

Received October 20, 2021, accepted November 5, 2021, date of publication November 8, 2021, date of current version November 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3126777

# Predicting the Risk of Alcohol Use Disorder Using Machine Learning: A Systematic Literature Review

ALI EBRAHIMI<sup>1</sup>, UFFE KOCK WIIL<sup>1</sup>, THOMAS SCHMIDT<sup>1</sup>, AMIN NAEMI<sup>1</sup>, ANETTE SØGAARD NIELSEN<sup>2</sup>, GHULAM MUJTABA SHAIKH<sup>3</sup>, AND MARJAN MANSOURVAR<sup>4</sup>

<sup>1</sup>SDU Health Informatics and Technology, The Maersk Mc-Kinney Møller Institute, University of Southern Denmark, 5230 Odense, Denmark

<sup>2</sup>Unit for Clinical Alcohol Research, Clinical Institute, University of Southern Denmark, 5230 Odense, Denmark

<sup>3</sup>Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan

<sup>4</sup>Department of Mathematics and Computer Science (IMADA), University of Southern Denmark, 5230 Odense, Denmark

Corresponding author: Ali Ebrahimi (aleb@mmmi.sdu.dk)

This work was supported by the European Union (EU) Interreg 5a Deutschland (DE)-Denmark (DK) Project Access & Acceleration (<https://www.interreg5a.eu/dk/blog/projekt/access-acceleration/>).

**ABSTRACT** The number of deaths caused by alcohol-related diseases may be reduced by predicting alcohol use disorder (AUD). Many researchers have worked on AUD prediction using machine learning (ML) techniques. However, to the best of our knowledge, there is a lack of a comprehensive systematic literature review (SLR) that summarizes the existing studies on AUD prediction using ML in the last ten years. To address this knowledge gap, this article provides an SLR of academic articles on AUD prediction using ML techniques dated from January 2010 to July 2021. This SLR highlights technical decision analysis related to five aspects: data collection site, characteristics, and type of dataset; data sampling and data pre-processing techniques; feature types and feature engineering techniques; and characteristics of ML techniques and evaluation metrics. Six bibliographic databases were searched, and the identified studies were rigorously reviewed based on the above five aspects. In the selected studies, public datasets were not used very often for AUD prediction. Given that, the current paper identified two different types of data collection sites for review. Imbalanced class distribution in datasets was the primary focus of the pre-processing and sampling steps. Various features, including demographics, family history, drinking behaviour, and electronic health records, were introduced as the more widely used AUD prediction features. The filter, wrapper, and embedded methods were identified as the primary feature selection methods. Support vector machine was the most widely employed algorithm for predicting AUD; however, the lack of deep neural network techniques is evident in this field. Moreover, considering gender disparities, early detection of AUD, and identifying trajectories towards AUD are suggested for future work. For the purpose of evaluating the performance of the prediction approaches, most studies considered the overall accuracy and the area under the receiver operating characteristic curve. Nevertheless, external validation was not performed in any of the selected studies. This paper also discusses challenges and open issues of AUD prediction for future research. This SLR represents a valuable resource for scholars investigating the prediction of AUD.

**INDEX TERMS** Systematic literature review, alcohol use disorder, prediction, machine learning, supervised learning.

## I. INTRODUCTION

Alcohol use disorder (AUD) is a broad term used to refer to problems caused by alcohol consumption. Affected

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

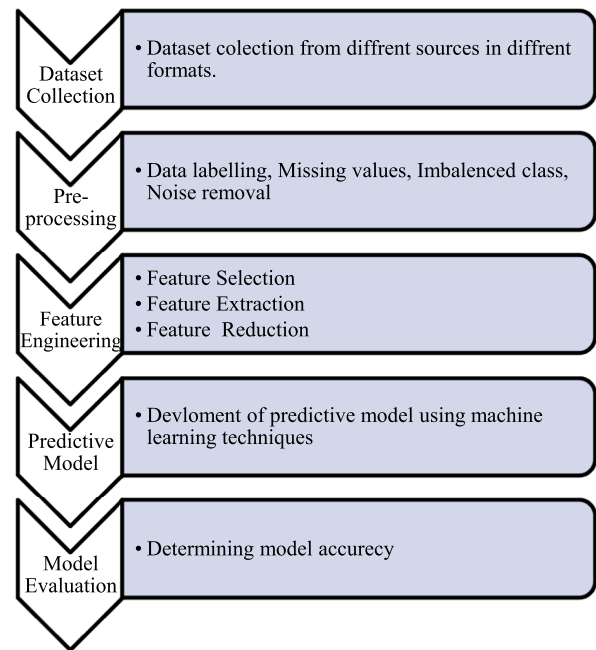
individuals tend to have a lack of control over their alcohol consumption; they continue to drink despite the serious adverse effects of alcohol on their health and on the lives of others, including family, friends, and co-workers. Extreme alcohol use leads to various diseases, such as liver cirrhosis, chronic pancreatitis, upper gastrointestinal

cancers, cardiomyopathy, polyneuropathy, and dementia. High alcohol consumption also affects decision-making abilities, in addition to accelerating cases of violence and harmful behaviours. The World Health Organization (WHO) report of 2014 mentioned that approximately 5.9% or 3.3 million deaths were caused by alcohol misuse [1]; alcohol misuse is also the world's fifth leading cause of death [2] and the leading risk factor for premature death and disability.

According to the diagnostic criteria stated in the Diagnostic and Statistical Manual of Mental Disorders (5th edition), 36.0% of adult males and 22.7% of adult females in the US, from 2012 to 2013, fulfilled the criteria for AUD at some point in their lives. In addition, 17.6% of men and 10.4% of women in past years were also noted to have fulfilled these criteria [3]. In Europe, 3.5% of people aged 18–64 years have been estimated to be alcohol dependent, with 11.1% of them being assessed as heavy drinkers [4]. Although AUD contributes to the second-highest level of all diseases linked to mental disorders after depression [5], treatment rates have been low, and most patients suffering from AUD have never received specialized treatment for their addiction [6]. Different studies have reported different reasons for the low treatment rate of AUD in Europe. [7] claimed that primary care physicians did not recognize AUD, delaying treatment. The reason for this is that conventional methods tend to detect alcohol-related problems through self-test reports [8]. Patients' dishonesty, lack of memory, the taboo surrounding the issue, and various other reasons may result in self-test reports that are used in the diagnosis of AUD being inaccurate.

Numerous factors are related to the increased risk of AUD. These factors have helped scientists predict AUD, for example, a history of alcoholism in biological family members [9], psychological factors, such as level of stress [10], and personality disorder [11], behavioural factors, such as gambling problems [12], and social influences [13]. Health records in hospitals also contain a substantial amount of information that can be related to AUD and thus potentially useful for AUD prediction. Advancements in machine learning (ML) methods may make the prediction of AUD based on health records even more precise and thereby helpful for staff medical decision-making. Through the development of ML methods, researchers can identify target groups for AUD interventions [14].

ML methods can classify documents or reports into one or more predefined categories [15]. These methods have been documented in numerous clinical documents and reports, such as electroencephalogram (EEG) reports, radiology reports, electronic health records (EHRs), and biomedical documents. These documents show that ML methods were used for detecting cancer stages [16], identifying paediatric traumatic brain injury [17], and predicting AUD [10] and [18]. Various types of ML methods exist, such as unsupervised machine learning (UML), semi-supervised machine learning (SSML), and supervised machine learning (SML). The primary goal of SML is to build an efficient mapping



**FIGURE 1.** Overall method of building a predictive model using machine learning techniques.

function to accurately predict or classify a dependent variable from the independent variables [19]. In contrast to SML, UML does not require a specific outcome variable. It is primarily used for clustering and dimensionality reduction. The aim of SSML methods, in comparison, is to optimize classification accuracy using only a few labelled records [20].

The basic methodology for developing a predictive model using ML usually consists of five phases: data collection, pre-processing, feature engineering, predictive model development, and model evaluation (see Figure 1). A dataset, including clinical reports and survey questionnaires, is initially collected from one source or multiple sources. Thereafter, these reports are labelled by experts or through clustering techniques of specific classes (e.g., AUD positive or AUD negative). Next, the pre-processing technique is employed to remove the dataset forms that are unnecessary or are noisy information. After the pre-processing stage, a numeric master feature vector is constructed as a result of feature engineering. This is done by selecting or extracting the most discriminative features from the dataset. The primary input of ML algorithms used for developing a predictive model or a classification model is the master feature vector. A variety of validation methods, precision, recall, F1 score, overall accuracy, receiver operating characteristic (ROC) curve, and area under the ROC curve (AUC), can then be employed to evaluate and validate the constructed ML algorithms [21].

In recent years, several studies have examined the prediction of AUD in particular. These studies have primarily provided reviews on AUD, predictions of AUD (e.g., duration of treatment and intervention type) [22]–[25], predictions of other addictions (smoking, gambling, and cocaine) [26],

and a scoping study on prediction of AUD [27]. However, the important aspects of AUD prediction using ML techniques, particularly the data collection process, types and characteristics of the datasets, sampling and pre-processing techniques, types of features and variables, the process of handling feature redundancy and high dimensionality, types of ML algorithms used, and evaluation processes and performance metrics, have not been systematically reviewed in the past 10 years. Therefore, the current systematic literature review (SLR) strives to evaluate the academic articles that explored AUD prediction using ML methods in the period from January 2010 to July 2021. In particular, the current SLR article aims to maximize the five general phases of building a predictive model or a classification model (Figure 1). To the best of our knowledge, this article is the first SLR to focus on AUD prediction over the last decade.

The primary contributions of this SLR are as follows:

- The paper reviews studies from the past decade using five different dimensions.
- It explores the different types of datasets used in predicting AUD using ML techniques; it also reviews the data pre-processing and sampling techniques that have been used to prepare datasets for AUD prediction. Moreover, it analyses the types of features and variables that contribute to the development of the AUD and the techniques used for extraction, selection, and the reduction of intended features, as well as ML algorithms that have been used for AUD prediction and their performances.
- Finally, it outlines open issues and research challenges related to ML-based AUD prediction.

The structure of this SLR is as follows: Section II presents the methodology applied in this study to systematically select the primary studies. Section III presents the comprehensive review and findings of selected studies. Section IV presents discussion on the findings. Section V offers some suggestions and research directions for future work. Finally, Section VI concludes this review.

## II. METHODS

The methodology of this paper was inspired by the SLR guidelines proposed by Kitchenham and Charters [28]. As shown in Figure 2, we adopted three steps from this methodology: planning, implementation, and reporting. The planning phase was discussed in Section I above. The implementation phase will be discussed in Section II, parts A to C. Reporting reviews that include critical analysis and data synthesis will serve as the final phase discussed in Section III.

