

Alatrany, AS, Hussain, A, Mustafina, J and Al-Jumeily, D

Machine learning approaches and applications in genome wide association study for Alzheimer's Disease: A systematic review

<https://researchonline.ljmu.ac.uk/id/eprint/17173/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Alatrany, AS ORCID logoORCID: <https://orcid.org/0000-0002-4504-1506>,
Hussain, A ORCID logoORCID: <https://orcid.org/0000-0001-8413-0045>,
Mustafina, J and Al-Jumeily, D ORCID logoORCID: <https://orcid.org/0000-0002-9170-0568> (2022) Machine learning approaches and applications in**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

Received May 23, 2022, accepted June 5, 2022, date of publication June 13, 2022, date of current version June 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3182543

Machine Learning Approaches and Applications in Genome Wide Association Study for Alzheimer's Disease: A Systematic Review

ABBAS SAAD ALATRANY^{1,2}, ABIR JAAFAR HUSSAIN^{1,3}, (Member, IEEE),
JAMILA MUSTAFINA⁴, AND DHIYA AL-JUMEILY¹, (Senior Member, IEEE)

¹School of Computer Science and Mathematics, Liverpool John Moores University, Liverpool L3 3AF, U.K.

²School of Engineering, University of Information Technology and Communications, Baghdad 00964, Iraq

³Department of Electrical Engineering, University of Sharjah, Sharjah, United Arab Emirates

⁴Naberezhnochnelnsky Institute, Kazan Federal University, 420008 Kazan, Russia

Corresponding author: Abbas Saad Alatrany (a.s.alatrany@2020.ljmu.ac.uk)

ABSTRACT Machine learning algorithms have been used for detection (and possibly) prediction of Alzheimer's disease using genotype information, with the potential to enhance the outcome prediction. However, detailed research about the analysis and the detection of Alzheimer's disease using genetic data is still in its primitive stage. The aim of this paper was to evaluate the scientific literature on the use of various machine learning approaches for the prediction of Alzheimer's disease based solely on genetic data. To identify gaps in the literature, critically appraise the reporting and methods of the algorithms, and provide the foundation for a wider research programme focused on developing novel machine learning based predictive algorithms in Alzheimer's disease. A systematic review of quantitative studies was conducted using three search engines (PubMed, Web of Science and Scopus), and included studies between 1st of January 2010 and 31st December 2021. Keywords used were 'Alzheimer's disease(s)', 'GWAS', 'Artificial intelligence' and their synonyms. After applying the inclusion/exclusion criteria, 24 studies were included. Machine learning methods in the reviewed papers performed in a wide range of ways (0.59 to 0.98 AUC). The main findings showed that high risk of bias in the analysis can be linked to feature selection, hyperparameter search and validation methods.

INDEX TERMS Machine learning, systematic review, Alzheimer's disease, GWAS, SNP, genome wide, review.

I. INTRODUCTION

One of the most significant scientific issues in the human genome is the study of genetic variants connected to complex illnesses. The bulk of genome-wide association studies (GWAS) [1] attempt to identify genetic variations that may be connected to complicated illnesses. Single nucleotide polymorphisms (SNP) are known to be the most prevalent genetic variations, with around 10 million SNPs in the human genome [2].

A single nucleotide site is one in which a significant fraction of the population has exactly two (of four) unique nucleotides. There are two known ways in which SNPs play a significant role in the disorders' complications. First, by changing the protein structure. Second, via altering

the protein quantity. This process is referred to as SNP functionality.

Genetic association research attempted to explore genetic risk factors by identifying statistical connections between genotypes and phenotypes (disease of interest). The most popular method for determining the genetic connections of complicated disorders is to conduct case-control studies in unrelated individuals.

Machine learning (ML) has emerged since a decade an alternative for genetic predictions and has shown prominence especially after the developments in deep learning [3]. This emerged in line with scaling-up of datasets and computational capacity. In statistical genetics, where the effects of a large number of factors on an outcome were difficult to anticipate, ML techniques have been intriguing because of their capability to operate in high dimensions and identify relations across genes [4]. There have also been more requests

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang¹.

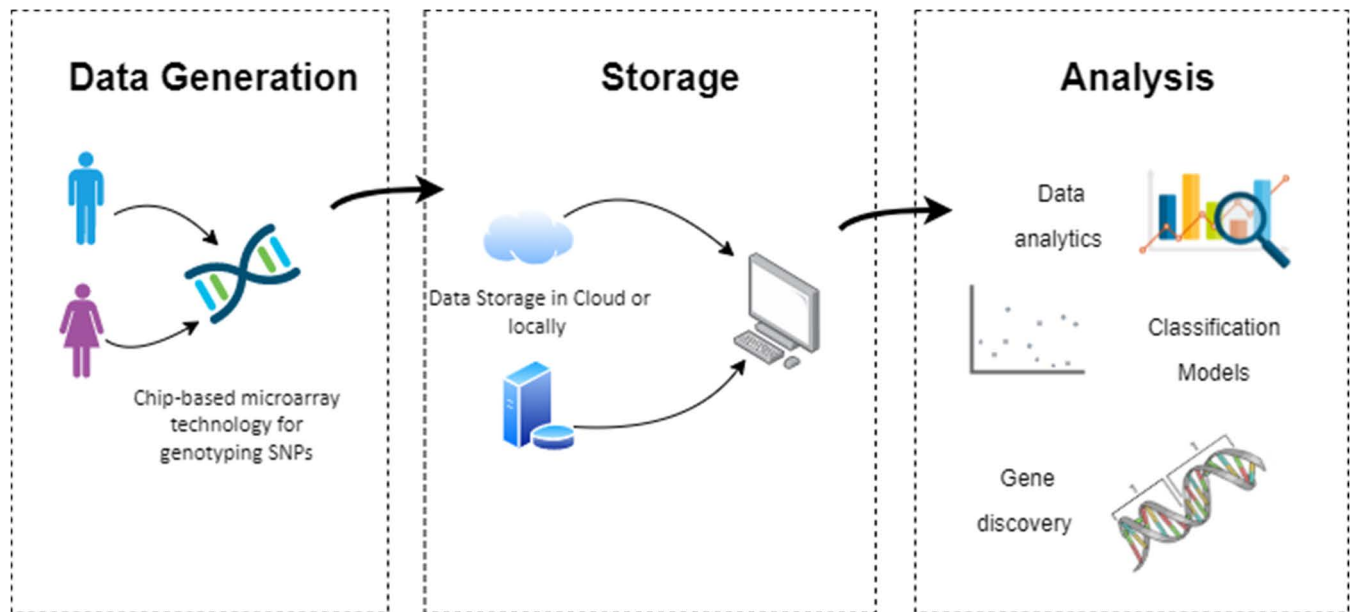


FIGURE 1. Big data in GWAS. A pipeline is depicted schematically, beginning with SNPs generated by genotypes platforms, and ending with the development of analytical models using various big data analytics solutions and tools.

to employ ML to handle the complexity of diseases such as Alzheimer's disease [5]. However, the accuracy of machine learning approaches in predicting Alzheimer's disease using genetics is still vague, and a new review of prediction models across a variety of outcomes and predictors discovered that logistic regression (LR) provided high accuracy, and hence the use of machine learning in this research field is questionable [6].

Various reviews have examined genome-wide association research and genetic prediction in relation to ML. The first conducted by Bracher-Smith et al [7] that examined machine learning algorithms for identifying mental health diseases based only on genetic information. Second, Madhukar et al. [8] discuss bioinformatics ideas for leveraging sequencing data to predict sample-specific medication susceptibility. Third, Upstill-Goddard et al. [9] provided a review of machine learning methods in genetic epidemiology for detecting gene-gene interactions. The most essential machine learning approaches and the circumstances that must be addressed when applying these algorithms to genomic challenges were discussed in [10]. The goal is to find and analyze GWAS concerns that require computational approaches instead of or in addition to biostatistical methods [11]. Data mining and machine learning computational methodologies, as well as bioinformatics methods for embedding pre-existing biological knowledge into data analysis algorithms, were the focus of other research work carried out by Wu and Zhao [10]. A review of illness prediction based on single-nucleotide polymorphisms has been undertaken by Ho et al. [12]. A recent systematic review of ML algorithms in SNP data of Alzheimer's disease is shown by Rowe et al. [13]. However, the main limitation of the research paper provided

by Rowe et al. [13] is the fact that they have utilized machine learning as a keyword for their search and presentation instead of investigating specific ML techniques. Hence, their review paper does not provide sufficient details on the contribution of various research in relation to the use of ML for the analysis of SNP data of AD. Their inclusion criteria involved studies which combined SNP data with other forms of data. As far as we are aware, there have been no reviews of studies which have developed ML models to predict AD outcomes from SNP data specifically.

As a result, in our review study, we have looked at extensive literature review for the ability of machine learning (ML) techniques to predict Alzheimer's disease risk using genetic based on the Genome Wide Associations study. The goal of the review is to identify gaps in the literature, critically appraised the reporting and methods of the algorithms and provide the foundation for a wider research program focused on developing novel machine learning based predictive algorithms in AD. It should be noted that studies in which models were also tested on simulated datasets or other chronic diseases alongside Alzheimer's disease are also considered in this comprehensive review paper. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) criteria were followed for writing this review.

II. THEORETICAL BACKGROUND

A. BIG DATA

Big data is a collection of big, structured or unstructured data sets that a typical database system struggle to manage. "Big data" refers to the tools and procedures that enable an organisation to produce, utilise, and store huge volumes of data with storage facilities [14]. Big data is usually defined by five

characteristics that, within the context of genetics, include: volume, variety, velocity, veracity and variability. In terms of volume, GWAS requires the genotyping of thousands or millions of genotypes for each participant which in total for a complete GWAS produce a massive amount of data. For variety, a mixture of data types needs to be stored and handled that is made across several files for genotype data and individuals' information. For velocity, the fast advancement of GWAS has been aided by the availability of genotyping technologies. These genotyping technologies were created expressly for assaying more than one million SNPs, such as sequencing the whole human genome in one day [15]. For veracity, errors in the genotyping process can result in data quality concerns that are difficult to spot, which can have a significant impact on a study's biological results [16]. For variability: GWAS dataset can be stored in different formats. However, it is preferable to save the data in a binary formatted file, which results in a large decrease in file size and a significant increase in computing performance. Figure 1 shows a graphical representation of the big data in GWAS.

B. MACHINE LEARNING

Machine Learning (ML) simulates human learning by allowing computers to recognise and gain knowledge from the actual world, as well as enhance performance on particular tasks depending on this new information. ML was explored as a separate discipline in the 1990s [17], despite the fact that the earliest notions of ML (with different terminologies) were developed in the 1950s. Apart from computer science, ML algorithms have been applied in a variety of fields, including business [18], advertising [19], and medicine [20].

