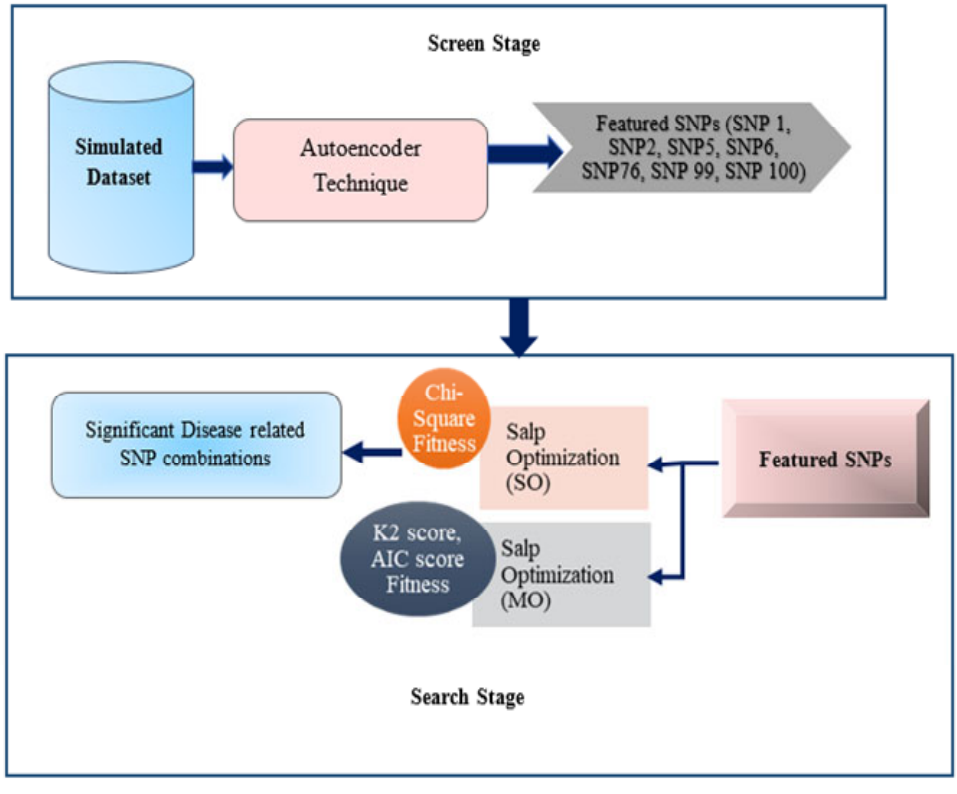
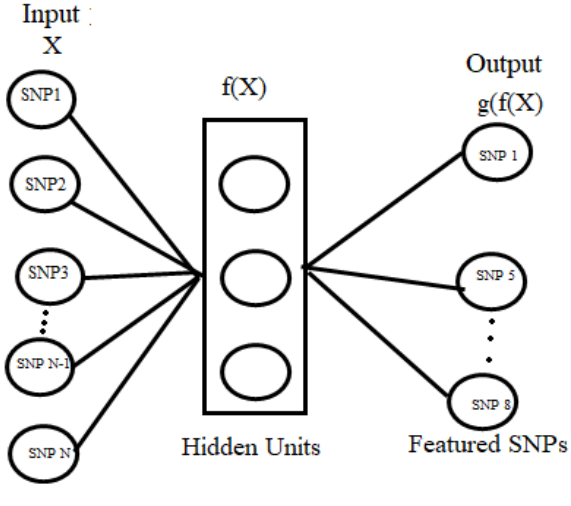



**Figure 3** Structure of autoencoder




The encoder function is represented as

$$f(x) = \sigma_1(XW^{(1)}) \tag{1}$$

The decoder function is presented as

$$X = g\left(f(x)\right) = \sigma_2\left(f(X)W^{(2)}\right) \tag{2}$$

where $\sigma_1$, $\sigma_2$ represents activation functions of hidden and output layer, respectively. The overall function of autoencoder is presented as $g(f(x))$.

In learning process, AE tries to minimise the loss function. The proposed model utilises least square loss in the proposed method for seeking the self-representation:

$$J(\theta) = 1/2m\left\|X - g\left(F(X)\right)\right\|_F^2 \tag{3}$$

where

$m$ denotes sample size.

$||.||_F$ is the Frobenius norm form of the matrices. By optimising $J(\theta)$, the autoencoder compresses the input matrix $X$ as low-dimensional data $f(X)$ and produces the decoded matrix as $g(f(X))$.

The weight matrix connecting the input layer and the hidden layer $W^{(1)} = [w_1, \cdots, w_d]^T$, where the $i^{\text{th}}$ row $w_i$ corresponds to the $i^{\text{th}}$ feature $x_i$. In order to select the most discriminative features from observed data $X$, the row-sparse regularisation on $W^{(1)}$ is exposed as, i.e.,

$$\|W^1\|_{2.1} = \sum_i^d \sqrt{\sum_j^h \left(W_{ij}^{(1)}\right)^2} \tag{4}$$

Equation (4) is reformulated as

$$J(\theta) = \tfrac{1}{2m}\left\|X - g\left(f(X)\right)\right\|_F^2 + \alpha\|W^1\|_{2.1} \tag{5}$$

Here, $\alpha$ represents the trade-off parameter of the reconstructed loss and the regularisation term. To reduce overfitting and improve convergence speed, a weight decay term is also necessary during the training process of a NN. Thus, equation (4) is combined with the weight decay regularisation to form the following object function:

$$J(\theta) = \frac{1}{2m}\left\|X - g\left(f(X)\right)\right\|_f^2 + \alpha\|W^1\|_{2,1} + \frac{\beta}{2}\sum\nolimits_{i=1}^{2}\|W^i\|_F^2 \tag{6}$$

where $\beta$ denotes the penalty parameter, By minimising the equation (6) over the train dataset, we can obtain an autoencoder network for selecting useful features. Figure 4 illustrates the screening stage approach.