### A. SEARCH STRATEGY AND SEARCH RESULTS

We shaped the search strategy stage into a five-group block, as shown in Table 1. Anne Faber Hansen (a professional librarian at the University of Southern Denmark) and Ali Ebrahimi (the corresponding author) prepared a list of potential keywords and formulated queries to search the literature. The search tools Medical Subject Headings (MeSH) and Boolean logic (OR, AND) were used to narrow and expand

the search terms. The formulated queries were applied to the keywords, title, and abstract of articles to identify potential journal and conference articles published (in English) from January 2010 to July 2021. Six high-quality academic databases in medical and engineering fields, including Medline, Embase, Inspec, ScienceDirect, Web of Science, and IEEE Xplore, were considered for the extraction of relevant literature. More details on query formation for each academic dataset can be found in the supporting materials.

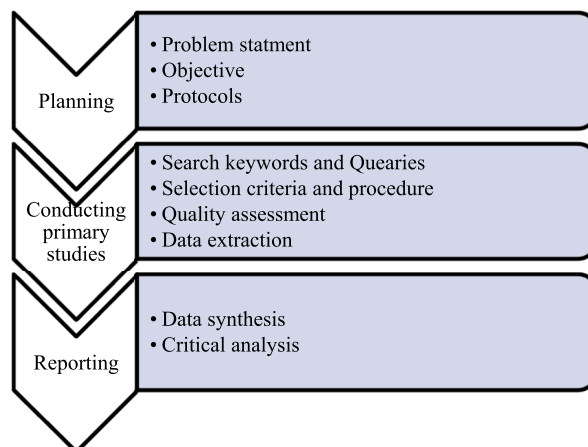


FIGURE 2. Systematic review steps (guidelines) adapted from Kitchenham and Charters [28].

A total of 3,736 studies were retrieved using the above-mentioned search query. Next, all extracted data were stored in Endnote [29], and duplicate studies were removed, which resulted in the exclusion of 1,381 studies. The details of the search results are presented in Figure 3.

### B. DATA SCREENING AND SELECTION CRITERIA

Twelve principal reasons were used as the basis of the inclusion and exclusion criteria, which are listed in Table 2. The most important and vital criteria were based on the primary aim and the goal of the studies. First, the primary aim of some studies was not AUD prediction. For example, one study [30] proposed a novel approach for automatically selecting features from a multivariate time series using a trace-based class separability criterion. Although an EEG dataset containing alcoholic and non-alcohol patients was used, the study's primary aim was not to analyse patients for AUD prediction. Second, many studies aimed to detect or make a diagnosis of AUD or alcoholism [31]. Prediction is an analysis based on a pattern from past and present data and information, whereas detection is mining for the extraction of information from a dataset when it is being processed. For example, processing past clinical records of patients can be used for the prediction of AUD. However, the detection of AUD can be accomplished by the analysis of a survey questionnaire [32]. Given that the primary focus of such studies was AUD detection rather than prediction, these studies were also excluded [33]–[36].

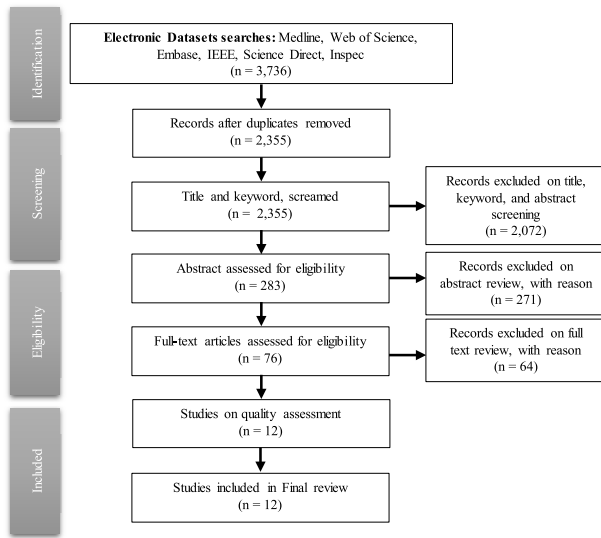


FIGURE 3. Data selection process based on the PRISMA flowchart [38].

Third, some studies, for example, [37] and [38], focused on prediction, but it was geared toward the abuse of other substances (e.g., smoking, cocaine and gambling) instead of alcohol. Fourth, some studies considered ML techniques for assessing alcoholism treatment, for example, [39]. Fifth, a few studies, such as [40], used ML techniques to assess the prediction of specific diseases of patients with an alcohol drinking problem. Other studies, such as [41], used ML techniques to analyse patterns in psychiatric disorders. All of these were excluded, including studies that used ML techniques on animals. Studies where full texts could not be accessed at the time of the SLR were also excluded. Moreover, a few studies used alcoholic and non-alcoholic patient datasets to indicate the performance of improved ML techniques [42]. These studies were also removed since they were not aimed at predicting AUD.

Following duplicate removal, the remaining 2,355 articles were screened based on the inclusion and exclusion criteria using the titles, abstracts, and keywords proposed by three authors (AE, AN, and GMS). Discrepancies regarding whether the articles should be included or excluded were resolved through majority voting. In the case of a tie, the authors (AE, UKW, ASN, and MM) made the final decision. Figure 3, based on PRISMA guidelines [43], illustrates the screening of articles. The article title and keyword screening resulted in the inclusion of only 283 articles out of 2,355 articles. Next, these 283 articles were included in the abstract and full-text screening phase. This was performed by three of the authors (AE, AN, and GMS) based on the inclusion and exclusion criteria (see Table 2). Finally, twelve articles were retained.

C. QUALITY ASSESSMENT AND DATA EXTRACTION

The quality assessment criteria based on the primary objective of this study were adjusted to identify the final quality

TABLE 1. Selected keywords and search strategies.

Selected keywords in the different groups	
<b>Group 1- Keywords related to the Alcohol Use Disorder</b>	
<ul style="list-style-type: none"> <li>• Alcoholism.</li> <li>• Alcohol Drinking.</li> <li>• Alcohol use disorder.</li> <li>• Alcohol abuse.</li> <li>• Alcohol-Related Disorders.</li> <li>• Drinking initiation.</li> <li>• Alcohol dependence.</li> </ul>	
<b>Group 2 - keywords related to prediction using machine learning techniques</b>	
<ul style="list-style-type: none"> <li>• Early diagnosis.</li> <li>• Prediction.</li> <li>• Prognosis.</li> <li>• Machine learning.</li> <li>• Supervised machine learning.</li> <li>• Unsupervised machine learning.</li> </ul>	
<b>Group 3 – Publication years</b>	
<ul style="list-style-type: none"> <li>• January 2010 to July 2021</li> </ul>	
<b>Group 4 – Document types</b>	
<ul style="list-style-type: none"> <li>• Journal and Conference Articles</li> </ul>	
<b>Group 5 – Language</b>	
<ul style="list-style-type: none"> <li>• English</li> </ul>	
<b>Final Search Query</b>	
<ul style="list-style-type: none"> <li>• (Group 1) AND (Group 2) AND (Group 3) AND (Group 4) AND (Group 5)</li> </ul>	

TABLE 2. List of the inclusion and exclusion criteria.

Inclusion Criteria
1. The main goal of the article must be the prediction of AUD.
2. The paper should describe the use of machine learning algorithms for the prediction of AUD.
3. The paper should have reported research to finding the features that may facilitate the prediction of AUD using machine learning algorithms.
4. Clinical reports, medical documents or other type of dataset should have been used.
5. The paper should be written in English. Nonetheless, papers on the processing of non-English documents were included.
6. Article must be published between 2010 and 2021. Article is either conference proceeding or journal article.
Exclusion Criteria
1. Studies which aimed the detection or diagnosis of the AUD
2. Studies which assessed the outcome of alcoholism treatment or the quality of post treatment.
3. Studies which aimed the prediction of other substance abuse than alcohol.
4. Studies which aimed to predict alcoholism lapses.
5. Studies on animals.
6. Studies which the full text is not available.

of the selected studies. A quality checklist of questions is shown in Table 1 in the Appendix. Three of the authors (AE, AN, and GMS) assessed the selected studies using the ‘Yes’ or ‘No’ response, which carried the weights of ‘1’ and ‘0’, respectively, for evaluating the selected studies. The final results of this evaluation were then discussed using the Delphi

method [44], and the threshold for inclusion of any study was set at five. All selected studies carried a score of six or above (50% or above); therefore, they were all included in the review. Three tables based on the five previously mentioned aspects were created to tabulate the twelve selected primary studies. These three tables include Table 3, which briefly describes the datasets, data collection sites, pre-processing, and sampling techniques. Table 4 describes the characteristics of the features and feature engineering process, and Table 5 describes the ML approaches used in the primary studies and the performance metrics used to evaluate them. In Section III, the extracted information from the primary studies is critically reviewed.

### III. REVIEW ON THE PREDICTION OF ALCOHOL USE DISORDER

Data generated from the twelve studies were critically reviewed based on the five different aspects mentioned above. Subsection A presents a review of the different datasets used for the prediction of AUD as well as the pre-processing techniques used to prepare the dataset for further analysis. Subsection B presents a review on the feature engineering techniques and the types of features used for the prediction of AUD, and Subsection C presents a review on the various ML techniques used for the prediction of AUD and the different performance metrics used to evaluate these techniques.

#### A. DATASET CHARACTERISTICS AND THEIR COLLECTION SITES, PRE-PROCESSING AND SAMPLING TECHNIQUES

ML methods have been applied to several types of datasets, such as survey questionnaires, clinical and educational datasets, and historical personal data, for the prediction of AUD. As shown in Table 3, most of the studies considered data from survey questionnaires [10], [45]–[52]. However, other sources of data, such as a combination of EEG samples and family history (FH) [53], a combination of students' health and clinical records, and educational datasets [54], EHRs [49], [50] and [52], genetic data [51], and MRI [47] and [48], were used in other studies. In terms of the collection sites, the primary selected studies can be categorized into single and multiple data collection sites. The category of each dataset is also shown in Table 3. Among all studies, there were two datasets that were collected from a single site, one used by [45] and [10] and the other used by [54]. On the other hand, another study [55] combined datasets from two different public schools in Portugal, while another [53] collected data from six different sites. More information regarding the datasets, such as year of study, age of participants, gender, and period of follow-up, can be found in Table 3.