Learning is the process of gaining information, because of their ability to reason, humans naturally learn from their experiences. Conventional computers, on the other hand, do not learn by thinking rather by following algorithms. There are several machine learning algorithms presented in the literature nowadays. They may be divided into groups based on how they approach the learning process, supervised, unsupervised, semi supervised, and reinforcement learning are the four primary classes [21]. Figure 2 shows ML types: a) Supervised learning works with data that has been labelled; in the instance of GWAS, the SNPs data is entered as inputs with corresponding labels, the ML model will automatically generate patterns and produce predictions for new unseen inputs; b) Unsupervised learning learns pattern from unlabeled data inputs. in the area of genetics, unsupervised learning can be used to cluster genes that have a common characteristic; c) Semi-supervised learning, the model accepts labelled and unlabeled datapoints; d) reinforcement learning is to feed the model with unlabeled data, then the model generates predictions, which can be approved by providing feedback to the model on whether that prediction was correct or not.

With the growth in processor speed and memory size, machine learning has become increasingly popular. As a

result, the discipline currently contains a wide variety of algorithms that learn, draw conclusions, or infer facts through mathematical or statistical analysis [22]. The number of scholarly articles proposing modifications or combinations of machine learning algorithms continues to rise [23], [24]. As a result, machine learning algorithms have been classified according to their intended use.

1) ARTIFICIAL NEURAL NETWORK

Artificial neural network (ANN) is a densely connected network of hundreds or even millions of fundamental processing nodes loosely modelled after the human brain. The vast majority of today's artificial neural networks are "feed-forward," meaning that information only goes one way through them and they are organized into layers of nodes. However, there exist other types of ANN that accept feedback connections. These are mainly known as recurrent neural networks. They are characterized by their "memory," which allows them to impact current input and output by using knowledge from previous inputs. While typical deep neural networks presume that inputs and outputs are independent of one another, recurrent neural networks' output is reliant on the sequence's prior components. While future occurrences may be useful in establishing the outcome of a series.

Nodes in a layer can be fully or partially connected to the nodes of a previous layer from which it obtains the data. Similarly, the nodes are connected and send data to the nodes of the succeeding layer. Figure 3 illustrates an ANN representation.

The process of training an ANN starts by randomly setting the values of weights and thresholds. The input layer receives the training data, which is subsequently multiplied and combined in different complex ways until it reaches the output layer. The values of weights are continually adjusted throughout the training process.

Initially, the perceptron of a very basic artificial neural network consisted of only two inputs and one output [25]. This setup enables the creation of a basic classifier that can discriminate between two groups. ANN then evolved into a Multilayer Perceptron, which consists of three layers: input, hidden and output. This development has allowed us to solve more complex non-linear problems [26].

Due to the increase in the volume of data and the complexity of the problems associated therewith, a new subset of machine learning algorithms was established known as deep learning. Deep learning (DL) excelled with its ability to automatically learn characteristics from data and the relationships between data points [27].

An ANN architecture with numerous hidden layers and neurons is the basic architecture in deep learning. Various designs have been suggested, and many of them have found success in various applications including the analysis of genetic data. Convolutional neural networks are deep learning structures inspired by human visual cortex models that have been widely used in image recognition. Recurrent artificial neural networks, which imbue neurons with dynamic behavior, have emerged as the most popular approach

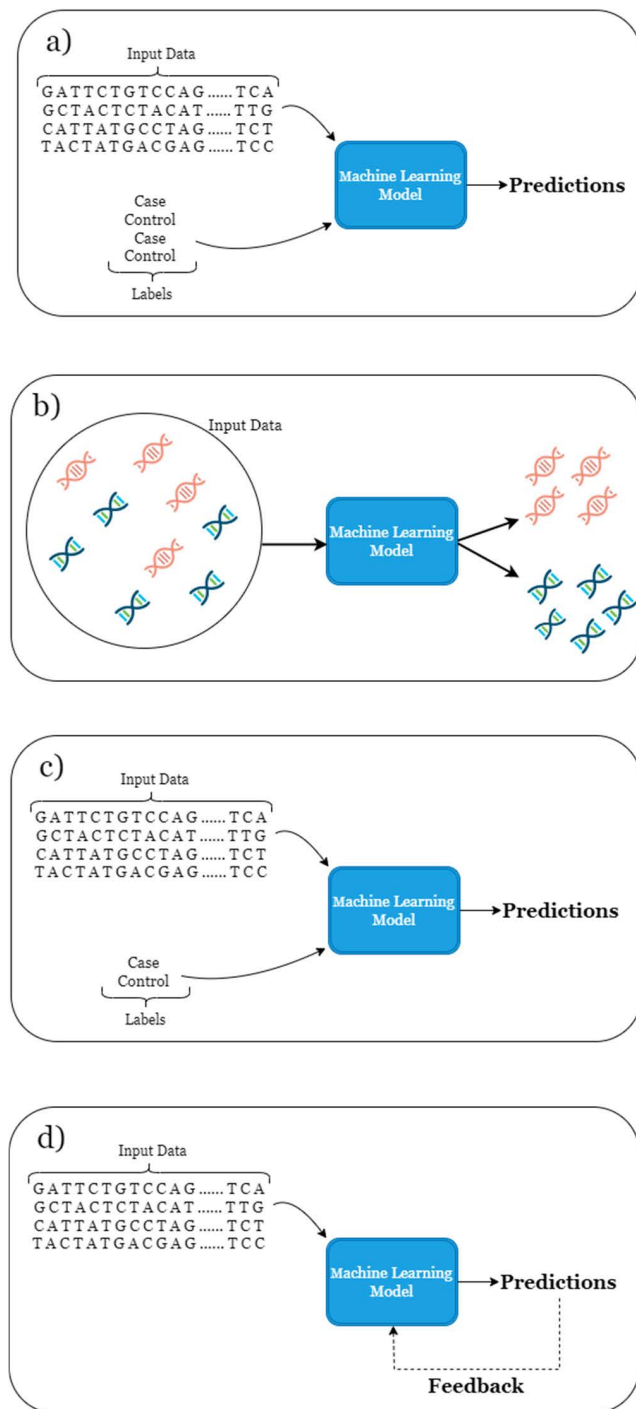


FIGURE 2. Machine Learning types: a) Supervised learning, b) Unsupervised learning, c) Semi-supervised learning and d) reinforcement learning.

for dealing with time series data and natural language processing [28].

Deep learning is a powerful tool for GWAS data analysis, mainly because the amount of data is enormous, far beyond our limited reasoning abilities. In genetic applications, a deep ANN can be built with nodes representing genetic

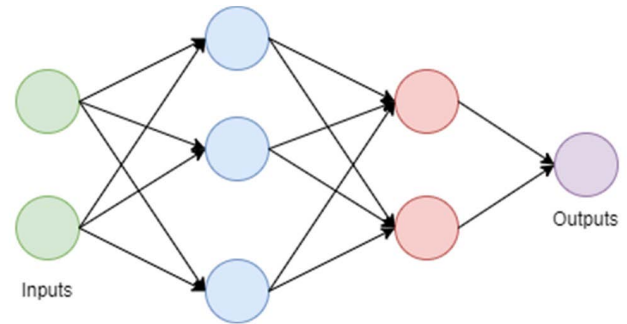


FIGURE 3. Representation of an artificial neural network.

elements (SNPs) and arcs indicating connections (interactions) between the elements.

2) SUPPORT VECTOR MACHINE (SVM)

SVMs are supervised learning algorithms that can be used to solve issues in classification and regression [29]. In high-dimensional spaces, where the number of features exceeds the number of observations, SVMs are well-known for their efficacy. The purpose of SVMs is to find a separating hyperplane with the maximum distance to each class. Different kernel functions are available for SVMs, and the selection of a kernel function depends on both the type of problem and the number of observations [30].

Data points can be classified into different classes based on where they fall - on which side of the hyperplane. A hyperplane's dimension depends on the number of features. If there are only two input features, then it is a line. However, if there are three input features, then it becomes a plane. As the number of features increases, similar to genetic data, it becomes increasingly difficult to imagine the dimensions of the hyperplane. A support vector is a data point that is closer to the hyperplane and is more influential on the position and orientation of the hyperplane. The main idea of SVM is to transfer data into higher dimensions in order to find a suitable boundary that separate the classes in nonlinear classification problems. However, as the number of dimensions increases, the computations inside that space become more expensive. SVM uses a kernel trick to overcome this burden, allowing it to calculate high-dimensional relationships without transferring data into higher dimensions. In the case of GWAS, usually there are two classes: cases and controls. Whereas SNPs represent the features, the SVM is capable of classifying a case and control given the feature set (SNPs). This is done when the SVM is given labelled data points, i.e., SNPs and the output. Figure 4 illustrates a support vector machine model of cases and controls.

3) RANDOM FOREST (RF)

A random forest is a learning algorithm that develops a powerful overall classifier through an ensemble of decision trees. The trees formed by random forests are usually trained using the bagging method. The bagging method explains that by

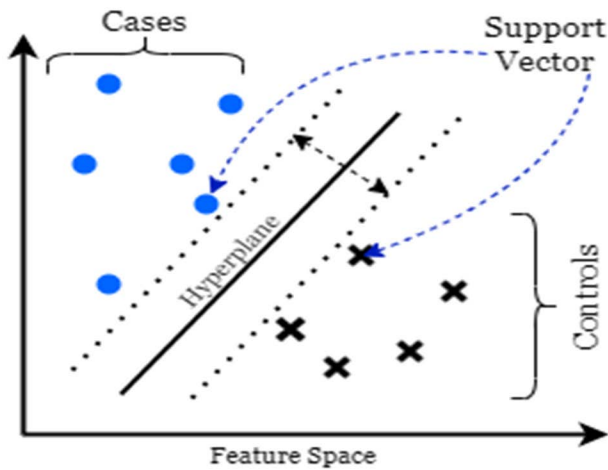


FIGURE 4. An example of SVM model.

combining different learning models, the overall performance can be improved [31].

Because datasets with big size subsets tend to increase computational complexity, a small subset size decreases the difficulty of deciding on the number of characteristics to be used in the training of the model enhances the algorithm's learning speed. Figure 5 illustrates the work process of a RF model. Firstly, the model randomly selects individuals from the original dataset to build new datasets. Each of these newly created datasets will contain the same number of features (SNPs) as the original one. These will be referred to as bootstrapped datasets. Several trees are constructed, and each tree is trained using random features (subset) from the feature set (input) of the bootstrapped datasets. When a RF makes a prediction on a new data point, it will pass the datapoint through each tree and the predictions are recorded. The model then checks all predictions and outputs the majority vote as the final prediction. The process of combining results from multiple models is known as aggregation.

The bootstrapping process ensures that the model does not use the same data in every tree, which helps the model to be more robust. On the other hand, the random feature selection helps reduce the correlation between the trees. RF can be mathematically summarized with the following equations [32]:

$$f(v) = \frac{1}{w} + \sum_{n=1}^w f(v, v_n q) \quad (1)$$

where $v_n q$ is the data variable and v represent the dependent variable.

$$f(v) = \log t_y \frac{1}{Y} \sum_{m=1}^Y \log t_m(y) \quad (2)$$

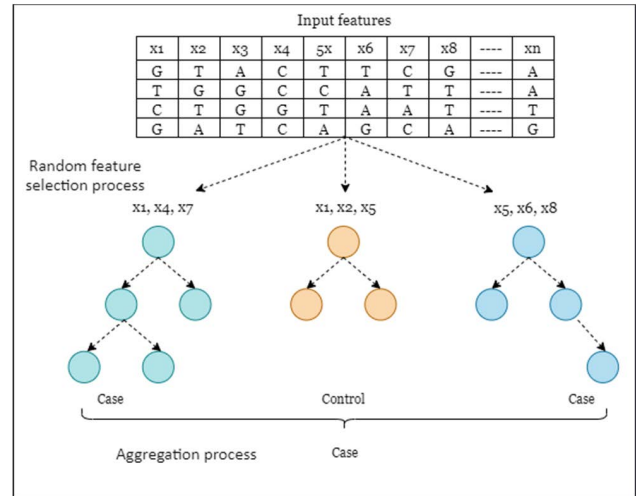


FIGURE 5. Random forest in genetics application.

where Y stands for the total number of classes and y stands for the particular class (in our case, case or control). Furthermore, tk is included in the fraction of total votes for class y .