### *3.2 Selection stage – salp optimiser for GGIs detection*

Alzheimer's disease is highly complex, involving multiple genetic and environmental factors. During the selection phase, disease-correlated genetic interactions are identified to detect disease risk or progression. Understanding these interactions can provide insights into the underlying biological mechanisms of Alzheimer's disease. By selecting optimal interactions, researchers can identify biomarker panels that offer improved sensitivity and specificity for detecting Alzheimer's disease and monitoring disease

progression. The selected features from the AE are forwarded to the selection phase in order to identify disease-correlated SNP pairs. The G-test is employed in SSA as a fitness criterion. Two variations such as SSA with single objective (SO) and SSA with multi-objective (MO) are chosen to discover the substantial disease related SNP pairs. SalpEpi-SO relies on the G-test as its measure of fitness, whereas SalpEpi-MO use the AIC and K2 scores (Priya and Manavalan, 2022). In SNP selection, the AIC is often employed to compare different models that include varying numbers of SNPs. Lower AIC values indicate better model fit, balancing the trade-off between model complexity and explanatory power. By selecting the model with the lowest AIC, researchers aim to identify the subset of SNPs that collectively provide the best explanatory power for the phenotype. In the context of genetic association studies, the G-test evaluates the independence between a candidate SNP and the phenotype of interest (e.g., disease status). The K2 score is applied to infer the conditional dependencies between SNPs and the phenotype. By evaluating the conditional probability of each SNP given the phenotype and other SNPs, whereas, the K2 score identifies the optimal subset of SNPs that collectively maximise the model's predictive accuracy.

**Figure 4** Pseudo code of screen stage for AE-SalpEpi-SO and AE-SalpEpi-MO

| **Stage 1 – Autoencoder for selecting featured SNPs for GGIs detection** |
|---|
| **Input**<br>Data: Simulated dataset<br>hlayer_size: size of the hidden layer<br>lambda: weight decay parameter<br>Beta: weight of sparsity penalty<br>pretrain_iter: Maximum iteration for training<br>**Output**<br>Fea_set: Selected indexes of features |
| **Screen Stage**<br>Step 1: Load the dataset and initialise the necessary parameters.<br>Step 2: Pretrain the autoencoder to minimise loss value.<br>Step 3: Get the sorted weight of input from the pretrained network<br>Step 4: Return the the index of selected features (SNPs) |

The Pareto optimum front technique (Tuo et al., 2016) assists in selecting non-dominated SNPs from the SNPs found by Salp-MO. The non-dominant SNPs are then subjected to the G-test to categorise 2-locus disease associated SNPs. Figure 5 depicts the pseudo-code of the selection stage for AESalp-Epi model.

**Figure 5** Pseudo code of screen stage for AE-SalpEpi-SO and AE-SalpEpi-MO

**Stage 2 – Salp optimiser for recognising significant SNP-SNP interactions**

**Input**

Data: Simulated Data based on selected featured indexed SNPs from screen stage

No: number of Salp's

io: interaction order

miter: Maximum iterations

**Output**

Optimal SNP combinations

**Selection Stage – SalpEpi-SO**

Step 1: Initialize the necessary parameters for SSA.

Step 2: Based on the SNPs in the feature set, arbitrarily place each salp in the population.

Step 3: Repeat the step until maximum iterations reached

Select a set of SNPs representing each salp in the solution space, and use the G-test procedure to produce a local optimal solution.

Choose a food source from the repository: F=SelectFood(repository)

For each salp (xi)

Check if i==1, then

Update the leader salp's position

else

Update the follower salp's position

end if

end for

The Salp updates the present solution by assessing new SNP combinations and comparing them to the solution space that was previously recorded.

End for

End while

**Selection Stage – SalpEpi-MO**

Steps 1 through 3 in SalpEpi-SO are the same in SalpEpi-MO.

Step 4: The Pareto optimum approach returns non-dominated SNPs.

Step 5: For $k$=1 to no. of SNPs in non-dominated solution

For $l$ =k+1 to no. of SNPs in non-dominated solution

*GGIs_pair* = G-test ($x_k$, $x_l$)

End For

SSA is a bio-inspired optimisation strategy (Mirjalili et al., 2017). The SSA mimics salps' social behaviour by gathering in a chain while sailing and searching for food in the sea. In SSA, there are two types of agents: the leader who leads the salp chain; The rest of the salps are only followers, and the leader decides where everyone goes; and supporters

who follows the leader. The description about the algorithm is exposed in (Priya and Manavalan, 2022).

## 4 Experimental results and discussion

To assess the efficacy of the suggested methodologies, two simulation models including disease loci without marginal effects (DNME) and disease with marginal effects (DME), as well as a real Alzheimer's disease dataset, are used. Sections 3.1 and 3.2 describe the datasets and evaluation measures, respectively. MATLAB R2018(b) is used to implement the suggested epistasis models. The experimental outcome of simulated models and real Alzheimer's datasets are portrayed in Section 3.3.

### *4.1 Datasets*

#### *4.1.1 Simulated dataset*

The AE-SalpEpi algorithm's efficacy is estimated using simulated datasets from several disease models. Disease models are characterised as the likelihood of having the disease from a combination of SNPs. GAMETES 2.0 is commonly used to build genotype simulation datasets. We created two-locus disease models in this study (Urbanowicz et al., 2012). Two unique types of epistatic models are produced for two-locus analysis to detect diseases: DME and DNME models. DME model describes disease's interaction and marginal impacts (Chen et al., 2019). The DNME model only exposes interaction effects and not marginal effects. Table 1 portrays the short description of the simulated dataset. The dataset's detailed description is provided in Priya and Manavalan (2022).

**Table 1** Simulated dataset description

| *Dataset* | *Simulated model category* | *No. of models* | *SNP details* | *Depiction* |
|---|---|---|---|---|
| Two-locus | Additive, multiplicative, and threshold models | Each model holds 4 variations | 2 diseased SNPs and 98 non-diseased SNPs | No. of datasets – 100 sample size – 1,600 including 800 cases and controls |
| | DNME | 10 variations of Models | | |

#### *4.1.2 Real Alzhemier's disease dataset*

The required data for research is collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The ADNI is a multi-centre longitudinal study aimed at developing clinical, imaging, genetic, and biochemical indicators for early identification and tracking of Alzheimer's disease (AD) (ADNI, n.d.). The data from ADNI 3 studies are taken to analyse the efficacy of proposed epistasis computational models. The various genotyping technologies are applied and tested over the ADNI samples. ADNI3 samples were genotyped using the Illumina Infinium Global Screening Array v2 (GSA2).

The ADNI3 Genotype data was downloaded in the binary plink format. PLINK is an open-source, command-line based whole-genome association analysis software designed to execute fundamental, large-scale analyses in a computationally efficient manner

(Purcell et al., 2007). After decompression of the dataset, PLINK 1.9 software is adapted to convert the binary files (.BED and .BAM) into .PED as well as the .MAP file format. PED and MAP files contain information about SNPs and the patients' genotype details.

The PLINK software is adapted to implement this research's quality control (QC) procedures. There are 7, 59, 993 SNPs in the ADNI3 study. Initially, 176,653 imputable variants in genetic type data were removed. Subsequently, 337,154 variants were eliminated due to MAF limits of less than 0.05. Finally, 327 people with 246,186 variants were successful in the QC procedures. Among these variants, 3467 variants were randomly chosen to analyse the efficacy of computational approaches.