One of the primary processes of the data mining task is data pre-processing. This is a process of formatting data so that they are usable by the ML algorithms. Depending on the dataset, pre-processing tasks differ, but all aim to improve the quality of the final predictive model by cleaning the dataset prior to model development. Data usually contain a high level of noise and sparsity, which can exist among medical

measures (such as blood pressure), clinical scores (heart rate), and clinical codes (such as diagnosis ICD codes). In this respect, noisy data can emerge when participants provide invalid responses or when the data are incorrectly encoded into the spreadsheets or databases. Therefore, comprehensive pre-processing of the datasets is necessary to form the variables of the collected data to be more meaningful for predictive models based on ML algorithms.

In the related literature, different pre-processing techniques were used to overcome these issues. Two studies, [45] and [10], applied the K-Medoids clustering algorithm [56] to label the datasets. However, some studies [48]–[52], [54] employed the results of the questionnaires to label their datasets. In terms of handling missing values, two studies, [10] and [47], used imputation techniques, one [51] applied the interpolation method, and two, [49] and [45], removed patients and participants with missing values from the dataset.

In terms of dealing with imbalanced class distribution in datasets, several [48]–[50], [55] studies prevented class imbalance in the dataset. Most of these studies applied synthetic minority oversampling technique (SMOTE) [57]-based techniques to handle this problem. Those studies that considered the data scaling process [10], [45], [48]–[50] used the normalization technique for rescaling all feature values to a range of between 0 and 1 [58].

#### B. REVIEW OF FEATURE ENGINEERING TECHNIQUES AND TYPE OF FEATURES

One of the most important aspects of field classification and prediction when using ML techniques is selection of the right feature or variable. A feature is an individual measurable property of the process being observed. Using a set of features, some ML algorithms can perform classification and prediction tasks [59]. In one study [27], variables such as FH and psychological and genetic factors were the features most widely used for the prediction and detection of AUD. Demographic features, including age, sex, family status, education level, income, and occupation, constitute another set of features widely used in ML studies [60]. A summary of the types of features used in the studies identified in the present systematic review is shown in Table 4.

Two studies, [45] and [10], employed factors such as drinking motives, academic performance, psychological factors, demographic items, and drinking behaviour during the last month as features. In comparison, another study [46] considered factors such as a history of drug use at some time in their lives, frequency of alcohol use in their lifetime, the past 12 months and past 30 days, self-reported impulsivity, motor impulsivity, choice impulsivity, self-reported sensation seeking, and executive function. One study [53] considered FH and EEG signals while another [54] considered demographics together with clinical records and a history of drug and alcohol use, and another [55] considered social, demographic, and school-related variables. One study [47] also considered demographic variables and FH of

**TABLE 3. Datasets and pre-processing description of the related literature.**

Number Authors [Reference number]	Year	Size of dataset	Period of follow-up	Female (%)	Age group	Source of data	Data collection site	Pre-processing
1. Zuba, et al. [45] 2. Bi, et al. [10]	2012 2013	350	30 days	52	College age	Survey questionnaire	Single	K-Medoids clustering used for labelling the dataset. Missing values handled by Multiple Imputation Package. Normalization technique for feature standardization.
3. O'Halloran, et al. [46]	2018	106	12 months	70	18–21 years old	Survey questionnaire	Multiple	NA
4. Kinreich, et al. [53]	2019	659	EEG signals recorded for 4 minutes	55	12–30 years old	Combination of EEG measures and genomic data	Multiple	Low-pass filter and digital notch filter was applied to remove line noises in the EEG.
5. Ngo, et al. [54]	2019	179,156	Six years academic	59	16–49 years old	Combination of Student Information System, Student Health Record Dataset, the ED's Patient Registration System, Clinical Data Repository and Incident Management Response System	Single	Stepwise multivariable logistic regression for calculation of the coefficients and odds ratios of each risk factor.
6. Sisodia, et al. [55]	2019	662	12 months	NA	15–22 years old	Survey questionnaire and students' school scores	Multiple	Synthetic Minority Oversampling technique was used to handle imbalance in the dataset.
7. Gowin, et al. [47]	2020	177	Different time frames considered in each source of data. Subjects spent two days at the hospital to complete samples.	27.7	22–35 years old	Combination of questionnaires, behavioural tasks, electronic health records, and MRI	Single	Missing values were handled by imputation using a proximity matrix.
8. Silveira, et al. [48]	2020	392	4 years	55	12–22 years old	Combinational of questionnaires and MRI	Single	MRI dataset pre-processed by slice time correction, data realignment to the first functional volume, field-map correction, individual rigid T1-to-EPI co-registration, segmentation and normalization of the co-registered T1 and EPI volumes in the MNI stereotactic space, and spatial smoothing using an isotropic Gaussian kernel with 8 mm FWHM. Physiological artefacts and residual subject movements were removed from the data using regression of realignment estimates to handle noise. Synthetic Minority Oversampling technique (SMOTE) and SMOTE Nominal Continuous were used to handle imbalance in the dataset.
9. Chen, et al. [51]	2020	317		Zero (All male)	24–59 years old	Combinational of questionnaires and genetic data.	Single	Missing values handled by interpolation.
10. Ebrahimi, et al. [49]	2020	2,571	18 months	53.5	18–101 years old	Combinational of questionnaires and electronic health records	Single	Categorical data was handled using one-hot-encoding. Imbalanced class distribution was handled using Synthetic Minority Oversampling. Missing values were handled by removing patient sample data with missing values. A standard normalization technique was used for data scaling.
11. Bonnell, et al. [52]	2020	43,545	17 years	51	Median age of unhealthy and low risk drinkers was 35 and 53 years old, respectively.	Combinational of questionnaires and electronic health records	Single	Radom Forest was modified to handle missing values.
12. Ebrahimi, et al. [50]	2021	2,571	18 months	53.5	18–101 years old	Combinational of questionnaires and electronic health records	Single	Imbalanced class distribution was handled using Synthetic Minority Oversampling combined with Wilson's Edited Nearest Neighbor Rule (SMOTEENN). Categorical data were handled using one-hot-encoding. A standard normalization technique was used for data scaling.

patients along with their psychological and health statuses. Among all selected studies, only one [51] considered genetic variables along with demographic variables and educational background. Among studies that considered EHRs as the primary features, two, [50] and [49], considered discharge data based on ICD-10 codes, while one [52] considered laboratory results.

One of the critical steps in developing a predictive model using ML techniques is feature engineering [61] and [62]. In the literature, this process is divided into two categories:

feature extraction and feature selection [63]. Their aims are to solve problems such as high dimensionality, data sparsity, feature redundancy, a vast number of features, and high noise levels. Feature extraction places project features into a new feature space with lower dimensionality, and newly constructed features are usually comprised of combinations of the original features.

Multiple correspondence analysis (MCA), principal component analysis (PCA), linear discriminant analysis (LDA), and canonical correlation analysis (CCA) are examples of

TABLE 4. Features and feature selection schemes used in the selected studies.

Number Authors [Reference number]	Features	Feature engineering type
1. Zuba, et al. [45]	School information: School year and semester, GPA, Demographic variables: Gender, Race, Dating status, Drinking behaviour, Personality, Social and Family support	Feature extraction: Multiple correspondence analysis to handle categorical variables, and principal component analysis Feature selection: Embedded method (Recursive feature elimination technique using support vector machine and simultaneously building a classifier to achieve the best validation performance)
2. Bi, et al. [10]	Drinking motives, academic performance, personality, recent depression and anxiety symptoms, negative life events	Feature extraction: Multiple correspondence analysis to handle categorical variables, and principal component analysis Feature selection: Wrapper method (1-norm support vector machine applied to identify the personality factors which were highly correlated between the classes)
3. O'Halloran, et al. [46]	Impulsivity variables	Feature extraction: Principal component analysis
4. Kinreich, et al. [53]	Family history: European American and African American ancestry Demographic features: Age, Gender Electroencephalography	Feature selection: Filter method (Least absolute shrinkage was used to generate coefficient and all features with a non-zero coefficient were retained for the analyses. The reduced set of most discriminant features was then fed into the classifier)
5. Ngo, et al. [54]	Demographic variables: Gender, Age, Ethnicity, Citizenship Academic affiliation, Clinical risk markers, Violence involvement	Feature selection: Embedded model (First the risk markers were identified based on the p-value rate and then selected features fitted into Stepwise multivariable logistic)
6. Sisodia, et al. [55]	Academic grades, social, demographic and school-related variables	NA
7. Gowin, et al. [47]	Psychological variables: Pittsburgh sleep quality index, stress, anxiety, social support, depressive episodes, conduct problems, depressive symptoms, emotional state Substance use history: 7-day retrospective reports for alcohol and tobacco use Health: Haematocrit level, blood pressure, thyroid stimulating hormone level, haemoglobin A1C percentage, weight, height, body mass index, and sleep quality Demographics variables: age, race, ethnicity, handedness, income, employment status, education, and relationship status Family history: history of psychosis, depression, bipolar disorder, anxiety, drug or alcohol problems, Alzheimer's disease, Parkinson's disease, or Tourette's syndrome	Feature selection: Filter methods (t-test and permutation testing were employed to assess how the psychological and neural variables differ in the models)
8. Silveira, et al. [48]	Demographic variables: age, sex, ethnicity, parental education, AUD family density; Substance use history; Eleven functional brain connectivity measures; Scanner type	NA
9. Chen, et al. [51]	Single nucleotide polymorphisms of HTR3A and HTR3B, age, education level and marital status	Feature selection: Filter method (Pearson correlation coefficients)
10. Ebrahimi, et al. [49]	Demographic variables: age and sex Electronic health records: Length of stay at the hospital, visitors of emergency departments, ICU (yes/no), ICD-10 codes indicating the reasons patients visited the hospital	Feature selection: Filter methods (chi-squared and frequency of variables)
11. Bonnell, et al. [52]	Demographic variables: sex, age History of smoking Electronic Health records: clinical information, chemistry, liver function, haematology, lipids	Feature selection: Filter method (Information Gain)
12. Ebrahimi, et al. [50]	Demographic variables: age and sex Electronic health records: Length of stay at the hospital, visitors of emergency departments, ICU (yes/no), ICD-10 codes indicating the reasons patients visited the hospital.	NA

feature extraction and dimensionality reduction techniques. In the literature, [45], [10], and [46] employed feature extraction techniques to overcome dimensionality problems in their datasets.

Feature selection approaches aim to select a small subset of features that minimize redundancy and maximize relevance to the target. These approaches can be divided into three groups: filter, wrapper, and embedded [63]. The filter

method selects features based on the statistical characteristics of features [64]. In the filter method, features are ranked according to specific criteria. They are then featured using the highest ranking to compose a feature subset. Some of the most common filter methods are information gain, mutual information, chi-square, Fisher score, and relief [65]. The filter method does not rely on the performance of the ML algorithms; therefore, the biases of classifiers cannot be taken into account in the classifier FS method [66]. In the primary studies, one [53] used the least absolute shrinkage [67], one [47] employed t-test and permutation testing, one [51] applied Pearson correlation coefficients, one [49] applied chi-square and considered the frequency of ICD-10 codes, and one [52] employed information gain as filter selection methods to select the best features.