4) NAIVE BAYES

Naive Bayes is a basic learning technique that use Bayes' rule in conjunction with the strong assumption that attributes are conditionally independent given the class [33]. Despite the fact that this independence assumption is frequently broken in practice, naive bayes classification accuracy is generally competitive. Because of this, as well as its computing efficiency, naive bayes is commonly used in practice.

Posterior probability is Bayes theorem is calculated as follows [33]:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (3)$$

where:

$P(c)$ is the prior probability of class.

$P(x|c)$ is the probability of predictor given class.

$P(x)$ is the prior probability of predictor.

In a genetic analysis scenario, let $P(c|x)$ be the probability for a new data point $X = \langle \text{SNP1}, \text{SNP2}, \dots, \text{SNPn} \rangle$ to belong to class c (case or control).

III. SEARCH STRATEGY

The systematic review relied on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [90] tool to review studies that have used machine learning algorithms in the analysis of genome wide association study data. Scopus, PubMed, and Web of Science were searched for conference and journal articles matching terms like "machine learning" and "genome wide association study". Full search queries are available in the supplementary material file. On 31 December 2021, articles were searched in their title, keywords, or abstract. The search was limited to the English language, and date was set between

1 January 2010 and 31 December 2021. The first author reviewed all the abstracts for inclusion. Full text was assessed when the abstract was relevant to the inclusion requirement.

IV. STUDY SELECTION

Through a two-stage screening procedure, we determined whether the studies retrieved by the search engine were eligible. We examined the paper titles and their abstracts first, then eliminated any papers that did not meet the inclusion criteria. Finally, the complete text of all studies that were considered relevant was evaluated for eligibility using the same screening process.

Studies were considered eligible if they met the following requirements:

- predicting Alzheimer's disease clinical outcomes or Alzheimer's disease alongside other diseases.
- only considering articles that use GWAS as dataset, excluding any articles based on text or imaging data for the detection of Alzheimer's disease.
- primary research rather than review papers.
- The entire manuscript is considered, instead of simply an abstract or notes.

These requirements were chosen to ensure inclusion of all studies investigating Alzheimer's disease, even when another disease is being studied. The restriction of the genetic data to be of type GWAS only is because we wanted to investigate the ability of ML models in this specific area of genetics, which has become popular during the past decade.

After selecting the relevant papers, an analysis of each paper was conducted, considering the following questions and conditions:

- 1) Which ML models were employed?
- 2) The type and source of the data.
- 3) Are there any pre-processing steps conducted?
- 4) What is the overall performance of the model?
- 5) Which hyperparameter optimization methods were utilized?
- 6) What feature selection techniques are used?
- 7) Are there any reported genetic markers?

Papers published between January 2010 and December 2021 were considered. Prior to this time period, ML approaches have been applied in research. However, in the last decade, there has been a surge in interest in machine learning in biological research; as a result, we only found a few research studies prior to that period.

Articles focusing on other forms of genetic data, such as gene expressions or uncommon variants, were not included. Authors that integrated SNPs with additional types of biological markers, such as MRI and PET, were excluded.

V. DATA EXTRACTION

A structured data collection form was created to help in the extraction of elements. The form includes extraction of general study characteristics such as author/s, type, study objective, and publication year as well as the characteristics

of the population used in the study, such as source of data, feature size, and sample size. Finally, other ML models and data pre-processing methods details were also extracted (unbalanced outcomes, other data pre-processing steps, ML models, performance measurements, feature selection methods, and results).

VI. RESULTS

Following an initial search, we identified a total of 283 articles from Scopus, Web of Science and PubMed databases. This number was reduced after removing duplicates, resulting in 165 studies. These were then further reduced to 65 after evaluating if both titles and abstracts met the inclusion criteria. The full texts were then subjected to a more in-depth study, with publications that failed to meet the inclusion requirements after a thorough examination being removed. At this point, 24 articles remained to be included in the review. This number of articles is expected due to the relatively new appearance of ML technologies in this specific area of human genetics as well as the difficulties in obtaining authorization for large GWAS datasets. Figure 6 shows a graphic representation of the study selection process.

As shown in Table 1, all of the included studies were conducted after 2010, with more than 50% of the included papers published after 2015, four studies published in 2020 [19]–[22] and three in 2021 [23]–[25].

A. AREA OF STUDY

Our extensive research indicated that there were three main areas where researchers apply machine learning techniques in genome-wide association studies. In the first domain, a prediction model was developed to classify healthy and unhealthy individuals based on their genetic data. The second domain involved developing multiple intelligent models to identify new genetic markers associated with a particular chronic disease of interest. The third method was to develop models to discover epistatic interactions. The percentage area of use of ML in GWAS for Alzheimer's disease is further illustrated in Fig 7.

B. DATA SOURCES

Large genome-wide association studies datasets were used for training the ML models in every paper. To test the association of each SNP with a phenotype, the genome data of each participant needs to be recorded.

The majority of the researchers (13 articles) used the genome wide association study dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Four articles [34]–[37] used the AD GWAS dataset from the Translational Genomics Research Institute. One article [38] used a dataset from the National Alzheimer's Coordinating Center. One article [39] used a late-onset AD dataset provided by the Harvard Brain Tissue Resource Center and Merck Research Laboratories. Two articles [40], [41] requested datasets from the database of genotype and phenotype. Further details about these datasets are available in Supplementary Table 1.

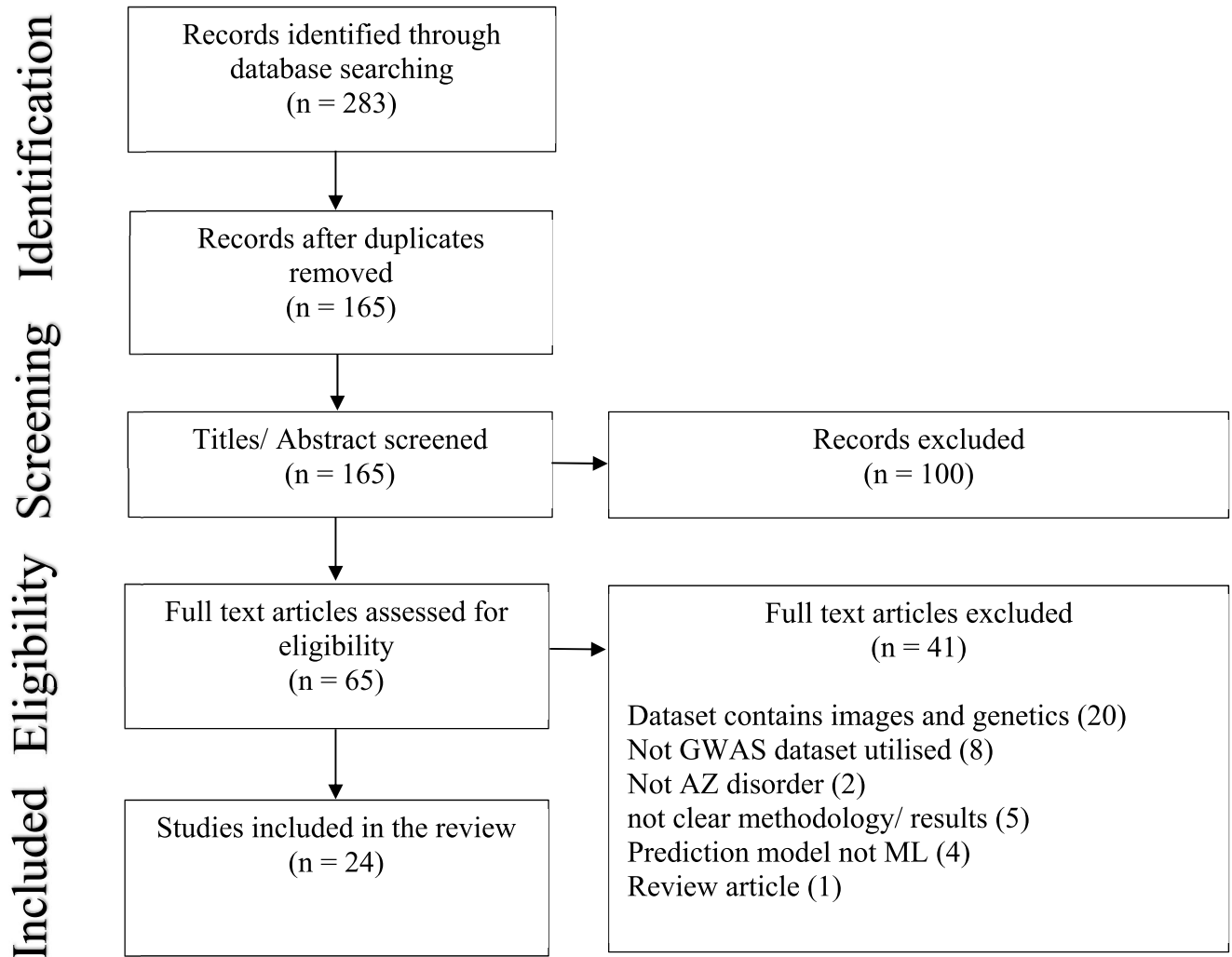


FIGURE 6. A visual breakdown of publication selection based on PRISMA guidelines.

C. DATA PRE-PROCESSING

The vast majority of researchers involved some data pre-processing steps before passing the data to the model. It is well-recognized that quality control in GWAS data is an essential pre-processing step for any analysis of genotype data, especially when studying phenotype associations, as it can have a strong influence on the end results [42]. QC and filtering procedures were performed on individuals and their SNPs in most of the papers using PLANK software. Only four articles [39], [40], [43], [44] did not conduct any quality control procedures.

A number of articles [43], [45]–[48] incorporated APOE (a gene whose polymorphic alleles were the most important genetic predictors of Alzheimer's disease risk [49]) genotyping into the dataset as a pre-processing step. The authors De Velasco Oriol et. al. [50] selected the top 2500 SNPs according to their p-value. An χ^2 statistical test was used in [51] to distinguish between the high and low informative

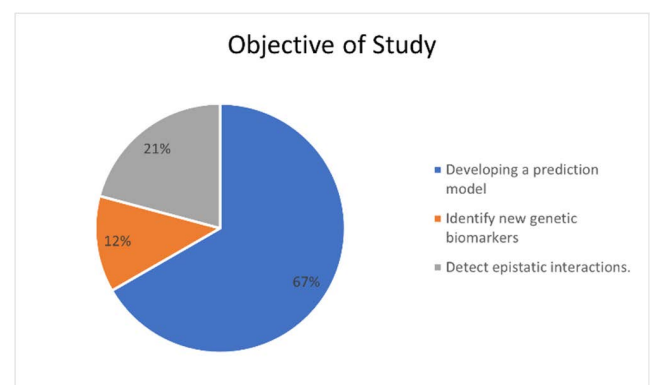


FIGURE 7. Percentage of papers in each area are of study.

subsets of SNPs. In [52], the authors retrieved only SNPs on the 19th chromosome to train their model.