Table 2 describes the QC procedures and the number of SNPs considered for testing the proposed epistasis computational approaches after QC.

**Table 2** Dataset description after QC

| *Dataset name* | *No. of SNPs* | *SNPs removed with missing genotypes* | *SNPs removed with MAF thresholds* | *Total no. of SNPs after QC* | *Number of SNPs taken for analysis* |
|---|---|---|---|---|---|
| ADNI 3 | 759,993 | 176,653 | 337,154 | 246,186 | 3,467 |

## *4.2 Performance metrics*

The suggested epistasis detection model's efficacy is assessed utilising evaluation criteria such as power. The statistical procedure of discovering true disease locus by neglecting the null hypothesis is known as power, and it is expressed as

$$Power = \frac{\#DSC}{TDSC}$$

where #DSC illustrates the quantity of data sets of are successful in the identification of disease-associated SNPs is reported among total datasets count (TDSC) (Aflakparast et al., 2014).

## *4.3 Simulation results and interpretation*

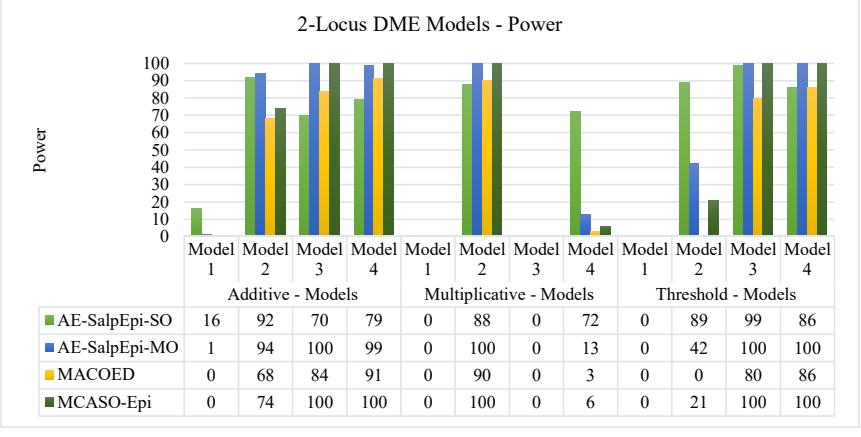
The goal of GWAS is to recognise relationships between SNPs and phenotypes. The detection of epistasis is critical for evaluating genetic disease susceptibility. AESalp-Epi's is contrasted to MACOED (Jing and Shen, 2014) and MCASO-Epi (Priya and Manavalan, 2021) techniques for epistatic identification.

### *4.3.1 Experimental analysis of DME models*

The power of AESalpEpi-SO and AESalpEpi-MO, MCASO-Epi and MACOED and for 12 DME disease models is focused in Figure 6. In additive model 1, AE-SalpEpi-SO and AE-SalpEpi-MO gained the power of 16% and 1%, respectively, whereas MACOED (Jing and Shen, 2014) and MCASO-Epi (Priya and Manavalan, 2021) did not found even a single disease-causing SNP pair. In the second additive model, the proposed strategy perform better than MACOED (Jing and Shen, 2014) and MCASO-Epi (Priya and Manavalan, 2021). Model 3 power was 100% for AE-SalpEpi-MO and MCASO-Epi (Priya and Manavalan, 2021). In contrast, MACOED (Jing and Shen, 2014) and

AE-SalpEpi-SO obtained the power of 84% and 70%, respectively. In model 4, the AE-SalpEpi-MO's power is 1% lower than MCASO-Epi. Because of the simulated setting of the dataset, none of the approaches discovers disease-causing SNP pairings from multiplicative models 1 and 3. In the second multiplicative model, MCASO-Epi (Priya and Manavalan, 2021) and AE-SalpEpi-SO had 100% power. When compared to other optimisation techniques, AE-based SO optimisation techniques such as AE-SalpEpi-SO outperform the competition in multiplicative model-4. In threshold model 3 and model 4, AE-SalpEpi-MO and MCASO-Epi (Priya and Manavalan, 2021) earned 100% power. At third threshold model, AE-SalpEpi-SO acquired a 89% power, which is better than other techniques. The evaluation of DME diseases models revealed that AE-SalpEpi-MO outperformed MCASO-Epi (Priya and Manavalan, 2021) and MACOED (Jing and Shen, 2014) in all scenarios. The comparative analysis showed that AE-SalpEpi-MO is better than MCASO-Epi (Priya and Manavalan, 2021) in 4 DME models. The efficiency of these two methods is similar in seven DME models. From the analysis, it is evidently found that feature selection based epistasis models such as AE-SalpEpi-SO and AE-SalpEpi-MO showed superior performance in DME models compared to MACOED (Jing and Shen, 2014) and MCASO-Epi (Priya and Manavalan, 2021).

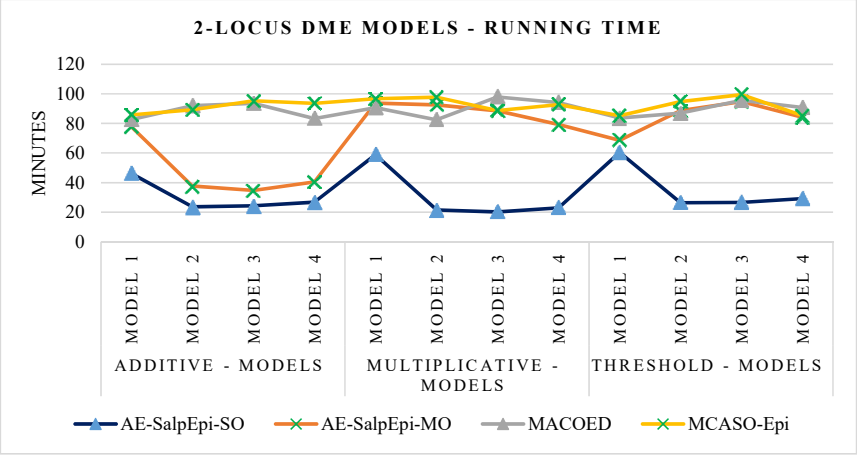
**Figure 6** Power comparison of DME models (see online version for colours)




| | Additive - Models | | | | Multiplicative - Models | | | | Threshold - Models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 1 | Model 2 | Model 3 | Model 4 | Model 1 | Model 2 | Model 3 | Model 4 |
| AE-SalpEpi-SO | 16 | 92 | 70 | 79 | 0 | 88 | 0 | 72 | 0 | 89 | 99 | 86 |
| AE-SalpEpi-MO | 1 | 94 | 100 | 99 | 0 | 100 | 0 | 13 | 0 | 42 | 100 | 100 |
| MACOED | 0 | 68 | 84 | 91 | 0 | 90 | 0 | 3 | 0 | 0 | 80 | 86 |
| MCASO-Epi | 0 | 74 | 100 | 100 | 0 | 100 | 0 | 6 | 0 | 21 | 100 | 100 |

Figure 7 illustrates the execution duration of 12 2-Locus DME simulated models. One of the proposed model AE-SalpEpi-SO, uses less duration for all 12 DME models since the model is assessed based on a SO optimisation approach in the selection stage. The MACOED (Jing and Shen, 2014) and MCASO-Epi (Priya and Manavalan, 2021) necessitates more running-time compared to AE-SalpEpi-MO and AE-SalpEpi-SO. As a result, feature selection-based techniques in DME disease models require less running time than MACOED (Jing and Shen, 2014) and MCASO-Epi (Priya and Manavalan, 2021). AE-SalpEpi-MO took the peak runtime of 94.74 minutes for the third threshold model, whereas AE-SalpEpi-MO requires the lowest runtime of 20.19 minutes for the third multiplicative model.