The wrapper method basically uses the accuracy of the predictive model to evaluate the quality of the selected features. In the wrapper method, all subsets of the features are first searched, and then their importance is evaluated based on the performance of the classifier. These steps will be repeated until the best features remain [68]. The primary disadvantage of this method is that they have to run the classifier many times to identify the best feature, which makes it computationally expensive for a dataset with a large number of features [69]. Among all identified studies, one [10] employed a 1-norm support vector machine as a wrapper method for selecting the best features. Finally, an embedded method that uses a filter method to pick candidate features and then employs a wrapper method to select the best features based on the accuracy of the ML model [70]; consequently, the disadvantages of filter and wrapper methods are mitigated by using such a hybrid method. In the literature, one study [54] performed the embedded method by applying stepwise multivariable logistics regression [71], [72]. Moreover, one [45] applied recursive feature elimination techniques as an embedded method to select the best feature set. More information regarding the feature engineering process of each study is provided in Table 4.

### C. REVIEW OF PREDICTIVE MACHINE LEARNING APPROACHES AND PERFORMANCE METRICS

ML is an important domain of artificial intelligence; it enables machines to learn and act on specific tasks. As discussed earlier, SML consists of techniques and algorithms that can predict future events or classify data by learning the patterns of existing data. Generally, discriminative and generative algorithms can be employed in SML models [73]. Logistics regression is an example of the SML algorithm, which is widely used in classification problems; by fitting a curve line between variables, it attempts to build a classifier [74]. Another famous classification algorithm is the support vector machine (SVM) algorithm [75], which attempts to identify the best data classifier for the given data; it can also achieve excellent predictive performance. K-nearest neighbor (KNN) is an algorithm that attempts to classify an unknown instance based on its neighbours' classification [76]. This means that

it labels targets by checking class labels of the K nearest points in the feature space. The decision tree is another powerful algorithm for modelling data that benefits from tree-like structures to classify given data [77]. Random forest is an example of ensemble learning that consists of several decision trees [78]. A neural network (NN) is another ML algorithm that has been identified as a powerful method for some complex ML tasks [79]. In the NN algorithms, a network of cells is produced, and connections between the cells are adjusted in a way that results in networks that can learn the structures of the training data. NN can extract a higher level of features from input data.

In the literature related to the prediction of AUD, SVM was the most popular among researchers, followed by logistic regression. As shown in Table 5, [10], [45], [47]–[49], [51]–[53], [55] employed SVM to build predictive models. Logistic regression was identified as the second most common algorithm among the literature related to the prediction of AUD [10], [48], [52], [54]. Some studies aimed to compare different algorithms to identify the one with the highest accuracy. In two studies, [55] and [52], six different ML algorithms were compared to identify the best accuracy for their study. This was also done by other studies [10], [47], [48], [51] on a smaller scale by comparing two, three, and five different algorithms. Three studies, [51], [52], and [50], employed different types of NN algorithms for the prediction of AUD, in which one [50] developed a deep NN (DNN) application for the prediction of patients with AUD, one [52] used several ML algorithms, including NN, to identify unhealthy drinkers, and another [51] developed the application of NN as well as a convolutional neural network (CNN) combined with long- and short-term memory (LSTM) to classify alcohol-dependent and non-alcohol-dependent groups.

A constructed predictive ML model's performance can be measured using several evaluation metrics, such as accuracy, F1 score, ROC curve and AUC ratio, sensitivity, specificity, and macro- and micro-averaging of accuracy. Four primary ratios can be used to compute the value of these metrics, including true positives (TPs), false-positives (FPs), false negatives (FNs), and true negatives (TNs). A detailed discussion on these performance metrics can be found in a previous study [80]. Different types of performance assessments were used to determine the predictive performance of the constructed ML models in the selected primary studies. The commonly used performance metrics were the ROC curve and AUC ratio, which were used by eight studies. Table 5 shows the ML algorithms used in the experiments of the relevant identified studies together with the performance measures, as well as the proportion of training and test sets.

### IV. DISCUSSION

This systematic review examined the application of ML-based methods for the prediction of AUD. The primary studies used for this SLR were extracted from six academic databases based on different exclusion and inclusion criteria



**TABLE 5. Machine learning techniques, performance metrics, and other related characteristics among the reviewed articles.**

Number Authors [Reference number]	Machine learning techniques	Modelling and validation set	Evaluation metrics	Results
1. Zuba, et al. [45]	Six different classifiers based on linear Support Vector Machine	2/3 of dataset reserved for training and 1/3 reserved for test sets	The area under the curve was used to calculate the performance of the classifiers	Heavy drinking classifier and low-risk drinking classifier for females performed the best among all models. The heavy drinking and low-risk drinking male classifiers were moderately accurate.  The moderate drinker classifiers for both males and females performed poorly and were closer to a random-guessing classified.
2. Bi, et al. [10]	In total, three models were developed, two Support Vector Machine and one Logistic Regression	2/3 of dataset reserved for training and 1/3 reserved for test sets	The area under the curve was used to calculate the performance of the classifiers	The ROC curves of the three classifiers showed that the two SVM algorithms outperformed the logistics regression. The temporally correlated SVM demonstrated the best performance.
3. O'Halloran, et al. [46]	Linear Regression	Tenfold cross-validation	R-squared value	Intoxication frequency was significantly predicted by the impulsivity variables. Consumption frequency was not significantly predicted by impulsivity variables.
4. Kinreich, et al. [53]	Linear-kernel Support Vector Machine	Tenfold cross-validation	Area under the curve and F-scores were used to evaluate the classification models	Combined model of EEG and SNP feature for female groups achieved the best accuracy.
5. Ngo, et al. [54]	Logistic Regression	NA	The sensitivity, specificity, positive predictive value, and negative predictive value were used to evaluate the classification models	The model with Top 50% performed with best.
6. Sisodia, et al. [55]	Naïve Bayes, Decision Tree, Simple Logistic, Random Forest, Bagging and Adaboost	Tenfold cross-validation	Precision Recall and F-Measure were used to evaluate the classification models	Simple Logistic and Random Forest performed better than the other classifiers.
7. Gowin, et al. [47]	Radial Support Vector Machine, Random Forest, and Elastic Net	75% of dataset reserved for training and 25% reserved for test sets	The model performance was calculated by the area under the curve, considering both sensitivity and specificity as well as the accuracy of models	Psychological models: Neuropsychosocial model had the best performance, AUC= 0.86, 95% and accuracy of 0.80. The psychosocial model achieved an AUC= 0.84, 95% and accuracy of 0.79. fMRI models: The motor task model achieved AUC= 0.58, 95% and accuracy of 0.63, but the language model achieved AUC= 0.70, 95% and accuracy of 0.66, while social models achieved AUC= 0.62, 95%.
8. Silveira, et al. [48]	Support Vector Machine, Logistic Regression, Adaboost, Gradient Boost, and Random Forest	Fivefold cross-validation	Precision, Recall, F1-score, and overall model accuracy	With an overall accuracy of 0.80, the Random Forest algorithm based on demographic, behavioural, and functional connectivity data provided the best classification results.
9. Chen, et al. [51]	Support Vector Machine, Convolutional Neural Network, and Convolutional Neural Network Combined with Long and Short-Term Memory	70% of dataset reserved for training and 30% reserved for test sets	The area under curve was used to calculate the specificity and sensitivity	Convolutional Neural Network achieved the best predictive accuracy (92.05%) followed by Convolutional Neural Network Combined with Long and Short-Term Memory and SVM with an accuracy of 87.50%.
10. Ebrahimi, et al. [49]	Support Vector Machine	80% of dataset reserved for training and 20% reserved for test sets	Precision, Recall, F-measure, and Accuracy	Two models were developed based on different feature selection methods. The model based on the chi-square technique achieved the best predictive accuracy (92%), while the one based on frequency of features archived 80%.
11. Bonnell, et al. [52]	Logistic Regression, Support Vector Machines, Neural Networks, K-Nearest Neighbor, Decision Trees, and Random Forests	65% of dataset reserved for training, 15% reserved for evaluation of initial models and 20% reserved as test set	The receiver operating characteristic curve and the area under the curve, as well as predictive accuracy, sensitivity, specificity, positive predictive value and negative predictive value were calculated to evaluate the predictive accuracy	The initial model based on the random forest model achieved the best AUC in both the training set (0.85) and the validation set (0.80) and outperformed the other machine learning methods in sensitivity, specificity, positive predictive value, negative predictive, overall accuracy, and savings in the validation set. The second-best AUC achieved by SVM, followed by neural networks, and K-nearest neighbor.
12. Ebrahimi, et al. [50]	Deep Neural Networks	5-fold cross-validation	The area under the curve and predictive accuracy was used to calculate the performance of the classifiers	Overall predictive accuracy: 88% Area under the curve: 96%, 82%, and 89% for Normal, Hazardous, and Harmful drinkers' classes.

and focused on different aspects, including database collection sites and the characteristics and types of datasets. Moreover, pre-processing and sampling techniques, types of features and techniques of feature reduction and selection of the most relevant subset of features, utilization of ML methods, and evaluation metrics used in performance evaluation of an ML model were evaluated.

The findings showed that researchers used different sources to collect datasets. This review revealed that many studies developed datasets based on single collection sites,

of which two major weaknesses were detected. First, the predictive model was developed using a one-dimensional dataset, which was collected from a single source of information, considered as one subject. Thus, a classification model trained using this kind of dataset cannot be used on a wide scale because many sources and subjects used for creating a sample dataset and for predicting AUD are available. Each source may have its own quality of the population, with different features and parameters. Therefore, multidimensional datasets are recommended for the development of a predictive

model. Multidimensional datasets collected from different sources, and a variety of subjects related to the prediction of AUD should be considered. Moreover, multidimensional datasets can produce accurate predictive models. They can also be used on a wider scale.