TABLE 1. Overview of Studies Selected Based on Prisma Analysis

Reference	Source of Data	Data Size	Feature Size	Validation	Best Algorithm
Abd El Hamid et al. (2019)	ADNI	391	500	10-fold cross-validation	k2
Ahmed et al. (2020)	ADNI	391	not reported	10-fold cross-validation	NB
(Alatrany et al., 2021)	NACC	364	Not reported	Training-test splitting	MLP
(Araujo et al., 2013)	ADNI	374	Not reported	out of bag	RF
(Briones and Dinu, 2012)	TGen	1411	199 SNPs	10-fold CV	RF
(Chang et al., 2020)	(ADNI)	364	4,916,249	2-fold cross validation leave-one-out CV five-fold cross validation	L1-regularized regression EBMC
(Cooper et al., 2010)	TGen	1411	500		
(De Velasco Oriol et al., 2019)	ADNI	471	1000	CV	ensemble
(El Hamid et al., 2017)	ADNI	391	21	10-fold cross validation	SVM
(Erdogan and Son, 2014)	GENADA / dbGAP	1480	958	11-fold-cross validation	DT
(Han et al., 2012)	TGen	1408	287,479		BN
(Li et al., 2021)	ADNI	1461	Not reported	5-fold cross validation	Deep Learning
(Moore et al., 2017)	ADNI	718	758		MDR
(Nguyen et al., 2015)	ADNI	364	Not reported	5-fold cross-validation	RF
(Osipowicz et al., 2021)	ADNI	485	24 998	10-fold cross validation training and test sets	RF
(Romero-Rosales et al., 2020)	NIA	1,830	1,106	CV	LASSO
(Sherif et al., 2015)	ADNI	330	435		NB
(Sherif et al., 2017)	ADNI	431	8360		MDR
(Stokes et al., 2014)	ADRC TGen	2,229	100	Five-fold cross validation.	KNN
(Wang et al., 2019)	ADNI LOAD**	1017	6970		Deep Learning
(Wei et al., 2011)	TGen	1411	20	fivefold cross-validation	MANB
(Xu et al., 2020)	(ADNI)	785	2943	10-fold cross-validation	TV-GroupSpAM
(Zou et al., 2012)	TGen	644	312316		RF
(Aflakparast et al., 2014)	TGen	364	76 755		CSE

ADNI: ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE, NACC: NATIONAL ALZHEIMER'S COORDINATING CENTER, TGEN: TRANSLATIONAL GENOMICS RESEARCH INSTITUTE, NIA: NATIONAL INSTITUTE ON AGING, dbGAP: DATABASE OF GENOTYPES AND PHENOTYPES, ADRC: ALZHEIMER'S DISEASE RESEARCH CENTER, LOAD**: LATE-ONSET AD DATASET PROVIDED BY HARVARD BRAIN TISSUE RESOURCE CENTRE AND MERCK RESEARCH LABORATORIES. MDR: MULTIFACTOR DIMENSIONALITY REDUCTION, NB: NAIVE BAYES, EBMC: EFFICIENT BAYESIAN MULTIVARIATE CLASSIFICATION ALGORITHM, KNN: K-NEAREST NEIGHBOURS, LASSO: LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR, BN: BAYESIAN NETWORKS, MLP: MULTI-LAYER PERCEPTRON, TV-GROUPSpAM: TIME-VARYING GROUP SPARSE ADDITIVE MODEL, RF: RANDOM FOREST, CSE: CUCKOO SEARCH EPISTASIS

As GWAS is extremely high in terms of dimensionality, all researchers tried to reduce their dimensions by selecting a subset of SNPs. In most papers, authors used logistic regression with different p-values, mainly between (10^{-2} and 10^{-8}) to find the most significant SNPs for a phenotype and, based on that, a subset of SNPs was retrieved for further analysis. Both [43], [44] used the Boruta algorithm to find an appropriate set of SNPs for their model. Whereas in [35] the ReliefF algorithm was utilized. In [53], [54], the authors extracted a set of known genes related to AD based on meta-analyses of GWAS catalogued on AlzGene database. While in reference [50], the authors sorted SNPs according

to p-value using summary statistics from the International Genomics of Alzheimer's Project (IGAP). El Hamid et al [46] selected SNPs from the top 10 AD candidate genes listed on the AlzGene database, a Chi Squared Attribute Evaluator with a ranker search method was used for ranking SNPs in order to select the most important SNPs to further reduce the feature set size. In [55], SNPs that resulted from the intersection of three statistical approaches were selected, including the Allelic test, the Genomic test, and the regression test. Number of features selection testing was used in [56] where they looked at the effectiveness of these tests through the classification task. Wang *et al.* [39] considered only the SNPs

that reside protein coding exons according to GENCODE (a public research consortium) then further excluded the SNPs on X-chromosome

D. TECHNICAL DETAILS

From a technical side, eight of the research papers used a single machine learning algorithm. In what follows, we will provide the details for the machine learning methods utilized, as well as model performance, Sample size and Hyperparameter Search. Summary of machine learning algorithms employed in the included studies is presented in Table 2.

1) MACHINE LEARNING METHODS

On genetic data from the Alzheimer's disease neuroimaging initiative phase 1 dataset, the Naive Bayes, tree augmented Naive Bayes, and K2 learning algorithms were used for early detection of the illness. Based on the p-value criterion (p-value 0.05), the greatest classification accuracy was attained with 500 SNPs. The NB and K2 learning algorithms reached an overall accuracy of 98 percent and 98.40 percent respectively [45]. Based on genetic data from 188 controls and 176 AD patients, a deep learning model for AD prediction was constructed and evaluated. Using Convolutional Neural Networks and Multilayer Perceptrons, the model attained an area under the curve of 0.9 and 0.93, respectively [38].

Araujo *et al.* suggested to use biologically motivated SNP selection as an input to RF for predicting patient risk of developing AD. The findings reveal that non-disease-related SNPs perform similarly to or better than disease-related SNPs. As the identification of novel relevant markers is the most important effort in GWAS. These findings suggest that SNPs from unrelated sets might be new candidates for Alzheimer's disease [53]. In a GWAS data collection of 550 controls and 861 cases, two distinct techniques were designed to find SNPs linked with AD. In the first technique, the authors employed logistic regression to filter the data depending on a p-value threshold, resulting in a subset of SNPs that were then used in by random forest to perform a multi-locus analysis. On the other hand, in the second technique, pre-select loci for input into the RF classifier using biological information and logistic regression analysis. The first method yielded 199 SNPs. Using 10-fold CV in RF modelling, these SNPs, together with other SNPs that associated to AD, produced a predictive subgroup for AD prediction with an average error of 9.8%. With the second method, 19 variations were discovered. These variations were incorporated of a model that includes APOE and GAB2 SNPs to predict AD risk, and the model achieved a 10-fold CV average error of 17.5 percent [34].

Chang *et al.* introduced GenEpi, a computational package that uses L1-regularized regression to find epistasis related with phenotypes. GenEpi uses a two-stage modelling methodology to identify both within-gene and cross-gene epistasis. The ML model was trained and evaluated on genetic data of 364 individuals. A total of 24 SNPs from 12 genes were used in the final model. This model has a 2-fold cross validation accuracy of 0.83 and a leave-one-out CV accuracy

of 0.83 [58]. When applied to a GWAS dataset of Alzheimer's disease that comprises 312,318 SNP of 1411 individuals, Cooper *et al.* compares the predicted performance and efficiency of a Bayesian approach to various traditional machine-learning methods. The findings reveal that the Bayesian algorithm predicts outcomes as well as traditional approaches while requiring less overall training time [35].

FRESA.CAD's (Feature Selection Algorithms for Computer Aided Diagnosis) benchmarking tool was employed to forecast a person's hereditary risk of acquiring AD by construction and evaluating several machine learning models, including Bootstrap Stage-Wise Model Selection (BSWiMS), Least Absolute Shrinkage and Selection Operator (LASSO), and Recursive partitioning and regression trees (RPART). The ROC AUC varied between 0.60 and 0.70. The BSWiMS, LASSO, and RPART performed similarly, and the ensemble of approaches performed best, with a ROC score of 0.719 [50]. Support Vector Machine (SVM) classifiers of various kernels were applied to ADNI data based on chosen 21 variations most related with AD using two techniques, Correlation-based and Chi-squared. The results show that an SVM trained model utilising an RBF kernel reached the highest accuracy of 76.70 percent [46].

The authors [44] ran two types of experiments to see whether information used to categorise one dataset may be effectively utilised to classify a completely separate patient group dataset. In the first experiment, the authors trained a random forest classifier from features selected from the first dataset. The resulting model is evaluated using the second dataset. Secondly, authors picked significant features based on the first dataset's training subset and then utilised the positions of the selected SNPs to create a new random forest model using the data from the second dataset. In both of the experiments, the resultant classifier's performance was evaluated on patients from the second dataset who had not utilised in training the tested classifier. The AUC values for both experiments decreased slightly when compared to the results from one dataset, but they remained well above 0.5, indicating that all of the models contain some universal information about genetic differences between AD cases and controls.

Nguyen *et al.* propose a ts-RF, a new two-stage strategy in random forests, for selecting a subset SNPs in GWAS. The suggested method has been found to be successful in identifying informative sets of SNPs that may relate to illnesses [51].

Authors present a novel technique for AD GWAS analysis that combines both random forests and enrichment analysis to identify new genetic biomarkers based on data from 527 controls and 117 cases [59]. Wang *et al.* proposed a deep learning model method that is capable of identifying interactions between SNPs in GWAS data. The model, named Deep Mixed Model, consists of two components, the first component acts as a confounding factor correction by using a CNN, while the second component utilize LSTM for genetic variants selection [39]. Romero-Rosales *et al.*

present a comparison of three ML models: genetic algorithm sand step-wise, as well as L1 (LASSO) approaches for developing models for Alzheimer's disease prediction trained on data of 813 cases and 1,017 controls. In practice, LASSO models outperformed the other two techniques [41]. Sherif *et al.* developed a framework for evaluating several Bayesian network algorithms (naive Bayes, tree augmented naive Bayes, Markov blanket MB, and minimum augmented Markov blanket) to predict the condition of the diagnostic variable, AD case or control. A total of 435 were considered predictors for naive and tree augmented naive networks. However, selecting only 11 and 13 SNPs to train and test Markov blanket and minimum augmented MB respectively showed improved accuracy [54].