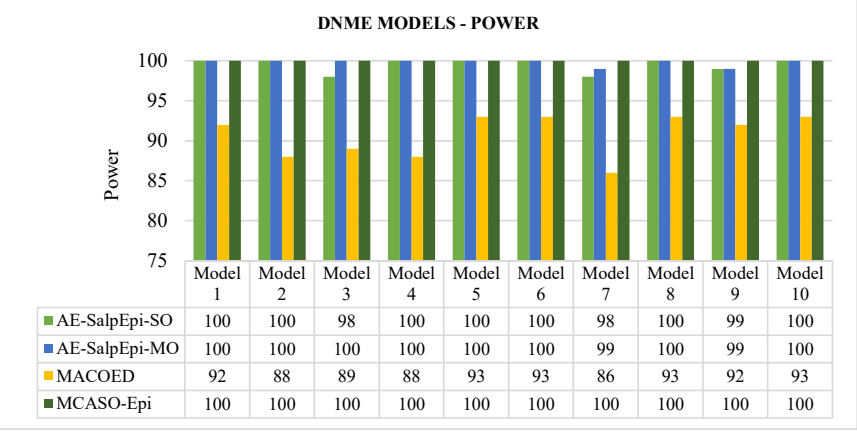
**Figure 7** Running time comparison of 2-Locus DME models (see online version for colours)




### *4.3.2 Experimental analysis of DNME models*

Figure 8 exposes the power evaluation of proposed approaches with MCASO-Epi (Priya and Manavalan, 2021) and MACOED (Jing and Shen, 2014) for the DNME models. Except for models 7 and 9, AE-SalpEpi-MO acquired 100% efficiency for all the ten DNME models. For seven different DNME models, with the exception of model 3, 7, and 9, the AE-SalpEpi-SO obtained 100 percent power. For all DNME epistasis models, the MACOED (Jing and Shen, 2014) did not produce 100% power, but the MCASO-Epi (Priya and Manavalan, 2021) produced maximum power for 10 DNME models. In DNME models 7 and 9, MCASO-Epi outperforms AE-SalpEpi-MO. The feature selection based approaches such as AE-SalpEpi-SO and AE-SalpEpi-MO are superior to MACOED (Jing and Shen, 2014) in DNME models.

**Figure 8** DNME power comparison of MSalp-Epi and MACOED (see online version for colours)




| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| AE-SalpEpi-SO | 100 | 100 | 98 | 100 | 100 | 100 | 98 | 100 | 99 | 100 |
| AE-SalpEpi-MO | 100 | 100 | 100 | 100 | 100 | 100 | 99 | 100 | 99 | 100 |
| MACOED | 92 | 88 | 89 | 88 | 93 | 93 | 86 | 93 | 92 | 93 |
| MCASO-Epi | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

The runtime of the DNME 2-locus genetic disease models is shown in Table 3. The MACOED (Jing and Shen, 2014) used the least amount of runtime compared to other approaches in DNME models. AE-SalpEpi-MO takes the longest running time for all ten variants of DNME models. The execution time of AE-SalpEpi-SO is superior to MCASO-Epi (Priya and Manavalan, 2021) and inferior to MACOED (Jing and Shen, 2014). Even though AE based feature selection approach requires the highest execution duration compared to MACOED (Jing and Shen, 2014), But it shows superior power.

**Table 3** Running time of DNME models

| *Model* / *Methods* | *DNME models – running time (minutes)* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M1* | *M2* | *M3* | *M4* | *M5* | *M6* | *M7* | *M8* | *M9* | *M10* |
| AE-SalpEpi-SO | 36.15 | 35.52 | 45.06 | 46.90 | 41.75 | 43.33 | 44.22 | 40.68 | 44.61 | 44.35 |
| AE-SalpEpi-MO | 88.60 | 83.80 | 57.01 | 74.76 | 85.58 | 77.61 | 77.61 | 55.26 | 76.72 | 62.37 |
| MACOED (Jing and Shen, 2014) | 38.35 | 37.81 | 34.58 | 35.33 | 37.67 | 36.74 | 34.37 | 35.13 | 35.44 | 35.20 |
| MCASO-Epi (Priya and Manavalan, 2021) | 39.17 | 36.33 | 46.24 | 44.51 | 47.71 | 48.79 | 48.64 | 43.36 | 47.15 | 47.16 |

## *4.4 Experimental results of real dataset*

Novel SNP interactions in Alzhemier's disease may provide key biological pathways or networks involved in disease development. Targeting these pathways or modulating the interactions between associated variants could lead to the development of more effective therapeutic interventions. Incorporating information about interactions between genetic variants allows for more precise risk stratification and personalised disease risk assessments.

The introduced epistasis detection technique such as AE-SalpEpi-SO and AE-SalpEpi-MO have been tested over the AD datasets by detecting disease-causing 2-locus interactions. From the literature study, the SNPs that are highly correlated with Alzheimer's disease are presented in Table 4, these SNPs are referred as reported SNPs. The number of 2-locus interactions discovered by the various introduced methods and the MACOED and MCASO-Epi is presented in Table 5, and the same is shown in Figure 9.

From the experiment, it is noted that the implemented epistasis approaches such as AESalpEpi-SO, AESalpEpi-MO, MCASO-Epi and MACOED have identified 1,215, 62, 15, 44 novel significant two-way interactions, respectively. The proposed approach AE-SalpEpi-SO identified 1,215 interactions, among which 49 interactions are identified disease-causing SNP interactions from the literature. Therefore, it is necessary to conduct wet-lab experiments or pathway enrichment analyses to validate the above number of novel genetic interactions discovered by the introduced methods.