Second, some studies used datasets that contained insufficient sample data [81]; hence, the reported predictive model may have suffered from overfitting or underfitting. Although the studies employed many techniques to overcome limitations, many of the datasets require additional records to achieve excellent predictive accuracy [82] for AUD. Some datasets that were used for predicting AUD in the selected studies suffered from class imbalance. This refers to a problem where the sample size across the classes has not been equally distributed. This issue usually results in a high rate of false negatives [83] in the final predictive model. For instance, one study [55] used a publicly available dataset to predict AUD. Class labels were used for the training set, which consisted of 595 samples of the majority of classes and only 67 samples of the minority class. In other words, approximately 89% of the sample data were class label 0 (non-event), and approximately 11% were class label 1 (event). Therefore, it is recommended that researchers who use imbalanced data perform an appropriate sampling method that is popular and convenient to apply for ML tasks. The basic idea of the sampling technique is to produce balanced classes by either adding or removing sample data from the primary datasets. Sampling methods may reduce learning time while accelerating execution time in supervised learning algorithms [84] and outperform bagging and boosting [85]. Sampling methods can be categorized into three groups: oversampling (to increase the size of the minority class to obtain balanced classes), undersampling (draw the random set of samples from the majority class to balance the classes), and hybrid sampling (a combination of undersampling and oversampling) [86].

In this paper, several features were identified from the studies on AUD prediction. The most commonly used features included demographic features, such as age and gender, drinking behaviour, school-related variables, psychological variables, and health-related information. Feature engineering techniques that were used in the process of predicting AUD were also reviewed. The filter methods were the most widely used feature selection approaches [47], [49], [51]–[53]. Feature selection for classification tasks can generally be categorized into filter methods, wrapper methods, and embedded methods [87]. Generally, in filter methods, the selected features are general characteristics of the dataset as well as measures (such as consistency, dependency, correlation, distance, and information) of the features in comparison to target values, and they are independent of the classifier [64]. Wrapper methods use a specific classifier to evaluate the quality of the selected features. They also report features with the highest quality. Previous studies [68], [88] reported that the wrapper model has a better predictive accuracy than the filter model, but it is computationally expensive.

Therefore, some studies used embedded models to extract the best features for classification tasks. Embedded models are hybrid techniques that have the advantage of both filter and wrapper models. Given that they combine feature selection and classifier construction and use filter models, they are less computationally intensive than wrapper methods [89]. In the identified studies, only one [54] used the embedded models for feature selection.

The ‘no free lunch’ theorem [90] indicates that having the most accurate ML-based application in a domain requires the testing of a variety of ML algorithms. However, the algorithm that outperforms others using a collected dataset should be determined. Each of the identified studies used its own customized datasets, with several experimental settings. Therefore, statistically evaluating the performance of such studies is not practical. Out of the twelve primary studies, four studies [10], [45], [49], [53] used the SVM for predicting AUD. One [46] used linear regression, one [54] used logistics regression, and one [50] employed DNN application for the prediction of patients with AUD. Among those studies that compared several algorithms, two, [55] and [52], compared six different classifiers, one [47] compared four classifiers, and one [51] compared three different classifiers for predicting AUD. The results indicate that traditional machine learning algorithms were more dominant than NN algorithms in AUD-related studies. The primary benefit of the NN, in comparison to traditional machine learning algorithms, is that feature engineering steps by human experts are not needed. In NN, through the learning process, features are automatically learned from the training dataset. However, a predictive model based on NNs is a black box [91], which in health care-related studies creates adoption difficulties for clinicians since they prefer to fully understand and justify the actions for which they are essentially responsible [92].

Following a review of the primary studies’ findings, it was determined that scientists prefer to use standard evaluation metrics such as accuracy, sensitivity, and specificity for determining the success of developed predictive models. Moreover, almost all studies preferred to use a part of the selected dataset as a training set and the other part as the test set. Furthermore, none of the primary studies conducted external validation using unseen datasets.

## V. FUTURE RESEARCH DIRECTIONS

This SLR reviewed twelve studies that used ML techniques for predicting AUD. The reviewed aspects were inspired by the basic ML development method, which is shown in Figure 1. In this section, we present various research directions and problems that the researchers of the selected studies were unable to address. These research directions will be described according to the four steps of building a predictive model using ML techniques, including dataset collation; pre-processing and sampling techniques; types of features, dimensionality and feature selection techniques for feature engineering; and ML technique utilization process.

These important research problems and challenges must be improved by applying different techniques to increase the feasibility and performance of AUD prediction. The research challenges that need to be addressed are discussed below.

#### **A. DATASET COLLECTION, PRE-PROCESSING AND SAMPLING FOR THE PREDICTION OF AUD**

A dataset's quality is one of the primary research problems that must be considered in future research. As shown in Table 3, only three primary articles used a dataset collected from multiple sites and a combination of multiple data sources. One of the main problems that may arise due to the lack of multiple-multiple (multidimensional) datasets is losing the chance of having different documentation patterns or styles in the development of the predictive model. These differences could lead to generalizations in the final constructed classifiers. Data collected from multiple sources or the use of multidimensional datasets may reduce the chance of generalizations in the final classification models. Moreover, multidimensional datasets can help to build a predictive model on one dimension and validate it on the other dimension. For example, a predictive model developed based on a dataset containing a combination of patients' EHRs and interviews from a Western country could be validated using a dataset from an Eastern country.

The lack of multidimensional datasets may increase the risk of bias in the developed predictive models. Although the proposed methods in the primary studies offered a reasonably accurate result, there might be a risk of bias in their results since most of them collected data from a single site. Nevertheless, since multidimensional datasets are usually in the form of big data, big data tools and techniques must also be considered. By considering multiple sources for data collection, researchers would be able to handle security and privacy issues, especially when dealing with data collected from hospitals.

Publicly available datasets are another issue that needs to be considered. Although a few studies did not provide details for the dataset used in their study, all reported a reasonable prediction accuracy on AUD. This means that their results may lead this domain to publication bias since low result experiments can be disclosed. To overcome this issue, some public standard datasets must be available for benchmarking. Therefore, future works should focus on publishing their datasets in the form of public corpora for the perfection of AUD prediction.

Furthermore, class distribution in many of the datasets was imbalanced. Numerous studies tried to overcome class bias in the majority class using sampling techniques. However, there are still major challenges regarding imbalanced data in ML studies [93]. Future studies in the prediction of AUD may employ various sampling techniques, including resampling and reweighting techniques, to overcome the challenges of imbalanced data [86].

#### **B. FEATURE ENGINEERING IN THE PREDICTION OF AUD**

Several studies reported that demographic features, drinking behaviour, and educational backgrounds represented the most useful features for AUD prediction. Reviews of selected studies have demonstrated that features extracted primarily from survey questionnaires tend to consider the filter model and the wrapper model as techniques. This means that future studies should consider using new datasets that include different types of features and try to apply a variety of feature extraction and feature selection techniques. This will help them present the potential features for the prediction of AUD to clinical scientists and improve the performance of ML applications. Future researchers may design feature selector frameworks that can select the best subset of features and present clinical biomarkers and risk factors for the development of AUD. This would help the research discipline keep up with new trends in the prediction of AUD.

The lack of using EHR as datasets for the prediction of AUD is another issue that must be considered. As shown in Table 4, only three primary articles considered EHRs for the development of predictive models for AUD. EHR datasets are usually the medical records of patients that are stored in hospital databases. Such datasets consist of information regarding patient admissions and visits to hospitals, along with diagnoses and treatments. All of these can be stored in the form of ICD codes. Moreover, these types of datasets consist of clinical records, including laboratory results in the form of numeric values, magnetic resonance imaging of the brain and the description of the result in the form of text, clinical reports by general practitioners in the form of text or audio, etc. Extracting features from such datasets can assist clinicians in having a better understanding of the most important features, which could have a huge impact on the accuracy of the predictive ML-based models for AUD.

#### **C. MACHINE LEARNING TECHNIQUES IN THE PREDICTION OF AUD**

As shown in Table 5, only three studies used NN algorithms for the prediction of AUD. Thus, future research may consider investigating NNs and DNNs for the prediction of AUD. One of the primary advantages of NN algorithms is that the feature engineering step by a human expert can be skipped [94]. In NN, during the general learning process, features are automatically learned from the training data. Therefore, NN approaches might be the most suitable approaches for use in ML-based predictive models in clinical problems with high-dimensional data when a human expert is unable to reduce high dimensionality among the datasets.

The primary health care facility is normally the first place where most patients with AUD have their initial communication with health care systems [95]. Nevertheless, the majority of the world's primary health care system lacks systematic screening for AUD [96]. One study [95] stated that interventions for the treatment of AUD could begin with brief advice

or guidance to discourage hazardous drinking behaviour. However, at harmful levels of drinking, more serious treatments, such as lifestyle changes and additional psychological and pharmacological treatments, are required [95], [97]. Integrating these into primary health care remains a challenge in many countries. This is due to the lack of qualified staff and the lack of familiarity with formal psychological therapies among general practitioners [97]. One of the factors leading to the early detection or prediction of harmful drinking is the detection of early initiation of hazardous drinking [96], [98].

Our findings indicate that this issue has not been considered in primary studies. Therefore, the early detection of AUD in patients based on the progression of their alcohol misuse would be a necessity since this has not been thoroughly analysed, nor has it been reported. For this to happen, we need to consider feeding the temporal sequence of the AUD data into the ML algorithms. Assuming that a national dataset was collected and labelled with sequences that would indicate the progression of AUD levels in patients (based on blood tests, clinical courses or questionnaires), early detection of patients with AUD could be achieved by designing and developing a sequential ML application.

Alcohol consumption has traditionally been a male-dominated activity, with men consuming more alcohol and causing more alcohol-related damage to themselves and others than women [99], [100]. This is because female drinkers drink only approximately one-third of the overall amount of alcohol consumed by male drinkers each year [100]. However, the gaps in AUD between men and women are shrinking [99]. The rising rate of AUD in women has become a concern, since women encounter the adverse health and behavioural effects of alcohol use earlier and at lower levels than men [101]. There are several factors distinguishing AUD between sexes. Women are typically smaller than men, with a lower overall body water level and higher total body fat. As a consequence, alcohol is absorbed more in a woman's body. Women's blood alcohol intake also increases more rapidly and remains higher and longer than for men [102]. Additionally, one study [102] mentioned that gender variations exist in brain structure, neurochemistry, and function. Therefore, we highly suggest that researchers consider gender disparities when building predictive models for the prediction of AUD.

EHRs usually contain patient admission and discharge data of each admission episode, which are stored in the form of ICD codes. Such data may often share information about risk factors for the development of a disease. Recently, due to the growing number of EHRs, social network analysis (SNA) has gained attention among scientists for predicting disease risk [103] as well as for identifying patterns and the nature of disease comorbidities [104]. SNA can help understand AUD progression by considering the comorbidities that occur over a period prior to the development of AUDs. This will also lead to an AUD trajectory, which can be used for the prediction of AUD with ML techniques.