Model-averaged naive Bayes (MANB) technique was used to predict AZ. The models were trained and tested on the genetics data of 1411 people. The outputs of the model were compared to the results of a naïve bayes MANB achieved an AUC of 0.72, which is much higher than NB's AUC of 0.59 while maintaining the same training time [37]. Aflakparast *et al.* presents a novel strategy, cuckoo search epistasis (CSE), for detecting epistatic interactions in case-control studies. This technique combines a Bayesian scoring function with a heuristic search algorithm. CSE was able to discover interactions that were reported in the literature [60]. Stokes *et al.* assess the efficacy of label propagation (LP), a multivariate graph-based approach for effectively ranking SNPs in genome-wide data. The top-ranked SNPs were evaluated in terms of classification performance, and prior evidence of being linked with AD. LP performed much better in categorization than other control approaches. There were 14 SNPs in one dataset among the 25 top-ranked SNPs found by LP that had evidence in the literature of being linked with AD [56].

Li *et al.* present a novel deep-learning genomics technique and apply it to the multitasking categorization of Alzheimer's disease progression. For the DLG model, the ResNet framework was employed using 1461 patients' genotyping data, and the results were compared to those obtained using a basic convolutional neural network structure. When applied to the course of Alzheimer's disease, the DLG model can obtain improved accuracy and sensitivity [47]. Moore *et al.* present Crush, a stochastic search technique to explore relations between genes in genome-wide data as an application of multifactor dimensionality reduction (MDR). Applying the approach to an AZ GWAS dataset, results showed that Crush-MDR was capable of identifying a collection of interacting genes with biological linkages to Alzheimer's disease [57]. Xu *et al.* employed a Time-Varying Group Sparse Additive Model (TV-GroupSpAM) to carry out a two-stage SNP selection. First, they divided the genotyping data into 30 subgroups, each with 5000 SNPs. This resulted in a filtered collection of 2943 SNPs for TV-GroupSpAM. The technique was then repeated, but this time just with the SNPs that had been chosen in the initial round. This has resulted in a final collection of 126 SNPs. TV-GroupSpAM completed the

analysis in 16 days, which was faster than the two control techniques [52].

RFs and NB were two prominent algorithms employed in the evaluated research. NB is known for its ease of construction. However, due to the nature of GWAS data, which contains strong interrelationships between SNPs, which runs counter to the naive notion that all input characteristics are independent [61]. Because of their capacity to avoid overfitting, RFs are a popular classifier [62]. However, because the classifier requires extensive hyperparameter searching, utilising RFs to predict disease risk may be difficult.

Neural networks are strong prediction algorithms that can understand non-linear relationships in large and complicated datasets. In certain cases, NN may infer data associations that are beyond the purview of other ML approaches. NNs are notorious for being difficult to employ, tune hyperparameters, and prone to overfitting. Especially, in the case where a dataset has many more features than observations [63]. As SVMs are well-known for their simplicity and predictive accuracy, and as a result, they are often used in prediction modelling [64]. Whereas Bayesian modelling approaches have various characteristics that make them valuable in a wide range of genetic data analysis tasks. They enable the merging of data and domain expertise. They also allow for easier understanding of the causal links between variables. While Bayesian models are a valuable technique to describe expert knowledge, getting the knowledge from the experts in a manner that can be turned into probability distributions may be problematic [65].

The process of choosing a model architecture entails balancing model underfitting and overfitting, often known as the bias-variance trade-off [66]. When a low-capacity model is adopted in relation to the issue complexity and dataset size, underfitting is prevalent. Underfitting can be mitigated by using a more parameter-rich model or using less regularisation during training. More serious is model overfitting, which occurs when the evaluator overestimates the generalisation performance on previously unknown data. A surprisingly poor performance on the test set compared to the training set is an indicator of overfitting. Overfitting is often avoided by employing a validation set inside the trainset for performance estimations, as well as numerous regularisation algorithms, for example, early stopping [67].

2) MODEL PERFORMANCE

AUC was used in over half of the research papers (55% models) as performance measurement of the models. There was also information on a variety of categorization metrics and model fit assessments such as accuracy, sensitivity and specificity. Out of bag error was used in few papers (Supplementary table 8).

Internal validation was reported by around 68% of models. The most of them used k-fold cross-validation, a resampling technique that includes testing a model on each of k distinct divisions of a dataset and then training on the remaining k-1 folds. The most popular method was 10-fold

TABLE 2. Summary of Machine learning algorithms

Ref	ML Models	Problem Addressed	Dataset Control/case	Model Performance	Feature Selection
[45]	Linear Quadratic Polynomial, cubic polynomial, Naïve Bayes, TAN, K2	Developing a classifier model	214/177	0.95-0.99	p-value threshold of 0.05 then selected top 100
[38]	MLP, CNN	Developing a classifier model	188/176	0.92-0.93	Logistic regression then Random Forest
[53]	RF	Developing a classifier model	205/169	Error out of Bag 0.45	Used set of known genes from AlzGene database
[34]	RF	Identify new genetic biomarkers	550/861	Not reported	Logistic regression
[58]	L1-regularized regression with stability selection	To explore epistasis interactions.	241/132	0.83-0.94	χ^2 test
[35]	LR, NB, SVM, EBMC	Developing a classifier model	550/861	0.61-0.72	ReliefF algorithm
[50]	BSWiMS, LASSO, RF, RPART, KNN with BSWiMS features, SVM with mRMR, ensemble	Developing a classifier model	230/241	0.494-0.719	p-value using summary statics
[46]	SVM	Developing a classifier model	214/177	0.62-0.77	SNPs of top 10 AD candidate genes listed on the AlzGene
[44]	Fandom forest	Developing a classifier model	266/219	0.67	Boruta algorithm
[51]	RF SVM	Developing a classifier model	188/176	0.62-0.97	Wilcoxon test
[59]	Random forests with enrichment analysis	Identify new genetic biomarkers	527/117	Not reported	-
[39]	Deep Mixed Model	To explore epistasis interactions	477/540	Not reported	SNPs that reside protein coding exons according to GENCODE
[41]	BSWiMS, Genetic Algorithms Vector Machine, GALGO, LASSO,	Developing a classifier model	1017/813	0.68-0.844	χ^2 test
[54]	Naïve Bayes Tree augmented naïve Bayes Markov blanket Minimal augmented Markov blanket	Developing a classifier model	282/48	0.62-0.66	Top ten AD candidate genes listed on the AlzGene database
[37]	NB FSNB MANB	Developing a classifier model	550/861	0.59-0.72	Laplace parameter
[60]	Cuckoo search epistasis	To explore epistasis interactions.	550/861	Not reported	χ^2 test
[56]	KNN	Identify new genetic biomarkers	938/1291	0.52-0.73	chi squared Sigmoid Weighted ReliefF
[47]	deep residual network ResNet, CNN	Developing a classifier model	622/366	0.58-0.99	Chi-square test
[57]	Crush-MDR	To explore epistasis interactions.	718	Not reported	Constructed a list of 734 genes using Ingenuity Pathway Analysis
[52]	TV-GroupSpAM	Identify new genetic biomarkers	785	0.53-0.63	TV-GroupSpAM

cross-validation (CV). A random split between training and testing sets was employed in one study for internal validation [38]. Internal validation was not mentioned in five of the total studies.

Within studies, the performance of models varied depending on the machine learning algorithm used, the sample size, and the number of features. The AUC for Alzheimer's disease models ranged from 0.59 to 0.98, with the greatest AUC observed utilizing a deep residual network [47]. On the ADNI phase 1 dataset, Nave Bayes produced a similarly high AUC (0.97 AUC) [43]. In [45] the authors reported the accuracy of 92.68 while using 100 SNPs a feature set size, this has risen to 98.4 by increasing the number of features to 500 SNPs. Despite the fact that there are few studies, deep learning outperforms other techniques of machine learning [38]. The outcomes of utilizing an ensemble of machine learning algorithms appear to be superior to using a single method [50].

A variety of techniques have been used to genotype-based categorization of Alzheimer's disease patients and healthy controls using GWAS data, with different accuracies reported. However, because the predicted effect of genotype on sporadic AD prevalence is minimal, these extremely high classification accuracies might be the product of overfitting. Osipowicz et al [44] found that if feature selection is undertaken before splitting the data into training and testing sets, the techniques are prone to overfitting. As a result, it was discovered that it was preferable to avoid selecting features used to create the model based on data contained in the testing set. According to the authors, the anticipated classifier performance for currently available dataset sizes is between 0.55 and 0.7 (AUC), and greater accuracies reported in the literature are most likely the consequence of overfitting.

Despite of performing well in terms of prediction accuracy in some datasets, ML still suffers from selecting informative SNPs and build accurate prediction models.

3) SAMPLE SIZE

In the study of brain illnesses, small sample numbers are a prevalent problem. The vast discrepancy between the number of characteristics and the amount of data accessible in genetic data adds another layer of complexity: A GWAS can include over a million SNPs, each of which can be considered a feature; however, the average number of participants is significantly lower.

The overall sample size was relatively small, with 11 of the total research papers using the dataset sizes ranging from 364 to 485. A dataset of over 1000 samples was utilised in 9 investigations. A dataset with sample sizes in the 700s was utilized in two research. One study involved the use of imbalance class distribution but did not report if any steps were taken to re-balance the distribution prior to the analysis [54] (supplementary table 6).

The amount of data utilised in several of the listed research was insufficient to properly explore the possibilities of machine learning technologies. Larger datasets can aid in the accurate selection of candidate SNPs, such as those used in

meta-analysis [68]. Research can be focused on developing machine learning models that works well for data were the number of features much larger than the number of samples, which is a typical challenge in GWAS. Simulated datasets can be a good starting point in developing ML models that can be then tested on real datasets.

4) HYPERPARAMETER SEARCH

The majority of hyperparameter searches went unreported or were ambiguous, few models being reported as having been utilized with default parameters. The type of search and tuning for a single model resulted in ambiguous reporting since it was unclear whether these circumstances applied to other models in the research. As a result, it's possible that most research considered many hyperparameter options but did not mention them (Supplementary Table 7).

E. REPORTED GENETIC MARKERS

Most of the studies (19 papers) reported associated SNPs with Alzheimer's disease according to their analysis. Five studies reported the SNP rs429358 as an Alzheimer's disease risk factor which is found in the ApoE gene's fourth exon. Han [36] found interaction among the SNPs: rs1931565 and rs4505578 with APOE. A list of high SNPs reported by each article is stated in Supplementary Table 10.

VII. DISCUSSION

This systematic review contributed to the findings in the literature by evaluating in detail the ML algorithms used for detection of Alzheimer's disease.

The majority of studies conducted for each model only reported the measurements of either discrimination, or classification. Only few of the studies publicized the measure of calibration used within the models [37]. Model building was based on many fundamental properties, including model calibration [69]. Model calibration has been identified as a fundamental aspect which is lacking in the field of generic prediction literature [70]. The Model Calibration compares the anticipated probability of the result happening with the observed probability. When looking at the presented work of authors, the discrimination measures should also be presented along with the classification's measures of specificity, accuracy, and sensitivity as they displayed all the available information on long projected probabilities. Within genetics and ML, the most commonly used and available discrimination metric was the AUC.

The hyperparameter optimization is used to determine how machine learning models navigate the bias-variance trade off and learn from the data [71]. It is surprising that this was unobserved or exposed to only a number of manual tests. To ensure any of the models are neither under fit nor over fit, the use of hyperparameters should be investigated in systematic ways. Hyperparameter selections often decrease overfitting, hence it is of utmost importance that search is used within candidate predictors such as genomics. This is

applicable to those areas which have a minor or small number of samples over features.