Table 6 reveals the interactions identified by the proposed approaches with at least one genetic variant documented in the literature. The reported SNPs found in the 2-way associations are highlighted in Table 6. Apolipoprotein E (ApoE) is a key cholesterol carrier in the brain that aids in lipid transport and injury repair. The two polymorphisms rs429358 and rs7412 are typically present in the APOE gene. GWAS identified that the any genetic variants interact with the APOE polymorphism increase the Alzheimer's

disease risk (Zhen et al., 2017). The results show that our proposed approaches also found the interactions associated with APOE genetic variants such as rs7412 or rs429358.

**Table 4** Reported SNPs from Literature

| *S. no.* | *Reported SNPs* | *Gene name* | *References* |
|---|---|---|---|
| 1 | rs190982 | *MEF2C-S1* | Arseniou et al. (2020), Belbin et al. (2019), Bodily et al. (2016), Granados et al. (2013), Jiang et al. (2010), Marioni et al. (2018), Tang et al. (2016), Xie et al. (2018), Zhen et al. (2017) |
| 2 | rs2718058 | *Unknown Gene* | |
| 3 | rs1476679 | *Unknown Gene* | |
| 4 | rs28834970 | *PTK2B* | |
| 5 | rs9331896 | *MIR6843* | |
| 6 | rs2230805 | *ABCA1* | |
| 7 | rs2230806 | *ABCA1* | |
| 8 | rs10838725 | *CELF1* | |
| 9 | rs983392 | *Unknown Gene* | |
| 10 | rs17125944 | *FERMT2* | |
| 11 | rs10498633 | *SLC24A4* | |
| 12 | rs429358 | *APOE* | |
| 13 | rs7412 | *APOE* | |
| 14 | rs7274581 | *CASS4* | |
| 15 | rs6733839 | *LOC105373605* | |
| 16 | rs11771145 | *EPHA1-AS1* | |
| 17 | rs3865444 | *CD33* | |

**Table 5** Two-way interactions results in ADNI3 datasets

| *Methods* | *No. of interactions identified* | *No. of reported SNPs identified* |
|---|---|---|
| AE-SalpEpi-SO | 1,188 | 30 |
| AE-SalpEpi-MO | 64 | 9 |
| MACOED | 15 | 5 |
| MCASO-Epi | 44 | 6 |

**Figure 9** 2-Locus analysis of proposed approaches in AD dataset (see online version for colours)

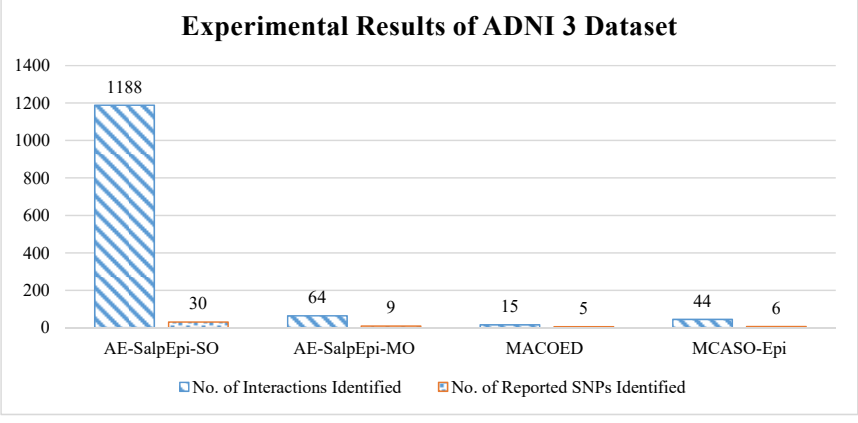


**Table 6** Topmost reported SNP variants associated 2-way interactions in AD dataset

| *SNP1* | *SNP2* | *P_value* | *Method* |
|---|---|---|---|
| rs11121600 (*) | **rs200458** (*DNAJC11*) | 0.001953 | AE-SalpEpi-SO |
| rs111530107 (*LOC105373903*) | **rs2718058** (*) | 0.002974 | |
| **rs10498633** (*SLC24A4*) | rs3935435 (*PLEKHG5*) | 0.008728 | |
| **rs10838725** (*CELF1*) | rs1452632 (*LOC124904510*) | 0.010744 | |
| rs10910045 (*SKI*) | **rs7412** (*APOE*) | 0.01278 | |
| **rs10498633** (*SLC24A4*) | rs771071 (*) | 0.014539 | |
| GSA-rs80072814 (*CRACDL*) | **rs28834970** (*PTK2B*) | 0.015918 | |
| **rs429358** (*APOE*) | **rs3865444** (*CD33*) | 0.0166 | |
| rs10036748 (*TNIP1*) | **rs28834970** (*PTK2B*) | 0.018268 | |
| **rs10498633** (SLC24A4) | **rs200458** (*DNAJC11*) | 0.022329 | |
| **rs28834970** (PTK2B) | rs6664880 (*) | 0.022788 | |
| rs17040590 (*) | **rs28834970** (*PTK2B*) | 0.023708 | |
| **rs3865444** (*CD33*) | rs763173 (*CAMTA1*) | 0.02559 | |
| **rs28834970** *(PTK2B*) | GSA-rs4908448 (*CAMTA1*) | 0.02584 | |
| **rs28834970** (*PTK2B*) | rs499416 (*) | 0.029486 | |
| rs7103026 (*OR1S1*) | **rs190982** (*MEF2C-S1*) | 0.033297 | |
| **rs10838725** (*CELF1*) | rs771071 (*) | 0.033683 | |
| **rs7412** (*APOE*) | rs6664880 (*) | 0.035734 | |
| **rs7103026** (*OR1S1*) | rs10498633 (SLC24A4) | 0.03675 | |
| **rs200458** (*DNAJC11*) | rs706007 (*CASZ1*) | 0.039278 | |
| **rs7412** (*APOE*) | rs763173 (*CAMTA1*) | 0.040114 | |
| **rs3865444** (*CD33*) | rs3796056 (*CNRIP1*) | 0.040392 | |
| **rs10838725** (*CELF1*) | rs2802720 (*AJAP1*) | 0.041643 | |
| **rs7412** (*APOE*) | rs7522140 (*) | 0.0434 | |
| **rs200458** (*DNAJC11*) | rs495768 (*) | 0.04403 | |
| **rs1476679** (*) | rs6664880 (*) | 0.0456 | |
| **rs10498633** (*SLC24A4*) | rs11121600 (*) | 0.046157 | |
| rs4548406 (*) | **rs200458** (*DNAJC11*) | 0.04701 | |
| rs771071 (*) | **rs200458** (*DNAJC11*) | 0.048342 | |
| rs16845368 (*) | **rs28834970** (*PTK2B*) | 0.049418 | |
| **rs200458** (*DNAJC11*) | rs11121600 (*) | 0.001953 | AE-SalpEpi-MO |
| rs12132314(*) | **rs7412** (*APOE*) | 0.006824 | |
| **rs200458** (DNAJC11) | **rs2230806** (*ABCA1*) | 0.010688 | |
| **rs200458** (DNAJC11) | rs873982 (*LOC124903830*) | 0.02823 | |
| **rs10838725** (*CELF1*) | rs771071 (*) | 0.033683 | |
| **rs2230806** (*ABCA1*) | rs12132314 (*) | 0.042218 | |
| **rs200458** (*DNAJC11*) | rs4548406 (*) | 0.04701 | |
| **rs200458** (*DNAJC11*) | rs771071 (*) | 0.048342 | |
| **rs200458** (*DNAJC11*) | **rs7412** (*APOE*) | 0.048873 | |

Note: *Gene name is unknown.