In the primary studies, several evaluation metrics, such as precision, recall, F1-Score, predictive accuracy, ROC, and

AUC, were used by researchers to evaluate the performance of the developed predictive models. However, almost all models were trained and tested using datasets derived from simple split (Holdout) or k-fold cross validation. If the models' performances were solely used as an evaluation metric, the developed model would not be justified as a real-world predictive model. The performance of a predictive model is dependent on a variety of considerations, including the specific application, the collected data, the sample size, and the quality of the data. We suggest that in our future works, researchers establish standard evaluation protocols, encourage comparative models, and use external datasets (national and internal, or multidimensional) to validate their predictive models' performance.

## VI. THREATS TO VALIDITY

This section discusses potential threats to our SRL's validity.

### A. SEARCH STRATEGY AND SEARCH RESULT

The search strategy included selecting digital libraries and searching for predefined keywords. This step may be jeopardized by factors such as missing or excluding relevant articles. To mitigate this risk, we used three strategies. First, the first author and a research librarian identified the primary search keywords and search queries. Second, we created search queries using different strings by creating various combinations of the selected keywords and synonyms based on MeSH and Boolean logic techniques. Third, to increase the possibility of identifying relevant articles, we ran search queries on the six digital libraries most relevant to our scope. We did not apply the snowballing process because our search yielded 3,355 papers, which we believe covered most of the papers relevant to our scope. Although we considered six digital libraries, we did not consider the Scoops digital library, which may have increased the chance of missing relevant study due to human error. Therefore, we cannot state that the results of this SLR cover all published studies on the prediction of AUD.

### B. DATA SCREENING AND SELECTION CRITERIA

Article screening and selection may potentially pose a threat to validity, as it may arise from the authors' objective judgement in selecting research, resulting in the exclusion of important articles or the inclusion of irrelevant publications. To mitigate this risk, we established the inclusion and exclusion criteria in advance, with all authors participating in the validation. Following that, we adhered strictly to these criteria throughout the article selection process. As stated in section II, this SLR utilized twelve main inclusion and exclusion criteria for data screening, which may violate the exclusion procedure and lead to an article being erroneously excluded.

### C. QUALITY ASSESSMENT AND DATA EXTRACTION

In the process of data extraction and quality assessment, threats may arise due to incomplete information extraction

TABLE 6. Quality assessment criteria of 6 studies.

Quality criteria	Authors											
	[Reference number]											
	Zuba, et al. [45]	Bi, et al. [10]	Kinreich, et al. [53]	O'Halloran, et al. [46]	Ngo, et al. [54]	Sisodia, et al. [55]	Gowin, et al. [47]	Silveira, et al. [48]	Chen, et al. [51]	Ebrahimi, et al. [49]	Bonnell, et al. [52]	Ebrahimi, et al. [50]
1. Are research objectives are clearly stated?	√	√	√	√	√	√	√	√	√	√	√	√
2. Are the ML methods defined to such degree that the study can be reproduced by someone else?	√	√	√	√	√	√	√	√	√	√	√	√
3. Is the origin of the data stated?	√	√	√	√	√	√	√	√	√	√	√	√
4. Is the minimum duration of data collection stated?	√	√	√	√	√	x	√	√	√	√	√	√
5. Is the number of training and testing data identified?	√	√	√	√	x	x	√	√	√	√	√	√
6. Are the pre-processing techniques used in the study clearly described and their selection justified?	√	√	√	x	x	x	√	√	√	√	√	√
7. Are the features used to build the ML model described clearly?	√	√	√	√	√	x	√	√	√	√	√	√
8. Is the process of feature engineering clearly stated?	√	√	√	√	√	x	√	x	√	√	√	x
9. Are the classifiers used in study clearly described?	√	√	√	√	√	√	√	√	√	√	√	√
10. Is there any comparison between the proposed approach with the existing approaches?	√	√	√	√	√	√	√	√	√	√	√	√
11. Were the performance measurements fully defined?	√	√	√	√	√	√	√	√	√	√	√	√
12. Does the conclusion reflect the research findings?	√	√	√	√	√	√	√	√	√	√	√	√
Overall Score	12	12	12	12	11	7	12	11	12	12	12	11

from the primary studies and their poor quality. To avoid threats to the validity of the quality assessment, a procedure was created and voted on by three authors. However, the existence of bias in the primary studies was not discussed in this SLR. To avoid threats to the validity of data extraction, we strictly followed the five aspects of building predictive models, which are presented in Figure 1, and then the extracted information was discussed by all authors.

VII. CONCLUSION

This SLR presented a comprehensive review of twelve studies that used ML techniques to predict AUD. The articles were published from 2010 to 2021 and were systematically extracted from six different academic databases. These studies were comprehensively reviewed from five different aspects, including collection sites, types and characteristics of datasets, data pre-processing and data sampling techniques, feature types, feature selection and feature extraction techniques, ML algorithm utilization and performance evaluation metrics. Several datasets with unique characteristics were identified and reported. In most of the selected studies, the investigators’ own collected datasets were used. Different sampling and pre-processing techniques were used to overcome the imbalanced class distribution in the datasets as well as removing noise and irrelevant information from the datasets. In terms of the features, demographic features together with FH, study-related features, and clinical features in the form of ICD codes were identified. To overcome high dimensionality and feature redundancy, several feature selection methods, including the filter method, wrapper method, and embedded method, were used. With respect to ML algorithms, most studies used SVM as the main ML algorithm for predicting AUD. However, the lack of deep learning

techniques for predicting AUD was evident and is suggested as one of the future research challenges for the prediction of AUD. Moreover, the lack of predictive models for the early detection of AUD, considering gender disparities, and the lack of a trajectory network or path towards AUD based on EHRs are other important research directions were suggested in this study. Focusing on performance metrics, it appears that the overall accuracy and ROC curve were the most popular evaluation methods among the studies. However, the lack of external validation in the primary studies is an important issue that must be considered in future works. The significance of these review findings was discussed in the separate discussion section. This comprehensive literature review provides unique insights into AUD prediction studies using ML techniques published during the past decade and outlines challenges and open issues that require additional attention in the future.

APPENDIX

Table 6: Quality assessment criteria of the twelve included studies.

ACKNOWLEDGMENT

The authors would like to thank Dr. Anne Faber Hansen, a Research Librarian at the University of Southern Denmark, for helping in the preparation of search keywords and queries.

REFERENCES

- [1] *Global Status Report on Alcohol and Health*, World Health Organization, Geneva, Switzerland, 2014.
- [2] S. S. Lim, T. Vos, A. D. Flaxman, G. Danaei, and K. Shibuya, “A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of disease study 2010,” *Lancet*, vol. 380, no. 9859, pp. 2224–2260, Dec. 2012.