In all of the papers reviewed, there was no presentation of any decision analytical methods, which are usually used to assess the clinical values of prediction models. The larger aim, objective of these research works is to assist doctors in determining the correct prognosis for patients and aiding them in treatment and planning the decision making. With this being said, no research has been seen to address the use of the model in real-world clinical trials and practices. A number of reasons varying in description can be given as to why machine learning cannot be seen and believed to be difficult to use in the healthcare profession and settings [72]. ML algorithms are often opaque in terms of how the prediction was formed and how various predictors contributed to the final decision. This may have a declining effect on the legitimacy and acceptability of the model's predictions within decision making for healthcare professionals. For model duplication, a degree of transparency is needed to be used in other datasets. It was brought to our attention that most studies model reporting, along with the model development, were not thorough enough in their transparency to be permitted to be used in other datasets. This therefore insinuates that most models will have an inadequate amount of evidence to support their accuracy in various settings, as well as being impractical when used in real-world healthcare settings [73].

Reporting criteria and guidelines for incorporating ML algorithms have a chance of increasing ML acceptability. Further studies in the future would profit from attempting to evaluate clinical utility along with the potential effects [74]. To begin with, the accumulation of information and reporting the recommendations for the verification and design of clinical prediction models would be a sensible starting point. New research [75]–[77], has dwelled and dived into the potential ethical issues which may come to light when using machine learning models. It is of utmost importance that algorithms and generated models be publicly accessible as well as comprehensive and transparent in their reporting. This would allow the promotion of clinical utility along with independent external validation across multiple contexts.

A. THE CURSE OF DIMENSIONALITY

A well-known problem with genetics data is that the number of attributes is much larger than the number of samples, introducing a challenge not only for ML algorithms but for DL and hence the general statistical approaches [78]. To avoid the curse of dimensionality, feature selection and feature extraction are often used [79], [80]. Some researchers [81]–[84] try using multiple datasets to provide a larger number of samples to balance the number of features and samples. Other researchers used association analysis techniques to reduce the dimensions of the original dataset to a significantly smaller number, mostly using logistic regression and selecting only SNPs that passed a threshold level of p-value [45].

B. MODEL INTERPRETATION

One of the main concerns when using ML approaches is the ambiguity in decision-making. Owing to the architecture of the machine learning algorithms, it is not yet clear how the learned patterns are formed, therefore making a full image of how the model reached a specific output from inputs a difficult task [85]. Researchers, especially those working in healthcare domain applications, prefer to use white box methods to understand the decisions made by these applications since they are very sensitive, and any small error could be costly [86]. In [47], the last convolutional layer of the last res-block was made transparent to extract features used to investigate the interpretability of the DL model.

C. HYPERPARAMETERS TUNING

One of the challenging tasks in ML is the tuning of the model's hyperparameters. For instance, ANNs require several hyper parameters, including learning rate. A very small learning rate value could take a long time to converge or get stuck at a local minimum. In contrast selecting a large value for the learning rate, it will process parameters faster but most likely lead to an oscillation. Thus, this is the most important hyperparameter and a careful decision should be made. Tuning the parameters is critical step of building an accurate ML model, there are few practices existing in choosing the right set of hyperparameters, such as the Grid search [83] for tuning the parameters of ML. Whereas Bellot et al [87] used genetic algorithm for hyperparameter optimization. Many researchers randomly test few combinations of hyperparameters and select the best values according to the model performance.

D. IMBALANCE CLASSES

Another obstacle that limits the ability of ML models is when the number of samples is extremely different in each class. As the aim of ML for classification problems, such as classifying healthy vs unhealthy [82] or responding to disease treatment vs unresponsive to the treatment [88], is to obtain an efficient model for such discrimination cases, a satisfactory number of samples per class should be provided.

VIII. RISK OF BIAS

Within each paper that aimed to develop a prediction model, the risk of bias was estimated for the optimal model. All of the models showed signs of bias, mainly in the area of analysis. Risk of bias (ROB) was assessed following PROBAST [89] guidelines. Participants, predictors, results, and analysis are the four areas that PROBAST is divided into. There are a total of 20 questions in these four categories to help with structured ROB judgement. The answers to these questions are recorded in supplementary ROB Table. Participants' ROB was rated low in all the included studies. PROBAST for predictors is designed to help the researcher determine if the processes for measuring biomarkers were the same for all study participants. Procedures for collecting predictors in

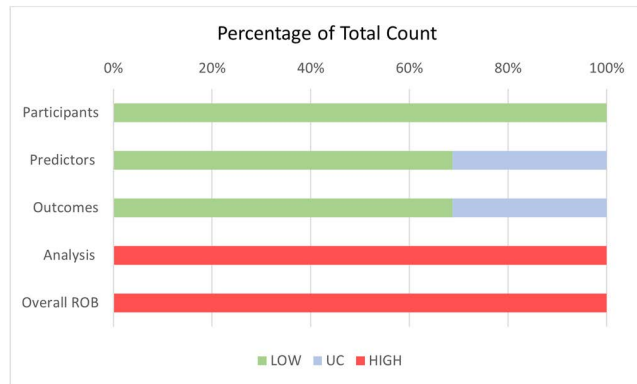


FIGURE 8. Risk of bias of studies included in the review.

references [35], [37], [40], [41], [56] were not provided. As a result, ROB for predictors was judged to be unclear. The procedures for collecting biomarkers are outlined in public materials provided by ADNI. For all subjects, predictors generated from blood samples or MRI images were gathered using the same techniques. As a result, the procedure of acquiring predictors was assessed to have a low ROB for articles that used the ADNI dataset.

During analysis, the models showed a high ROB (as illustrated in Fig 8). It should be noted that the amount of the data used to create the model, the incorrect or unjustifiable handling of missingness, the removal of registered individuals prior to analysis, predictor selection using univariable approaches, and inability to account for overfitting were all common causes.

IX. CONCLUSION

In order to decide where to focus their research efforts, GWAS researchers need a comprehensive picture of the trends in ML algorithm utilization. This paper offers a rigorous examination of the machine learning techniques employed in the GWAS of Alzheimer's disease.

A total of 24 research publications were included following meticulous filtering based on exclusion criteria. There has been, and currently is a rising interest in the utilization of ML to aid in predicting AD results. Deep learning algorithms are becoming more popular in GWAS analysis, especially when conventional Artificial Neural Networks are used. Approaches of transfer learning are still a study topic. The trends and opportunities described are confirmed by a chronology of the number of primary studies published in recent years.

Regardless of the incursion of research publications, A very limited number of papers met the standard criteria of clinical prediction tools, with non-making their models obtainable in a format which is either evaluable or usable. When speaking of the improvements and enhancements of current machine learning prediction algorithms and how they are built and verified, we can identify a few aspects which can be looked into, beginning with the use of a larger scale data source which is also richer and more diverse in the data it

withholds, along with an improvement in model architecture, and finally, providing thorough reports on the development method used and the final model. To be able to use such approaches and evaluate their use, highly required improvements are needed within reporting and ML research design. Within this specific framework, the use of guidelines, along with reporting requirements used for implementing ML algorithms, could aid in the surge in value of investigations.

Future work will investigate ML models that have been applied to genetic and image data for Alzheimer's disease.

REFERENCES

- [1] M.-L. Hwang, Y.-D. Lin, L.-Y. Chuang, and C.-H. Yang, "Determination of the SNP-SNP interaction between breast cancer related genes to analyze the disease susceptibility," *Int. J. Mach. Learn. Comput.*, vol. 4, no. 5, p. 468, 2014.
- [2] L. Bertram, C. M. Lill, and R. E. Tanzi, "The genetics of Alzheimer disease: Back to the future," *Neuron*, vol. 68, no. 2, pp. 270–281, Oct. 2010.
- [3] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," presented at the 27th Int. Conf. Neural Inf. Process. Syst., Montreal, Canada, vol. 2, 2014, pp. 3104–3112.
- [4] H. J. Cordell, "Detecting gene-gene interactions that underlie human diseases," *Nature Rev. Genet.*, vol. 10, no. 6, pp. 392–404, Jun. 2009.
- [5] M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, M. Prasad, and C. T. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 1, pp. 1–35, 2020.
- [6] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster, "A systematic review shows, no. performance benefit of machine learning over logistic regression for clinical prediction models," *J. Clin. Epidemiol.*, vol. 110, pp. 12–22, Jun. 2019.
- [7] M. Bracher-Smith, K. Crawford, and V. Escott-Price, "Machine learning for genetic prediction of psychiatric disorders: A systematic review," *Mol. Psychiatry*, vol. 26, no. 1, pp. 70–79, Jan. 2021, doi: [10.1038/s41380-020-0825-2](https://doi.org/10.1038/s41380-020-0825-2).
- [8] N. S. Madhukar and O. Elemento, "Bioinformatics approaches to predict drug responses from genomic sequencing," *Methods Mol. Biol.*, vol. 1711, pp. 277–296, Jan. 2018, doi: [10.1007/978-1-4939-7493-1_14](https://doi.org/10.1007/978-1-4939-7493-1_14).
- [9] R. Upstill-Goddard, D. Eccles, J. Fliege, and A. Collins, "Machine learning approaches for the discovery of gene-gene interactions in disease data," *Briefings Bioinf.*, vol. 14, no. 2, pp. 251–260, Mar. 2013.
- [10] J. Wu and Y. Zhao, "Machine learning technology in the application of genome analysis: A systematic review," *Gene*, vol. 705, pp. 149–156, Jul. 2019, doi: [10.1016/j.gene.2019.04.062](https://doi.org/10.1016/j.gene.2019.04.062).
- [11] J. H. Moore, F. W. Asselbergs, and S. M. Williams, "Bioinformatics challenges for genome-wide association studies," *Bioinformatics*, vol. 26, no. 4, pp. 445–455, Feb. 2010, doi: [10.1093/bioinformatics/btp713](https://doi.org/10.1093/bioinformatics/btp713).
- [12] D. S. W. Ho, W. Schierding, M. Wake, R. Saffery, and J. O'Sullivan, "Machine learning SNP based prediction for precision medicine," *Frontiers Genet.*, vol. 10, p. 267, Mar. 2019.
- [13] T. W. Rowe, I. K. Katzourou, J. O. Stevenson-Hoare, M. R. Bracher-Smith, D. K. Ivanov, and V. Escott-Price, "Machine learning for the life-time risk prediction of Alzheimer's disease: A systematic review," *Brain Commun.*, vol. 3, no. 4, Oct. 2021, Art. no. fcab246.
- [14] R. H. Hariri, E. M. Fredericks, and K. M. Bowers, "Uncertainty in big data analytics: Survey, opportunities, and challenges," *J. Big Data*, vol. 6, no. 1, p. 44, Dec. 2019, doi: [10.1186/s40537-019-0206-3](https://doi.org/10.1186/s40537-019-0206-3).
- [15] S. Behjati and P. S. Tarpey, "What is next generation sequencing?" *Arch. Disease Childhood*, vol. 98, no. 6, p. 236, 2013, doi: [10.1136/archdischild-2013-304340](https://doi.org/10.1136/archdischild-2013-304340).
- [16] F. Pompanon, A. Bonin, E. Bellemain, and P. Taberlet, "Genotyping errors: Causes, consequences and solutions," *Nature Rev. Genet.*, vol. 6, no. 11, pp. 847–859, Nov. 2005.
- [17] T. M. Mitchell, "Artificial neural networks," *Mach. Learn.*, vol. 45, pp. 81–127, Jan. 1997.
- [18] W. Kratsch, J. Manderscheid, M. Röglinger, and J. Seyfried, "Machine learning in business process monitoring: A comparison of deep learning and classical approaches used for outcome prediction," *Bus. Inf. Syst. Eng.*, vol. 63, no. 3, pp. 261–276, 2021, doi: [10.1007/s12599-020-00645-0](https://doi.org/10.1007/s12599-020-00645-0).