The main genetic risk factor for LOAD is the APOE gene, which is found on chromosome 19q13.2 (Giri et al., 2016). The primary carrier of cholesterol in the brain is the APOE protein. The APOE gene contains SNPs rs429358 and rs7412, whose combined genotypes determine the APOE allelic status, which is recognised to be the strongest genetic variant associated with late onset Alzheimer's disease (LOAD) (Corder et al., 1993). The results show that our proposed approaches also found the interactions associated with APOE genetic variants such as rs7412 or rs429358. APOE and its genetic variants interacts with the SNPs rs10910045 from the gene SKI, rs3865444 (*CD33*), rs763173 (*CAMTA1*), rs200458 (*DNAJC11*), these interactions influence the Alzheimer's risk.

CD33 mediates interactions between cells, promotes clathrin-independent endocytosis, suppresses immune cell functions, and controls cell proliferation and survival by inducing apoptosis (Jandus et al., 2011). Recent GWAS have clearly pointed out that *CD33* is a genetic locus that has a strong association with LOAD (Hollingworth et al., 2011). Two SNPs such as rs3865444 and rs3826656, located in the *CD33* gene cluster have been linked to LOAD. The minor form of the *CD33* SNP rs3865444 was linked to lower levels of both *CD33* and amyloid plaque in the brains of people with AD. AE-SalpEpi-SO found 3 disease-causing interactions from the SNP variant rs3865444 in gene *CD33*.

The central nervous system is characterised by a high level of PTK2B expression. Alzheimer's risk was linked to an SNP in PTK2B called rs28834970 (Jiao et al., 2015). The proposed AE-SalpEpi-SO discovered that the gene PTK2B and its variant rs28834970 were linked to 8 SNP–SNP interactions.

*CELF1* is a protein that is part of the *CELF/BRUNOL* family. It is on chromosome *11p11* and is involved in regulating pre-mRNA splicing, mRNA translation, and mRNA editing (Jiao et al., 2015). The CELF1 intron SNP rs10838725 was found to increase susceptibility to LOAD (Beck et al., 2014). The proposed approaches found that *CELF1* (rs10838725) interact with the genetic variants rs1452632 (*LOC124904510*), rs771071, rs2802720 from the gene *AJAP1* and increase the disease risk.

The brain development, hair colour, and skin pigmentation are all connected to *SLC24A4*. LOAD risk is related to SNP rs10498633 in SLC24A4's intron (Larsson et al., 2011). Yu et al. (2015) discovered an association between *SLC24A4* and methylation, and DNA methylation in the brain contributes to the pathogenesis of Alzheimer's disease. Five SNP-SNP interactions involving the gene *SLC24A4* and its variant rs10498633 are found by one of the suggested approach AE-SalpEpi-SO.

## 5 Conclusions

The genetic factors and their interaction effects are speculated to contribute to human diseases. In GWAS, the detection of potential genetic interactions is a crucial problem. Alzheimer's and Parkinson's disease are both caused by deteriorated brain cells. Both disorders can lead to dementia and other mental health problems, as well as insomnia. In this paper, a two-step methodology is implemented to detect epistasis effects in Alzheimer's disease, referred to as AE-SalpEpi-SO and AE-SalpEpi-MO. The AE-SalpEpi uses an autoencoder technique to choose the featured SNPs in the screening stage, while Salp optimisation discovers disease-correlated SNPs in the selection step. In both DME and DNME models, experimental results showed that AE-SalpEpi-SO and

AE-SalpEpi-MO performed better than MCASO-Epi and MACOED. Several genes have been previously reported to be directly related to Alzheimer's disease. The proposed approach identified many interactions from the reported genes such as *APOE*, *CD33*, *DNAJC11*, *ABCA1*, *SLC24A4*, and *PTK2B*. The proposed paradigm also identified many novel SNP interactions, which need further validation, including pathway enrichment analysis. In the future, this study could be expanded to include multi-locus interaction analysis for the pathogenesis of Alzheimer's disease risk.

## Acknowledgements

Data used in the preparation of this article were obtained in part from the ADNI database (http://www.loni.ucla.edu/ADNI). Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organisation is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## References

'2020 Alzheimer's disease facts and figures' (2020) *Alzheimer's & Dementia*, Vol. 16, No. 3, pp.391–460, John Wiley & Sons, Ltd.

Alzheimer's Disease Neuroimaging Initiative (ADNI) (n.d.) 'ADNI', *Alzheimer's Disease Neuroimaging Initiative* [online] https://adni.loni.usc.edu/ (accessed 20 June 2022).

Aflakparast, M., Salimi, H., Gerami, A., Dubé, M.P., Visweswaran, S. and Masoudi-Nejad, A. (2014) 'Cuckoo search epistasis: a new method for exploring significant genetic interactions', *Heredity*, Vol. 112, No. 6, pp.666–674.

Anter, A.M., Azar, A.T., Hassanien, A.E., El-Bendary, N. and ElSoud, M.A. (2013) 'Automatic computer aided segmentation for liver and hepatic lesions using hybrid segmentations techniques', *2013 Federated Conference on Computer Science and Information Systems*, pp.193–198.

Arseniou, S., Siokas, V., Aloizou, A., Stamati, P., Mentis, A-F., Tsouris, Z., Dastamani, M. et al. (2020) 'SLC2A3 rs12842 polymorphism and risk for Alzheimer's disease', *Neurological Research*, Vol. 42, No. 10, pp.1–9.

Assareh, A., Volkert, L.G. and Li, J. (2012) 'Feature selections using AdaBoost: application in gene-gene interaction detection', *Proceedings – 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops, BIBMW 2012*, pp.831–837.

Azar, A.T., Banu, P.K.N. and Inbarani, H.H. (2013) 'PSORR – an unsupervised feature selection technique for fetal heart rate', *2013 5th International Conference on Modelling, Identification and Control (ICMIC)*, pp.60–65.