- [3] B. F. Grant, R. B. Goldstein, T. D. Saha, and S. P. Chou, "Epidemiology of DSM-5 alcohol use disorder: Results from the national epidemiologic survey on alcohol and related conditions III," *JAMA Psychiatry*, vol. 72, no. 8, pp. 757–766, 2015.
- [4] J. Rehm, P. Anderson, J. Barry, P. Dimitrov, Z. Elekes, F. Feijão, U. Frick, A. Gual, G. Gmel, Jr., L. Kraus, S. Marmet, J. Ranninen, M. X. Rehm, E. Scafato, K. D. Shield, M. Trapencieris, and G. Gmel, "Prevalence of and potential influencing factors for alcohol dependence in Europe," *Eur. Addiction Res.*, vol. 21, no. 1, pp. 6–18, 2015.
- [5] J. Rehm, J. Manthey, P. Struzzo, A. Gual, and M. Wojnar, "Who receives treatment for alcohol use disorders in the European Union? A cross-sectional representative study in primary and specialized health care," *Eur. Psychiatry*, vol. 30, no. 8, pp. 885–893, Nov. 2015.
- [6] R. Kohn, S. Saxena, I. Levav, and B. Saraceno, "The treatment gap in mental health care," *Bull. World Health Org.*, vol. 82, pp. 858–866, Feb. 2004.
- [7] T. B. Üstün and N. Sartorius, *Mental Illness in General Health Care: An International Study*. Hoboken, NJ, USA: Wiley, 1995.
- [8] H.-J. Rumpf, "Screening for alcohol use disorders and at-risk drinking in the general population: Psychometric performance of three questionnaires," *Alcohol Alcoholism*, vol. 37, no. 3, pp. 261–268, May 2002.
- [9] K. G. Chartier, N. S. Thomas, and K. S. Kendler, "Interrelationship between family history of alcoholism and generational status in the prediction of alcohol dependence in US Hispanics," *Psychol. Med.*, vol. 47, no. 1, pp. 137–147, Jan. 2017.
- [10] J. Bi, J. Sun, Y. Wu, H. Tennen, and S. Armeli, "A machine learning approach to college drinking prediction and risk factor identification," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 1–24, Sep. 2013.
- [11] T. Rosenström, F. A. Torvik, E. Ystrom, N. O. Czajkowski, N. A. Gillespie, S. H. Aggen, R. F. Krueger, K. S. Kendler, and T. Reichborn-Kjennerud, "Prediction of alcohol use disorder using personality disorder traits: A twin study," *Addiction*, vol. 113, no. 1, pp. 15–24, Jan. 2018.
- [12] A. M. Harvanko, L. R. N. Schreiber, and J. E. Grant, "Prediction of alcohol and gambling problems in young adults by using a measure of decision making," *J. Addiction Med.*, vol. 7, no. 5, pp. 314–319, 2013.
- [13] A. Gabbiadini, F. Cristini, L. Scacchi, and M. G. Monaci, "Testing the model of goal-directed behavior for predicting binge drinking among young people," *Substance Misuse*, vol. 52, no. 4, pp. 493–506, Mar. 2017.
- [14] F. Ferreri, A. Bourla, S. Mouchabac, and L. Karila, "E-addictology: An overview of new technologies for assessing and intervening in addictive behaviors," *Frontiers Psychiatry*, vol. 9, p. 51, Mar. 2018.
- [15] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, no. 1, pp. 3–24, 2007.
- [16] H.-J. Yoon, L. Roberts, and G. Tourassi, "Automated histologic grading from free-text pathology reports using graph-of-words features and machine learning," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Feb. 2017, pp. 369–372.
- [17] B. Shin, F. H. Chokshi, T. Lee, and J. D. Choi, "Classification of radiology reports using neural attention models," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4363–4370.
- [18] X. Zhu, X. Du, M. Kerich, F. W. Lohoff, and R. Momenan, "Random forest based classification of alcohol dependence patients and healthy controls using resting state MRI," *Neurosci. Lett.*, vol. 676, pp. 27–33, May 2018.
- [19] T. Hastie, R. Tibshirani, and J. Friedman, "Overview of supervised learning," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009, pp. 9–41.
- [20] X. Zhu and A. B. Goldberg, *Introduction to Semi-Supervised Learning* (Synthesis Lectures on Artificial Intelligence and Machine Learning), vol. 14. San Rafael, CA, USA: Morgan & Claypool, 2009.
- [21] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, Montreal, QC, Canada, 1995, vol. 14, no. 2, pp. 1137–1145.
- [22] E. S. Giesen, H. Deimel, and W. Bloch, "Clinical exercise interventions in alcohol use disorders: A systematic review," *J. Substance Abuse Treatment*, vol. 52, pp. 1–9, May 2015.
- [23] A. K. Singal, S. Kodali, L. A. Vucovich, V. Darley-Usmar, and T. D. Schiano, "Diagnosis and treatment of alcoholic hepatitis: A systematic review," *Alcoholism, Clin. Exp. Res.*, vol. 40, no. 7, pp. 1390–1402, Jul. 2016.
- [24] L. K. Schmidt, A. B. Bojesen, A. S. Nielsen, and K. Andersen, "Duration of therapy—Does it matter: A systematic review and meta-regression of the duration of psychosocial treatments for alcohol use disorder," *J. Substance Abuse Treatment*, vol. 84, pp. 57–67, Jan. 2018.
- [25] S. J. Adamson, J. D. Sellman, and C. M. A. Frampton, "Patient predictors of alcohol treatment outcome: A systematic review," *J. Substance Abuse Treatment*, vol. 36, no. 1, pp. 75–86, Jan. 2009.
- [26] K. K. Mak, K. Lee, and C. Park, "Applications of machine learning in addiction studies: A systematic review," *Psychiatry Res.*, vol. 275, pp. 53–60, May 2019.
- [27] A. Ebrahimi, A. S. Nielsen, U. K. Will, and M. Mansourvar, "The prediction of alcohol use disorder: A scoping review," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2019, pp. 1062–1067.
- [28] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Univ. Durham, U.K., EBSE Tech. Rep. EBSE-2007-01, 2007.
- [29] *Endnote X8*, C. Analytics, Philadelphia, PA, USA, 2019.
- [30] M. Han and X. Liu, "Feature selection techniques with class separability for multivariate time series," *Neurocomputing*, vol. 110, pp. 29–34, Jun. 2013.
- [31] S. Shah, M. Sharma, D. Deb, and R. B. Pachori, "An automated alcoholism detection using orthogonal wavelet filter bank," in *Proc. Int. Conf. Mach. Intell. Signal Process. (MISP)*, vol. 748. Indore, India: Springer-Verlag, 2019, pp. 473–483.
- [32] M. A. Schuckit, "Alcohol-use disorders," *Lancet*, vol. 373, no. 9662, pp. 492–501, 2009.
- [33] R. Whelan, "Decisions, decisions: Machine learning as a tool to identify alcohol-use disorder treatment seekers," *EClin. Med.*, vol. 12, pp. 4–5, Jul. 2019.
- [34] J. D. C. Rodrigues, P. P. R. Filho, E. Peixoto, and V. H. C. de Albuquerque, "Classification of EEG signals to detect alcoholism using machine learning techniques," *Pattern Recognit. Lett.*, vol. 125, pp. 140–149, Jul. 2019.
- [35] A. Anuragi and D. S. Sisodia, "Alcoholism detection using support vector machines and centered correlogram features of brain EEG signals," in *Proc. Int. Conf. Inventive Comput. Informat. (ICICI)*, Nov. 2017, pp. 1021–1026.
- [36] N. Ahmadi, Y. Pei, and M. Pechenizkiy, "Detection of alcoholism based on EEG signals and functional brain network features extraction," in *Proc. IEEE 30th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2017, pp. 179–184.
- [37] M. Mete, M. Devous, J. Spence, and B. Adinoff, "A support vector machines model to classify cocaine patients," *Alcoholism, Clin. Exp. Res.*, vol. 1, p. 396A, Jun. 2012.
- [38] E. Echeburúa, M. Gómez, and M. Freixa, "Prediction of relapse after cognitive-behavioral treatment of gambling disorder in individuals with chronic schizophrenia: A survival analysis," *Behav. Therapy*, vol. 48, no. 1, pp. 69–75, Jan. 2017.
- [39] M. Symons, G. F. X. Feeney, M. R. Gallagher, R. M. Young, and J. P. Connor, "Machine learning vs addiction therapists: A pilot study predicting alcohol dependence treatment outcome from patient data in behavior therapy with adjunctive medication," *J. Substance Abuse Treatment*, vol. 99, pp. 156–162, Apr. 2019.
- [40] T. B. Wray, X. Luo, and P. M. Monti, "Using smartphone data to identify targets of just-in-time interventions for heavy drinking among gay and bisexual men at high risk for HIV," *Alcoholism-Clin. Exp. Res.*, vol. 43, p. 333A, Jun. 2019.
- [41] C. Walss-Bass, R. Suchting, R. L. Olvera, and D. E. Williamson, "Inflammatory markers as predictors of depression and anxiety in adolescents: Statistical model building with component-wise gradient boosting," *J. Affect. Disorders*, vol. 234, pp. 276–281, Jul. 2018.
- [42] A. C. de Sousa Silva, A. I. C. Arce, A. R. B. Tech, and E. J. X. Costa, "Quantifying electrode position effects in EEG data with Lempel-Ziv complexity," in *Proc. Annu. Int. Conf. Eng. Med. Biol.*, Aug. 2010, pp. 4002–4005.
- [43] A. Liberati, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *Ann. Internal Med.*, vol. 151, no. 4, pp. 65–94, Aug. 2009.
- [44] N. Dalkey, O. Helmer, and O. Helmer, "An experimental application of the Delphi method to the use of experts," *Manage. Sci.*, vol. 9, no. 3, pp. 458–467, Apr. 1963.

- [45] M. Zuba, J. Gilbert, Y. Wu, J. Bi, H. Tennen, and S. Armeli, "1-norm support vector machine for college drinking risk factor identification," in *Proc. 2nd ACM SIGHIT Symp. Int. Health Informat. (IHI)*, Miami, FL, USA, 2012, pp. 651–660.
- [46] L. O'Halloran, B. Pennie, L. Jollans, H. Kiiski, N. Vahey, L. Rai, L. Bradley, R. Lalor, and R. Whelan, "A combination of impulsivity subdomains predict alcohol intoxication frequency," *Alcoholism, Clin. Exp. Res.*, vol. 42, no. 8, pp. 1530–1540, Aug. 2018.
- [47] J. L. Gowin, P. Manza, V. A. Ramchandani, and N. D. Volkow, "Neuropsychosocial markers of binge drinking in young adults," *Mol. Psychiatry*, vol. 2020, pp. 1–3, Apr. 2020.
- [48] S. Silveira, R. Shah, K. B. Nooner, B. J. Nagel, and S. F. Tapert, "Impact of childhood trauma on executive function in adolescence—Mediating functional brain networks and prediction of high-risk drinking," *Biol. Psychiatry, Cogn. Neurosci. Neuroimag.*, vol. 5, no. 5, pp. 499–509, May 2020.
- [49] A. Ebrahimi, U. K. Wiil, K. Andersen, M. Mansourvar, and A. S. Nielsen, "A predictive machine learning model to determine alcohol use disorder," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2020, pp. 1–7.
- [50] A. Ebrahimi, U. K. Wiil, M. Mansourvar, A. Naemi, K. Andersen, and A. S. Nielsen, "Deep neural network to identify patients with alcohol use disorder," *Stud. Health Technol. Informat.*, vol. 281, pp. 238–242, May 2021.
- [51] F. Chen, M. Xiao, C. Chen, C. Chen, Z. Yan, H. Han, S. Zhang, F. Yue, R. Gao, and X. Lv, "Discrimination of alcohol dependence based on the convolutional neural network," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0241268.
- [52] L. N. Bonnell, B. Littenberg, S. R. Wshah, and G. L. Rose, "A machine learning approach to identification of unhealthy drinking," *J. Amer. Board Family Med.*, vol. 33, no. 3, pp. 397–406, May 2020.
- [53] S. Kinreich, J. L. Meyers, A. Maron-Katz, C. Kamarajan, and A. K. Pandey, "Predicting risk for alcohol use disorder using longitudinal data with multimodal biomarkers and family history: A machine learning study," *Mol. Psychiatry*, vol. 26, pp. 1133–1141, Oct. 2019.
- [54] D. A. Ngo, S. V. Rege, N. Ait-Daoud, and C. P. Holstege, "Development and validation of a risk predictive model for student harmful drinking—A longitudinal data linkage study," *Drug Alcohol Dependence*, vol. 197, pp. 102–107, Apr. 2019.
- [55] D. S. Sisodia, R. Agrawal, and D. Sisodia, "A comparative performance of classification algorithms in predicting alcohol consumption among secondary school students," in *Proc. Int. conference Mach. Intell. Signal Process. (MISP)*, vol. 748. Indore, India: Springer-Verlag, 2019, pp. 523–532.
- [56] A. P. Reynolds, G. Richards, and V. J. Rayward-Smith, "The application of K-medoids and PAM to the clustering of rules," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Berlin, Germany: Springer, 2004, pp. 173–178.
- [57] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [58] L. Al Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *J. Comput. Sci.*, vol. 2, no. 9, pp. 735–739, 2006.
- [59] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
- [60] C. Shimasaki, "Biotechnology products and their customers: Developing a successful market strategy," in *Biotechnology Entrepreneurship*. Amsterdam, The Netherlands: Elsevier, 2014, pp. 229–241.
- [61] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.
- [62] J. Heer, J. M. Hellerstein, and S. Kandel, "Predictive interaction for data transformation," in *Proc. CIDR*, 2015, pp. 1–7.
- [63] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classification: Algorithms Appl.*, p. 37, Jan. 2014.
- [64] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Boca Raton, FL, USA: CRC Press, 2007.
- [65] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, and S. Fong, "Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy," *Pertanika J. Sci. Technol.*, vol. 26, no. 1, pp. 1–14, 2018.
- [66] M. Goodarzi, B. Dejaegher, and Y. V. Heyden, "Feature selection methods in QSAR studies," *J. AOAC Int.*, vol. 95, no. 3, pp. 636–651, May 2012.
- [67] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
- [68] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.
- [69] V. Kumar and S. Minz, "Feature selection: A literature review," *Smart Comput. Rev.*, vol. 4, no. 3, pp. 211–229, 2014.
- [70] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
- [71] J. Li, L. Liu, M. Zhou, J.-J. Yang, S. Chen, H. Liu, Q. Wang, H. Pan, Z. Sun, and F. Tan, "Feature selection and prediction of small-for-gestational-age infants," *J. Ambient Intell. Hum. Comput.*, vol. 2018, pp. 1–15, Jun. 2018.
- [72] Y.-H. Chou, C.-M. Tiu, G.-S. Hung, S.-C. Wu, T. Y. Chang, and H. K. Chiang, "Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis," *Ultrasound Med. Biol.*, vol. 27, no. 11, pp. 1493–1498, Nov. 2001.
- [73] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.
- [74] R. E. Wright, "Logistic regression," in *Reading and Understanding Multivariate Statistics*, L. G. Grimm and P. R. Yarnold, Eds. American Psychological Association, 1995, pp. 217–244.
- [75] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [76] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, "K-nearest neighbor classification," in *Data Mining in Agriculture*. New York, NY, USA: Springer, 2009, pp. 83–106.
- [77] Y.-Y. Song and L. Ying, "Decision tree methods: Applications for classification and prediction," *Shanghai Arch. Psychiatry*, vol. 27, no. 2, p. 130, 2015.
- [78] C. Smith, *Decision Trees and Random Forests: A Visual Introduction for Beginners*. Vancouver, BC, Canada: Blue Windmill Media, 2017.
- [79] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [80] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [81] R. D. Riley, K. I. E. Snell, J. Ensor, and D. L. Burke, "Minimum sample size for developing a multivariable prediction model: PART II—binary and time-to-event outcomes," *Statist. Med.*, vol. 38, no. 7, pp. 1276–1296, 2019.
- [82] S. Safari, A. Baratloo, M. Elfil, and A. Negida, "Evidence based emergency medicine: Part 5 receiver operating curve and area under the curve," *Emergency*, vol. 4, no. 2, p. 111, 2016.
- [83] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci.*, vol. 513, pp. 429–441, Mar. 2020.
- [84] L. Cao and Y. Zhai, "Imbalanced data classification based on a hybrid resampling SVM method," in *Proc. IEEE 12nd Int. Conf. Ubiquitous Intell. Comput.*, Aug. 2015, pp. 1533–1536.
- [85] B. W. Yap, K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin, and N. N. Abdullah, "An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets," in *Proc. 1st Int. Conf. Adv. Data Inf. Eng. (DaEng)*. Singapore: Springer, 2014, pp. 13–22.
- [86] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–36, 2019.
- [87] 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.
- [88] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. New York, NY, USA: Springer, 2012.
- [89] Y. Saeyns, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
- [90] D. H. Wolpert and W. G. Macready, "No free lunch theorems for search," Santa Fe Inst., Santa Fe, NM, USA, Tech. Rep. SFI-TR-95-02-010, 1995.
- [91] J. Wang, *Data Mining: Opportunities and Challenges*. Monticano, Italy: Idea Group, 2003.
- [92] M. Tayefi, *Challenges and Opportunities Beyond Structured Data in Analysis of Electronic Health Records* (Computational Statistics). Hoboken, NJ, USA: Wiley, 2021, p. e1549.
- [93] E. A. Felix and S. P. Lee, "Systematic literature review of preprocessing techniques for imbalanced data," *IET Softw.*, vol. 13, no. 6, pp. 479–496, Dec. 2019.
- [94] G. Zhong, X. Ling, and L. N. Wang, "From shallow feature learning to deep learning: Benefits from the width and depth of deep architectures," *Data Mining Knowl. Discovery*, vol. 9, no. 1, p. e1255, 2019.
- [95] A. F. Carvalho, M. Heilig, A. Perez, C. Probst, and J. Rehm, "Alcohol use disorders," *Lancet*, vol. 394, no. 10200, pp. 781–792, 2019.