- [19] J.-A. Choi and K. Lim, "Identifying machine learning techniques for classification of target advertising," *ICT Exp.*, vol. 6, no. 3, pp. 175–180, Sep. 2020.
- [20] M. L. Giger, "Machine learning in medical imaging," *J. Amer. College Radiol.*, vol. 15, no. 3, pp. 512–520, Mar. 2018.
- [21] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015. [Online]. Available: <http://science.sciencemag.org/content/sci/349/6245/255.full.pdf>
- [22] V. Kumar and M. L., "Predictive analytics: A review of trends and techniques," *Int. J. Comput. Appl.*, vol. 182, no. 1, pp. 31–37, Jul. 2018.
- [23] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, Jul. 2019.
- [24] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Gener. Comput. Syst.*, vol. 117, pp. 47–58, Apr. 2021.
- [25] S. I. Gallant, "Perceptron-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 1, no. 2, pp. 179–191, Jun. 1990.
- [26] T. Kim and T. Adali, "Fully complex multi-layer perceptron network for nonlinear signal processing," *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 32, nos. 1–2, pp. 29–43, 2002.
- [27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, Nov. 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [28] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Dec. 2021, doi: [10.1186/s40537-021-00444-8](https://doi.org/10.1186/s40537-021-00444-8).
- [29] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, May 2000.
- [30] D. A. Piser and D. M. Schnyer, "Support vector machine," in *Machine Learning*, A. Mechelli and S. Vieira, Eds. New York, NY, USA: Academic, 2020, ch. 6, pp. 101–121.
- [31] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, Jun. 2016.
- [32] M. Khalaf, H. Alaskar, A. J. Hussain, T. Baker, Z. Maamar, R. Buyya, P. Liatsis, W. Khan, H. Tawfik, and D. Al-Jumeily, "IoT-enabled flood severity prediction via ensemble machine learning models," *IEEE Access*, vol. 8, pp. 70375–70386, 2020, doi: [10.1109/ACCESS.2020.2986090](https://doi.org/10.1109/ACCESS.2020.2986090).
- [33] N. Kleanthous, A. J. Hussain, W. Khan, J. Sneddon, A. Al-Shamma'a, and P. Liatsis, "A survey of machine learning approaches in animal behaviour," *Neurocomputing*, vol. 491, pp. 442–463, Jun. 2022, doi: [10.1016/j.neucom.2021.10.126](https://doi.org/10.1016/j.neucom.2021.10.126).
- [34] N. Briones and V. Dinu, "Data mining of high density genomic variant data for prediction of Alzheimer's disease risk," *BMC Med. Genet.*, vol. 13, no. 1, Dec. 2012, Art. no. 7, doi: [10.1186/1471-2350-13-7](https://doi.org/10.1186/1471-2350-13-7).
- [35] G. F. Cooper, P. Hennings-Yeomans, S. Visweswaran, and M. Barmada, "An efficient Bayesian method for predicting clinical outcomes from genome-wide data," in *Proc. Annu. Symp.*, 2010, pp. 127–131.
- [36] B. Han, X.-W. Chen, Z. Talebizadeh, and H. Xu, "Genetic studies of complex human diseases: Characterizing SNP-disease associations using Bayesian networks," *BMC Syst. Biol.*, vol. 6, no. S3, Dec. 2012, Art. no. S14, doi: [10.1186/1752-0509-6-S3-S14](https://doi.org/10.1186/1752-0509-6-S3-S14).
- [37] W. Wei, S. Visweswaran, and G. F. Cooper, "The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data," *J. Amer. Med. Informat. Assoc.*, vol. 18, no. 4, pp. 370–375, 2011, doi: [10.1136/amiajnl-2011-000101](https://doi.org/10.1136/amiajnl-2011-000101).
- [38] A. Alatrany, A. Hussain, J. Mustafina, and D. Al-Jumeily, "A novel hybrid machine learning approach using deep learning for the prediction of Alzheimer disease using genome data," in *Proc. Int. Conf. Intell. Comput.*, vol. 12838, 2021, pp. 253–266.
- [39] H. Wang, T. Yue, J. Yang, W. Wu, and E. P. Xing, "Deep mixed model for marginal epistasis detection and population stratification correction in genome-wide association studies," *BMC Bioinf.*, vol. 20, no. S23, Dec. 2019, Art. no. 656, doi: [10.1186/s12859-019-3300-9](https://doi.org/10.1186/s12859-019-3300-9).
- [40] O. Erdogan and Y. A. Son, "Predicting the disease of Alzheimer with SNP biomarkers and clinical data using data mining classification approach: Decision tree," presented at the e-Health Continuity Care, 2014.
- [41] B.-L. Romero-Rosales, J.-G. Tamez-Pena, H. Nicolini, M.-G. Moreno-Treviño, and V. Treviño, "Improving predictive models for Alzheimer's disease using GWAS data by incorporating misclassified samples modeling," *PLoS ONE*, vol. 15, no. 4, Apr. 2020, Art. no. e0232103, doi: [10.1371/journal.pone.0232103](https://doi.org/10.1371/journal.pone.0232103).
- [42] P. Zeng, Y. Zhao, C. Qian, L. Zhang, R. Zhang, J. Gou, J. Liu, L. Liu, and F. Chen, "Statistical analysis for genome-wide association study," *J. Biomed. Res.*, vol. 29, no. 4, p. 285, 2015.
- [43] H. Ahmed, H. Soliman, and M. Elmogy, "Early detection of Alzheimer's disease based on single nucleotide polymorphisms (SNPs) analysis and machine learning techniques," in *Proc. Int. Conf. Data Anal. Bus. Ind., Way Towards Sustainable Economy*, 2020, pp. 1–6, doi: [10.1109/ICDABI51230.2020.9325640](https://doi.org/10.1109/ICDABI51230.2020.9325640).
- [44] M. Osipowicz, B. Wilczynski, and M. A. Machnicka, "Careful feature selection is key in classification of Alzheimer's disease patients based on whole-genome sequencing data," *NAR Genomics Bioinf.*, vol. 3, no. 3, Sep. 2021, Art. no. lqab069, doi: [10.1093/nargab/lqab069](https://doi.org/10.1093/nargab/lqab069).
- [45] M. M. Abd El Hamid, M. S. Mabrouk, and Y. M. K. Omar, "Developing an early predictive system FOR identifying genetic biomarkers associated TO Alzheimer's disease using machine learning techniques," *Biomed. Eng., Appl., Basis Commun.*, vol. 31, no. 5, 2019, Art. no. 1950040, doi: [10.4015/S1016237219500406](https://doi.org/10.4015/S1016237219500406).
- [46] M. M. A. El Hamid, Y. M. K. Omar, and M. S. Mabrouk, "Identifying genetic biomarkers associated to Alzheimer's disease using support vector machine," in *Proc. 8th Cairo Int. Biomed. Eng. Conf.*, 2017, pp. 5–9, doi: [10.1109/CIBEC.2016.7836087](https://doi.org/10.1109/CIBEC.2016.7836087).
- [47] L. Li, Y. Yang, Q. Zhang, J. Wang, and J. Jiang, "Use of deep-learning genomics to discriminate healthy individuals from those with Alzheimer's disease or mild cognitive impairment," *Behavioural Neurol.*, vol. 2021, Jul. 2021, Art. no. 3359103, doi: [10.1155/2021/3359103](https://doi.org/10.1155/2021/3359103).
- [48] H. Ahmed, H. Soliman, and M. Elmogy, "Early detection of Alzheimer's disease based on single nucleotide polymorphisms (SNPs) analysis and machine learning techniques," in *Proc. Int. Conf. Data Anal. Bus. Ind., Way Towards Sustain. Economy (ICDABI)*, Oct. 2020, pp. 1–6.
- [49] C.-C. Liu, T. Kanekiyo, H. Xu, and G. Bu, "Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy," *Nature Rev. Neurol.*, vol. 9, no. 2, pp. 106–118, Feb. 2013, doi: [10.1038/nrneuro.2012.263](https://doi.org/10.1038/nrneuro.2012.263).
- [50] J. De Velasco Oriol, E. E. Vallejo, K. Estrada, and J. G. Taméz Peña, "Benchmarking machine learning models for late-onset Alzheimer's disease prediction from genomic data," *BMC Bioinf.*, vol. 20, no. 1, pp. 1–17, 2019, doi: [10.1186/s12859-019-3158-x](https://doi.org/10.1186/s12859-019-3158-x).
- [51] T.-T. Nguyen, J. Z. Huang, Q. Wu, T. T. Nguyen, and M. J. Li, "Genome-wide association data classification and SNPs selection using two-stage quality-based random forests," *BMC Genomics*, vol. 16, no. 2, pp. 1–11, 2015, doi: [10.1186/1471-2164-16-S2-S5](https://doi.org/10.1186/1471-2164-16-S2-S5).
- [52] H. Xu, X. Li, Y. Yang, Y. Li, J. Pinheiro, K. Sasser, H. Hamadeh, X. Steven, and M. Yuan, "High-throughput and efficient multilocus genome-wide association study on longitudinal outcomes," *Bioinformatics*, vol. 36, no. 10, pp. 3004–3010, May 2020, doi: [10.1093/bioinformatics/btaa120](https://doi.org/10.1093/bioinformatics/btaa120).
- [53] G. S. Araujo, M. R. B. Souza, J. R. M. Oliveira, and I. G. Costa, "Random forest and gene networks for association of SNPs to Alzheimer's disease," presented at the Adv. Bioinf. Comput. Biol., 2013.
- [54] F. F. Sherif, N. Zayed, and M. Fakhr, "Discovering Alzheimer genetic biomarkers using Bayesian networks," *Adv. Bioinf.*, vol. 2015, 2015, Art. no. 639367, doi: [10.1155/2015/639367](https://doi.org/10.1155/2015/639367).
- [55] F. F. Sherif, N. Zayed, M. Fakhr, M. A. Wahed, and Y. M. Kadah, "Integrated higher-order evidence-based framework for prediction of higher-order epistasis interactions in Alzheimer's disease," *Int. J. Biol. Biomed. Eng., Article*, vol. 11, pp. 16–24, Jan. 2017.
- [56] M. E. Stokes, M. Barmada, M. Kamboh, and S. Visweswaran, "The application of network label propagation to rank biomarkers in genome-wide Alzheimer's data," *BMC Genomics*, vol. 15, no. 1, p. 282, 2014, doi: [10.1186/1471-2164-15-282](https://doi.org/10.1186/1471-2164-15-282).
- [57] J. H. Moore, P. C. Andrews, R. S. Olson, S. E. Carlson, C. R. Larock, M. J. Bulhoses, J. P. O'Connor, E. M. Greytak, and S. L. Armentrout, "Grid-based stochastic search for hierarchical gene-gene interactions in population-based genetic studies of common human diseases," *BioData Mining*, vol. 10, no. 1, Dec. 2017, doi: [10.1186/s13040-017-0139-3](https://doi.org/10.1186/s13040-017-0139-3).
- [58] Y.-C. Chang, J.-T. Wu, M.-Y. Hong, Y.-A. Tung, P.-H. Hsieh, S. W. Yee, K. M. Giacomini, and C.-Y. Chen, "GenEpi: Gene-based epistasis discovery using machine learning," *BMC Bioinf.*, vol. 21, no. 1, Dec. 2020, Art. no. 68, doi: [10.1186/s12859-020-3368-2](https://doi.org/10.1186/s12859-020-3368-2).
- [59] L. Zou, Q. Huang, A. Li, and M. Wang, "A genome-wide association study of Alzheimer's disease using random forests and enrichment analysis," *Sci. China Life Sci.*, vol. 55, no. 7, pp. 25–618, Jul. 2012, doi: [10.1007/s11427-012-4343-6](https://doi.org/10.1007/s11427-012-4343-6).