Beck, T.N., Nicolas, E., Kopp, M.C. and Golemis, E.A. (2014) 'Adaptors for disorders of the brain? The cancer signaling proteins NEDD9, CASS4, and PTK2B in Alzheimer's disease', *Oncoscience*, Vol. 1, No. 7, pp.486–503, USA.

Belbin, O., Morgan, K., Medway, C., Warden, D., Cortina-Borja, M., Duijn, C., Adams, H. et al. (2019) 'The epistasis project: a multi-cohort study of the effects of BDNF, DBH, and SORT1 epistasis on Alzheimer's disease risk', *Journal of Alzheimer's Disease*, Vol. 68, No. 4, pp.1535–1547.

Bodily, P.M., Fujimoto, M.S., Page, J.T., Clement, M.J., Ebbert, M.T.W., Ridge, P.G. and Initiative, the ADN (2016) 'A novel approach for multi-SNP GWAS and its application in Alzheimer's disease', *BMC Bioinformatics*, Vol. 17, No. 7, p.268.

*BrainFacts* (n.d.) 'Alzheimer disease' [online] https://www.brainfacts.org/diseases-and-disorders/neurological-disorders-az/diseases-a-to-z-from-ninds/alzheimer-s-disease%0A%0A (accessed 20 June 2022).

Chattopadhyay, A. and Lu, T-P. (2019) 'Gene-gene interaction: the curse of dimensionality', *Annals of Translational Medicine*, Vol. 7, No. 24, p.813, AME Publishing Company.

Chelliah, B.J., Malik, M.M., Kumar, A., Singh, N. and Regin, R. (2023) 'Similarity-based optimised and adaptive adversarial attack on image classification using neural network', *International Journal of Intelligent Engineering Informatics*, Vol. 11, No. 1, pp.71–95, Inderscience Publishers.

Chen, Q., Zhang, X. and Zhang, R. (2019) 'Privacy-preserving decision tree for epistasis detection', *Cybersecurity*, Vol. 2, No. 1, https://doi.org/10.1186/s42400-019-0025-z.

Chen, R-C., Dewi, C., Huang, S-W. and Caraka, R.E. (2020) 'Selecting critical features for data classification based on machine learning methods', *Journal of Big Data*, Vol. 7, No. 1, p.52.

Chen, Y., Xu, F., Pian, C., Xu, M., Kong, L., Fang, J., Li, Z. et al. (2021) 'Epimoga: an epistasis detection method based on a multi-objective genetic algorithm', *Genes*, Vol. 12, No. 2, pp.1–18.

Chifu, V.R., Chifu, E.S., Pop, C.B., Salomie, I. and Lupu, M. (2019) 'Evolutionary-based method for risk stratification of diabetic patients', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 1, p.37.

Consortium, W.T.C.C. (2007) 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature*, Vol. 447, No. 7145, pp.661–678.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D. et al. (1993) 'Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families', *Science*, Vol. 261, No. 5123, pp.921–923, New York, NY, USA.

Elshazly, H.I., Elkorany, A.M., Hassanien, A.E. and Azar, A.T. (2013) 'Ensemble classifiers for biomedical data: performance evaluation', *2013 8th International Conference on Computer Engineering & Systems (ICCES)*, pp.184–189.

Emary, E., Zawbaa, H.M., Hassanien, A.E., Schaefer, G. and Azar, A.T. (2014) 'Retinal blood vessel segmentation using bee colony optimisation and pattern search', *2014 International Joint Conference on Neural Networks (IJCNN)*, pp.1001–1006.

Ganesan, J., Inbarani, H. and Azar, A.T. (2014) 'Hybrid tolerance rough set: PSO based supervised feature selection for digital mammogram images', *International Journal of Fuzzy System Applications (IJFSA)*, Vol. 3, No. 4, pp.15–30.

Giri, M., Zhang, M. and Lü, Y. (2016) 'Genes associated with Alzheimer's disease: an overview and current status', *Clinical Interventions in Aging*, Vol. 11, No. 17, pp.665–681, New Zealand.

Gonçalves, F., Santander-Jiménez, S., Sousa, L., Granado-Criado, J.M. and Ilic, A. (2020) 'Parallel evolutionary computation for multiobjective gene interaction analysis', *Journal of Computational Science*, Vol. 40, https://doi.org/10.1016/j.jocs.2019.101068.

Granados, E.A.O., Vásquez, L.F.N. and Granados, H.A. (2013) 'Characterizing genetic interactions using a machine learning approach in Colombian patients with Alzheimer's disease', *Proceedings – 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013*, pp.1–2.

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J-C., Carrasquillo, M.M., Abraham, R. et al. (2011) 'Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease', *Nature Genetics*, Vol. 43, No. 5, pp.429–435, USA.

Humaidi, A.J., Najem, H.T., Al-Dujaili, A.Q., Pereira, D.A., Ibraheem, I.K. and Azar, A.T. (2021) 'Social spider optimization algorithm for tuning parameters in PD-like interval type-2 fuzzy logic controller applied to a parallel robot', *Measurement and Control*, Vol. 54, Nos. 3–4, pp.303–323, SAGE Publications Ltd.

Jandus, C., Simon, H-U. and Gunten, S. (2011) 'Targeting Siglecs-A novel pharmacological strategy for immuno- and glycotherapy', *Biochemical Pharmacology*, Vol. 82, No. 4, pp.323–332.

Jiang, X., Neapolitan, R.E., Barmada, M.M., Visweswaran, S. and Cooper, G.F. (2010) 'A fast algorithm for learning epistatic genomic relationships', *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium*, Vol. 2010, pp.341–345.

Jiao, B., Liu, X., Zhou, L., Wang, M.H., Zhou, Y., Xiao, T., Zhang, W. et al. (2015) 'Polygenic analysis of late-onset Alzheimer's disease from Mainland China', *PLOS ONE, Public Library of Science*, Vol. 10, No. 12, p.e0144898.

Jing, P-J. and Shen, H-B. (2014) 'MACOED: a multi-objective ant colony optimization algorithm for SNP epistasis detection in genome-wide association studies', *Bioinformatics*, Vol. 31, Oxford, England, https://doi.org/10.1093/bioinformatics/btu702.

Kumar, M. and Sharma, A. (2023) 'Progressive global best artificial bee colony algorithm for wind farm layout optimisation problem', *International Journal of Intelligent Engineering Informatics*, Vol. 11, No. 3, pp.272–297, Inderscience Publishers.

Kumar, S.S., Inbarani, H.H., Azar, A.T. and Hassanien, A.A. (2015) 'Rough set-based meta-heuristic clustering approach for social e-learning systems', *International Journal of Intelligent Engineering Informatics*, Vol. 3, No. 1, pp.23–41, Inderscience Publishers.