- [96] J. P. Connor, P. S. Haber, and W. D. Hall, "Alcohol use disorders," *Lancet*, vol. 387, no. 10022, pp. 988–998, 2016.
- [97] J. Rehm, P. Anderson, J. Manthey, K. D. Shield, P. Struzzo, M. Wojnar, and A. Gual, "Alcohol use disorders in primary health care: What do we know and where do we go?" *Alcohol Alcoholism*, vol. 51, no. 4, pp. 422–427, Jul. 2016.
- [98] J. McCambridge, J. McAlaney, and R. Rowe, "Adult consequences of late adolescent alcohol consumption: A systematic review of cohort studies," *PLoS Med.*, vol. 8, no. 2, Feb. 2011, Art. no. e1000413.
- [99] A. White, "Gender differences in the epidemiology of alcohol use and related harms in the united states," *Alcohol Res., Current Rev.*, vol. 40, no. 2, p. 1, 2020.
- [100] *Global Status Report on Alcohol and Health*, World Health Organization, Geneva, Switzerland, 2019.
- [101] K. T. Foster, B. M. Hicks, W. G. Iacono, and M. McGue, "Alcohol use disorder in women: Risks and consequences of an adolescent onset and persistent course," *Psychol. Addictive Behaviors*, vol. 28, no. 2, p. 322, 2014.
- [102] M. E. McCaul, D. Roach, D. S. Hasin, C. Weisner, G. Chang, and R. Sinha, "Alcohol and women: A brief overview," *Alcoholism, Clin. Exp. Res.*, vol. 43, no. 5, p. 774, 2019.
- [103] F. Folino, C. Pizzuti, and M. Ventura, "A comorbidity network approach to predict disease risk," in *Information Technology in Bio and Medical Informatics* (Lecture Notes in Computer Science), vol. 6266, S. Khuri, L. Lhotska, and N. Pisanti, Eds. Berlin, Germany: Springer-Verlag, 2010, pp. 102–109.
- [104] M. Guo, Y. Yu, T. Wen, X. Zhang, B. Liu, J. Zhang, R. Zhang, Y. Zhang, and X. Zhou, "Analysis of disease comorbidity patterns in a large-scale China population," *BMC Med. Genomics*, vol. 12, no. S12, p. 177, Dec. 2019.



**ALI EBRAHIMI** received the M.Sc. degree in computer science from the University of Malaya (UM), Kuala Lumpur, Malaysia, and the Ph.D. degree in computer engineering from the University of Southern Denmark (SDU), Odense, Denmark. He worked as a Software Developer and a Research Assistant with Research and Development. He is currently a Postdoctoral Researcher with the SDU. He has authored or coauthored several published articles in peer-reviewed international conferences and journals. His research interests include artificial intelligence, machine learning, natural language processing, predictive models for early detection and prognosis based on electronic health records, and social network analysis.



**UFFE KOCK WIIL** received the M.Sc. and Ph.D. degrees from Aalborg University, Denmark, in 1990 and 1993, respectively. Since 2004, he has been a Full Professor with The Maersk Mc-Kinney Moller Institute, University of Southern Denmark. From 2012 to 2018, he was a Grant Holder and a Principal Investigator of Patient@home aimed at developing products and services that contributed actively to help resolve the great challenges facing the healthcare system and society (budget: >300m DKK). He has published over 200 research papers (over 150 of these in peer-reviewed international journals, conference proceedings, and books). His publications have been cited more than 2300 times (H-index = 24) according to Google Scholar. His research interests include data analytics and visualization, user involvement, software technologies, social network analysis and mining, and innovation management with a particular focus on application within data-driven health technology (intelligent monitoring and decision support; predictive models for prevention, early detection, diagnosis, treatment and prognosis; digital health systems offering novel personalized data collection, treatment plans and interventions; healthcare ecosystems). He is serving on the Editorial Boards for the *Journal of Network and Computer Applications* (Elsevier), the *Social Network Analysis and Mining* (Springer), the *Journal of Healthcare Informatics Research* (Springer), and the *BMC Emergency Medicine* journal. He is the Editor-in-Chief of the *Network Modeling Analysis in Health Informatics and Bioinformatics* (Springer).



**THOMAS SCHMIDT** received the Bachelor of Science (B.Sc.) and the Master of Science (M.Sc.) degrees in computer systems engineering from the University of Southern Denmark (SDU), in 2007, and the Doctor of Philosophy (Ph.D.) degree in software engineering from SDU in collaboration with the Emergency Department, Odense University Hospital, in 2016. Subsequently, he worked in industry as a Software Developer and a Product Manager, until late 2012, when he initiated his study to the Ph.D. degree. He is currently an Associate Professor with the Center for Health Informatics and Technology, SDU. His research interests include implementation and evaluation of clinical decision support systems and cyber security in healthcare. He serves as a member for the Nordic eHealth Research Network, and holds board positions in the Danish Biomedical Society, as well as the Danish Society for Digital Health.



**AMIN NAEMI** received the B.Sc. degree in computer hardware engineering from the K. N. Toosi University of Technology, Iran, and the M.Sc. degree in artificial intelligence from Amirkabir University (Tehran Polytechnic), Iran. He is currently pursuing the Ph.D. degree with the Health Informatics and Technology Unit, University of Southern Denmark. His research interests include machine learning, biomedical signal processing, and natural language processing.



**ANETTE SØGAARD NIELSEN** is currently a Professor WSR at the Unit of Clinical Alcohol Research, University of Southern Denmark, and the Region of Southern Denmark. Till 2013, she was in charge of the Alcohol Treatment Facility of the Municipality of Odense, and previously the Chief of the Alcohol Treatment Center of Fuenen County, while at the same time working in research regarding the treatment of alcohol dependence. She is particularly interested in factors of importance to the outcome of treatment, the implementation of research results and the patients' perception of their encounter with the treatment system.



**GHULAM MUJTABA SHAIKH** received the Ph.D. degree in computer science from the University of Malaya, Kuala Lumpur, Malaysia, in 2018. Since 2006, he has been associated with Sukkur IBA University. He is currently working as an Associate Professor with the Department of Computer Science, Sukkur IBA University. He is also working as the Director of the Center of Excellence for Robotics, Artificial Intelligence and Blockchain (CRAIB). His research interests include artificial intelligence, machine learning, online social networking, text mining, text classification, image classification, and deep learning. He has authored or coauthored several articles in academic journals indexed in well reputed databases, such as Web of Science, IEEE Xplore, ScienceDirect, Springer, and Scopus.



**MARJAN MANSOURVAR** received the M.Sc. and Ph.D. degrees in computer science from the University of Malaya (UM), Malaysia, the oldest and highest-ranking Malaysian institution of higher education, in 2011 and 2014, respectively. She is currently a Computational Scientist with the Department of Mathematics and Computer Science (IMADA), University of Southern Denmark. Her research interests include big data analysis, artificial intelligence, and machine learning in different applications to address real problems in society, health, and industry. Her research has resulted in the publication record in high impact, international journals, ranked in the top quartiles in the field, and international conferences.

...