- [60] M. Aflakparast, H. Salimi, A. Gerami, M. P. Dube, S. Visweswaran, and A. Masoudi-Nejad, "Cuckoo search epistasis: A new method for exploring significant genetic interactions," *Heredity*, vol. 112, no. 6, pp. 666–674, Jun. 2014, doi: [10.1038/hdy.2014.4](#).
- [61] T. A. Shinde and J. R. Prasad, "IoT based animal health monitoring with Naive Bayes classification," *Int. J. Emerg. Trends Technol.*, vol. 1, no. 2, pp. 252–257, 2017.
- [62] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," *Int. J. Comput. Sci. Issues*, vol. 9, p. 272, Sep. 2012.
- [63] M. Pavlou et al., "How to develop a more accurate risk prediction model when there are few events," *BMJ, Brit. Med. J.*, vol. 351, p. h3868, 2015, doi: [10.1136/bmj.h3868](#).
- [64] L. Auria and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis," DIW Berlin, Discuss. Paper 811, 2008, doi: [10.2139/ssrn.1424949](#).
- [65] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecol. Model.*, vol. 203, nos. 3–4, pp. 312–318, May 2007, doi: [10.1016/j.ecolmodel.2006.11.033](#).
- [66] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, no. 4. Cham, Switzerland: Springer, 2006.
- [67] P.-H.-C. Chen, Y. Liu, and L. Peng, "How to develop machine learning models for healthcare," *Nature Mater.*, vol. 18, no. 5, pp. 410–414, May 2019, doi: [10.1038/s41563-019-0345-0](#).
- [68] I. E. Jansen, J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K. Karlsson, S. Hägg, L. Athanasiu, and N. Voyle, "Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk," *Nature Genet.*, vol. 51, no. 3, pp. 404–413, 2019.
- [69] E. W. Steyerberg, *Clinical Prediction Models*. Cham, Switzerland: Springer, 2019.
- [70] A. C. J. Janssens, J. Ioannidis, S. Bedrosian, P. Boffetta, S. M. Dolan, N. Dowling, and I. Fortier, "Strengthening the reporting of genetic risk prediction studies (GRIPS): Explanation and elaboration," *Eur. J. Epidemiol.*, vol. 26, no. 4, p. 313, 2011.
- [71] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Cham, Switzerland: Springer, 2013.
- [72] F. Ahamed and F. Farid, "Applying Internet of Things and machine-learning for personalized healthcare: Issues and challenges," in *Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE)*, Dec. 2018, pp. 19–21.
- [73] J. Futoma, M. Simons, T. Panch, F. Doshi-Velez, and L. A. Celi, "The myth of generalisability in clinical research and machine learning in health care," *Lancet Digital Health*, vol. 2, no. 9, pp. e489–e492, Sep. 2020, doi: [10.1016/S2589-7500\(20\)30186-2](#).
- [74] D. Shillan, J. A. C. Sterne, A. Champneys, and B. Gibbison, "Use of machine learning to analyse routinely collected intensive care unit data: A systematic review," *Crit. Care*, vol. 23, no. 1, pp. 1–11, Dec. 2019.
- [75] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care-addressing ethical challenges," *New England J. Med.*, vol. 378, no. 11, p. 981, 2018.
- [76] E. Vayena, A. Blasimme, and I. G. Cohen, "Machine learning in medicine: Addressing ethical challenges," *PLOS Med.*, vol. 15, no. 11, Nov. 2018, Art. no. e1002689.
- [77] S. J. MacEachern and N. D. Forkert, "Machine learning for precision medicine," *Genome*, vol. 64, no. 4, pp. 416–425, Apr. 2021.
- [78] L. Wang, Y. Wang, and Q. Chang, "Feature selection methods for big data bioinformatics: A survey from the search perspective," *Methods*, vol. 111, pp. 21–31, Dec. 2016.
- [79] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018, doi: [10.1016/j.neucom.2017.11.077](#).
- [80] O. O. Aremu, D. Hyland-Wood, and P. R. McAree, "A machine learning approach to circumventing the curse of dimensionality in discontinuous time series machine data," *Rel. Eng. Syst. Saf.*, vol. 195, Mar. 2020, Art. no. 106706, doi: [10.1016/j.res.2019.106706](#).
- [81] B. Abdulaimma, P. Fergus, C. Chalmers, and C. C. Montanez, "Deep learning and genome-wide association studies for the classification of type 2 diabetes," in *Proc. Int. Joint Conf. Neural Netw.*, 2020, pp. 1–8.
- [82] J. Kim, J. Kim, M. J. Kwak, and M. Bajaj, "Genetic prediction of type 2 diabetes using deep neural network," *Clin. Genet.*, vol. 93, no. 4, pp. 822–829, Apr. 2018, doi: [10.1111/cge.13175](#).
- [83] T. Sun, Y. Wei, W. Chen, and Y. Ding, "Genome-wide association study-based deep learning for survival prediction," *Statist. Med.*, vol. 39, no. 30, pp. 4605–4620, Dec. 2020, doi: [10.1002/sim.8743](#).
- [84] E. Lin, P.-H. Kuo, Y.-L. Liu, Y. W.-Y. Yu, A. C. Yang, and S.-J. Tsai, "A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers," *Frontiers Psychiatry*, vol. 9, Jul. 2018, Art. no. 290, doi: [10.3389/fpsy.2018.00290](#).
- [85] A. Ghorbani, A. Abid, and J. Zou, "Interpretation of neural networks is fragile," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3681–3688.
- [86] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief Bioinform.*, vol. 18, no. 5, pp. 851–869, 2017.
- [87] P. Bellot, G. de los Campos, and M. Pérez-Enciso, "Can deep learning improve genomic prediction of complex human traits?" *Genetics*, vol. 210, no. 3, pp. 809–819, Nov. 2018, doi: [10.1534/genetics.118.301298](#).
- [88] E. Lin, P.-H. Kuo, Y.-L. Liu, Y. W.-Y. Yu, A. C. Yang, and S.-J. Tsai, "A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers," *Frontiers Psychiatry*, vol. 9, p. 290, 2018, doi: [10.3389/fpsy.2018.00290](#).
- [89] R. F. Wolff, K. G. Moons, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett, "PROBAST: A tool to assess the risk of bias and applicability of prediction model studies," *Ann. Internal Med.*, vol. 170, no. 1, pp. 51–58, 2019.
- [90] D. Moher, et al., "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Int. J. Surg.*, vol. 8, no. 5, pp. 336–341, 2010.



ABBAS SAAD ALATRANY received the B.Sc. degree from Al-Mustansiriyah University, Baghdad, Iraq, in 2017, and the master's (Diploma) degree from the University of Leicester, Leicester, U.K., in 2020. He is currently pursuing the Ph.D. degree with Liverpool John Moores University, Liverpool, U.K. His research interests include deep learning, machine learning, and bioinformatics.



ABIR JAAFAR HUSSAIN (Member, IEEE) received the Ph.D. degree from The University of Manchester (UMIST), U.K., in 2000, with a thesis title polynomial neural networks for image and signal processing. She is currently a Professor in biomedical science and a member of the Applied Computing Research Group, Faculty of Engineering and Technology. She has published numerous refereed research papers in conferences and journals in the research areas of neural networks, signal prediction, telecommunication fraud detection, and image compression. She has worked with higher order and recurrent neural networks and their applications to e-health and medical image compression techniques. She has developed with her research students a number of recurrent neural network architectures. She is also a Ph.D. Supervisor and an External Examiner for research degrees, including M.Phil. and Ph.D. degrees. She is one of the initiators and the chairs of the Development in e-Systems Engineering (DeSE) series.



JAMILA MUSTAFINA received the Ph.D. degree, in 2007. She is currently a Professor and the Leader of the iTech Research Laboratory, Kazan Federal University. She received a Full Professorship, in 2012. She has published over 100 peer-reviewed scientific articles, four books, and more than eight book chapters. She has graduated five Ph.D. students and is currently supervising three Ph.D. students and two postdoctoral students. Her research interests include the interdisciplinary approach towards the improvement of the social spheres of human life using technology and applied computing. She is also acting as the Deputy Editor-in-Chief of the peer-reviewed scientific journal "Education and Self Development." She has been awarded a number of research grants nationally and internationally. She is the International Committee Chair for IEEE International Conference Series on Developments in eSystems Engineering DeSE (www.dese.org.uk).



DHIYA AL-JUMEILY (Senior Member, IEEE) is currently a Professor in artificial intelligence and the President of eSystems Engineering Society. He has extensive research interests covering a wide variety of interdisciplinary perspectives concerning the theory and practice of applied artificial intelligence in medicine, human biology, environment, intelligent community, and health care. He has published well over 300 peer reviewed scientific international publications, 12 books and

14 book chapters, in multidisciplinary research areas, including machine learning, neural networks, signal prediction, telecommunication fraud detection, AI-based clinical decision-making, medical knowledge engineering, human-machine interaction, intelligent medical information systems, sensors and robotics, wearable and intelligent devices, and instruments. But, his current research passion is decision support systems for self-management of health and medicine. He has successfully supervised over 20 Ph.D. student's studies and has been an external examiner to various U.K. and overseas universities for undergraduate, postgraduate, and research degrees. He is also a successful Entrepreneur. He is the Head of Enterprise for the Faculty

of Engineering and Technology. He has been awarded various commercial and research grants, nationally and internationally, over £5M from overseas research and educational partners, U.K., through British Council and directly from industry with portfolio of various knowledge transfer programs between academia and industry. He has been the Founder and the General Series Chair of the IEEE International Conference on Developments in eSystems Engineering DeSE (dese.org.uk), since 2007. He has been actively involved as a member of editorial board and review committee for a number peer reviewed international journals. He is on program committee or as the general chair for a number of international conferences. In December 2020, he was promoted and appointed by The Queen to the Most Excellent Order of the British Empire, OBE—Ordinary Officers of the Civil Division of the said Most Excellent Order for the Services to Scientific Research. (<https://www.thegazette.co.uk/notice/3703557>).

...