Larsson, M., Duffy, D.L., Zhu, G., Liu, J.Z., Macgregor, S., McRae, A.F., Wright, M.J. et al. (2011) 'GWAS findings for human iris patterns: associations with variants in genes that influence normal neuronal pattern development', *American Journal of Human Genetics*, Vol. 89, No. 2, pp.334–343, USA.

Manavalan, R. and Priya, S. (2021) 'Genetic interactions effects for cancer disease identification using computational models: a review', *Medical & Biological Engineering & Computing*, Vol. 59, No. 4, pp.733–758, USA.

Marioni, R.E., Harris, S.E., Zhang, Q., McRae, A.F., Hagenaars, S.P., Hill, W.D., Davies, G. et al. (2018) 'GWAS on family history of Alzheimer's disease', *Translational Psychiatry*, Vol. 8, No. 1, p.99.

Marzouki, B., Belkahla Driss, O. and Ghédira, K. (2018) 'Multi-agent model based on combination of chemical reaction optimisation metaheuristic with tabu search for flexible job shop scheduling problem', *International Journal of Intelligent Engineering Informatics*, Vol. 6, No. 3, p.242.

Mayo Clinic (2020) 'Alzheimer's disease – symptoms and causes' [online] https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447 (accessed 20 June 2022).

Mirjalili, S., Gandomi, A., Mirjalili, S.Z., Saremi, S., Faris, H. and Mirjalili, S. (2017) 'Salp swarm algorithm: a bio-inspired optimizer for engineering design problems', *Advances in Engineering Software*, Vol. 114, pp.1–29.

Najm, A.A., Ibraheem, I.K., Azar, A.T. and Humaidi, A.J. (2020) 'Genetic optimization-based consensus control of multi-agent 6-DoF UAV system', *Sensors*, https://doi.org/10.3390/s20123576.

Parenti, I., Rabaneda, L.G., Schoen, H. and Novarino, G. (2020) 'Neurodevelopmental disorders: from genetics to functional pathways', *Trends in Neurosciences*, Vol. 43, No. 8, pp.608–621.

Priya, S. and Manavalan, R. (2021) 'Multi-objective chaotic atom search optimization for epistasis detection in genome-wide association studies', in Singh, D., Awasthi, A.K., Zelinka, I. and Deep, K. (Eds.): *BT – Proceedings of International Conference on Scientific and Natural Computing*, Springer Singapore, Singapore, pp.11–22.

Priya, S. and Manavalan, R. (2022) 'MSalp-Epi: multi-objective salp optimisation for epistasis detection in genome-wide association studies', *International Journal of Intelligent Engineering Informatics*, Vol. 10, No. 1, pp.1–28, Inderscience Publishers.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J. et al. (2007) 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *American Journal of Human Genetics*, Vol. 81, No. 3, pp.559–575.

Sapin, E., Keedwell, E. and Frayling, T. (2015) 'An ant colony optimization and tabu list approach to the detection of gene-gene interactions in genome-wide association studies [research frontier]', *IEEE Computational Intelligence Magazine*, Vol. 10, No. 4, pp.1–21.

Sayed, G.I. and Hassanien, A.E. (2024) 'Explainable AI and sand cat optimisation algorithm for water quality classification', *International Journal of Intelligent Engineering Informatics*, Vol. 12, No. 1, pp.60–84, Inderscience Publishers.

Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X. and Chen, X. (2016) 'A sparse auto-encoder-based deep neural network approach for induction motor faults classification', *Measurement: Journal of the International Measurement Confederation*, Vol. 89, pp.171–178, Elsevier Ltd.

Surya, S. and Muthukumaravel, A. (2023) 'Efficient feature extraction on mammogram images using enhanced grey level co-occurrence matrix', *International Journal of Intelligent Engineering Informatics*, Vol. 11, No. 1, pp.35–53, Inderscience Publishers.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. and Meyre, D. (2019) 'Benefits and limitations of genome-wide association studies', *Nature Reviews Genetics*, Vol. 20, No. 8, pp.467–484.

Tang, S-S., Wang, H-F., Zhang, W., Kong, L-L., Zheng, Z-J., Tan, M-S., Tan, C-C. et al. (2016) 'MEF2C rs190982 polymorphism with late-onset Alzheimer's disease in Han Chinese: a replication study and meta-analyses', *Oncotarget, Impact Journals LLC*, Vol. 7, No. 26, pp.39136–39142.

Tripathi, N. and Patheja, P.S. (2022) 'Segmentation free text recognition for overlapping characters using spectral features and bidirectional recurrent wavelet neural network', *International Journal of Intelligent Engineering Informatics*, Vol. 10, No. 6, pp.464–483, Inderscience Publishers.

Tuo, S., Zhang, J., Yuan, X., Zhang, Y. and Liu, Z. (2016) 'FHSA-SED: two-locus model detection for genome-wide association study with harmony search algorithm', *PLoS ONE*, Vol. 11, No. 3, pp.1–27.

Ucsfhealth.org (n.d.) *Neurology and Neurosurgery* [online] http://www.ucsfhealth.org/conditions/neurological_disorders/ (accessed 20 June 2022).

Urbanowicz, R.J., Kiralis, J., Sinnott-Armstrong, N.A., Heberling, T., Fisher, J.M. and Moore, J.H. (2012) 'GAMETES: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures', *BioData Mining*, Vol. 5, No. 1, https://doi.org/10.1186/1756-0381-5-16.

Wang, L. and Nikouei Mehr, M. (2019) 'An optimization approach to epistasis detection', *European Journal of Operational Research*, Vol. 274, No. 3, pp.1069–1076, Elsevier B.V.

Wu, T., Chen, Y., Hastie, T., Sobel, E. and Lange, K. (2009) 'Genomewide association analysis by lasso penalized logistic regression', *Bioinformatics*, Vol. 25, pp.714–721, Oxford, England.

Xie, T., Akbar, S., Stathopoulou, M., Oster, T., Masson, C., Yen, F. and Visvikis-Siest, S. (2018) 'Epistatic interaction of apolipoprotein E and lipolysis-stimulated lipoprotein receptor genetic variants is associated with Alzheimer's disease'.

Yu, L., Chibnik, L.B., Srivastava, G.P., Pochet, N., Yang, J., Xu, J., Kozubek, J. et al. (2015) 'Association of Brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease', *JAMA Neurology*, Vol. 72, No. 1, pp.15–24, USA.

Zhen, J., Huang, X., Van Halm-Lutterodt, N., Dong, S., Ma, W., Xiao, R. and Yuan, L. (2017) 'ApoE rs429358 and rs7412 polymorphism and gender differences of serum lipid profile and cognition in aging Chinese population', *Frontiers in Aging Neuroscience, Frontiers Media S.A.*, Vol. 9, p.248.