Data-Driven Approaches to Smoking Cessation: Unraveling Predictors of Quitting Through Machine Learning

Srinath Reddy Ch\textsuperscript{1}, Kotthoju Nagendra Chary\textsuperscript{2}

\textsuperscript{1,2}Assistant professor, Department of Computer Science & Engineering, Sreenidhi Institute of Science and Technology Hyderabad

Abstract:
In order to define the usefulness of machine learning in this domain and to pinpoint the machine learning techniques that have been used, a comprehensive review of the literature has been conducted. Multiple searches in MEDLINE, Science Citation Index, Social Science Citation Index, EMBASE, CINAHL Plus, APA PsycINFO, PubMed, Cochrane Central Register of Controlled Trials, and IEEE Xplore were conducted for the current study through December 9, 2022. Studies reporting cigarette smoking cessation results (smoking status and cigarette consumption) as well as a variety of experimental designs (such as cross-sectional and longitudinal) were considered as inclusion criteria. The effectiveness of behavioral markers, biomarkers, and other predictors was evaluated as a predictor of smoking cessation outcomes. Twelve papers were found in our systematic review that met our inclusion criteria. This review includes.

Keywords: Machine learning; Systematic review; Smoking cessation;

1. INTRODUCTION
Smoking causes about half a million deaths yearly, including nearly 30\% of all cancer deaths in the US, making it the most preventable cause of death and disease [1–2]. These fatalities outnumber all deaths linked to alcohol, illegal drug use, violence, AIDS, and suicide put together [2]. Smoking cigarettes results in significant health and productivity costs that surpass $300 billion USD annually, which is not surprising. Reducing cigarette use can enhance public health by lowering mortality and morbidity. Fortunately, the majority of smokers want to stop, and there are effective smoking cessation treatments available. These medicines, such as bupropion, behavioral therapy, and nicotine replacement, lead to higher quitting rates than control or placebo treatments [3]. These therapies do, however, still have a great deal of space for development. For instance, smokers relapse at a rate of 70\% within 6 months of finishing the most intensive evidence-based treatments under tightly monitored conditions [4, 5]. Consequently, identifying cessation predictors may make it easier to match patients with treatments and enhance treatment results.

Machine learning is one method for finding predictors of smoking cessation. A branch of artificial intelligence called machine learning "gives computers the ability to learn without explicitly being programmed"[6]. Machine learning techniques can be divided into two broad categories: supervised learning, which involves fitting a model to data that has been labeled through experimental
measurements or assignments, and unsupervised learning, which involves spotting patterns in unlabeled data. However, hybrid methods like semi-supervised learning aim to combine the advantages of both supervised and unsupervised learning (see Review [8] for more information).

Typically, these techniques develop and train a model using sample data sets before testing that model on new samples. While there have been more machine learning applications to health behaviors, as shown by various studies (e.g., [10, 11]), there haven't been many reviews of machine learning methods for tackling substance and cigarette use.

For instance, a review of machine learning in addiction research found significant variation in predicting current substance use [12] and identified the lack of data or inadequate validation measures as contributing factors. Another example reported that most research had a moderate to high risk of bias due to missing data or a lack of external validation [13] and conducted a systematic evaluation of machine learning approaches to alcohol use disorder. The scoping review of tobacco research, which is most pertinent to the present review, demonstrated the variety of applications of machine learning (e.g., content analysis of tobacco on social media, classification of smoking status, and prediction of tobacco-related outcomes) and came to the conclusion that it was a potent tool that could advance research and policy addressing tobacco control [14].

![PRISMA diagram](image)

**Fig. 1.** PRISMA diagram.
The purpose of this systematic study is to describe how machine learning techniques are used and their capacity to find predictors of smoking cessation outcomes. We highlighted knowledge gaps and potential for innovation for machine learning research in the field of smoking cessation in this review.

2. Methods
Using Preferred Reporting Items for Systematic Reviews (PRISMA; Fig. 1), we report this systematic review. The systematic review protocol is listed in PROSPERO as (CRD42022323796).

2.1. Search strategies
PubMed MEDLINE, Science Citation Index, Social Science Citation Index, CINAHL Plus, APA PsycINFO, PubMed, Cochrane Central Register of Controlled Trials, and IEEE Xplore were among the databases that were searched for papers that were published before December 9, 2022. The search used full text, peer-reviewed, and English language filters to find publications on machine learning and smoking cessation. In particular, the following terms were looked up: ("machine learning "OR "reinforcement learning "OR "deep learning "OR "text mining ") AND (smoking OR tobacco OR nicotine OR cigarette OR e-cigarette OR "electronic nicotine delivery system "OR "smoking cessation "OR "smoking reduc "OR "tobacco reduc "OR cotinine OR "carbon monoxide " Studies have to be peer-reviewed, full-text, and available in English.

2.2. Criteria for inclusion and exclusion
The current review comprised observational studies (such as cross-sectional, longitudinal, and case series) and clinical trials (randomized or not) assessing smoking cessation results. Studies were also required to assess those results using machine learning. The participants in the study were smokers. Studies involving non-human subjects, studies on animals, and any gray literature were all disregarded.

2.3. Outcome measures
Machine learning meta-classes and methodologies that were employed to find predictors of smoking cessation results were our main metrics of interest.

2.4. Data collection and processing
L.N.A. and R.F.L. gathered the search results and put them into Covidence for further processing. The titles and abstracts were examined independently by two reviewers (C.L.D. and W.H.C.). Conflicts were settled by a third viewer (D.C.T.). Full-text screening was conducted independently by M.M. and Y.-H.Y., two reviewers. Disagreements were resolved in the full-text review by a third reviewer (D.C.T.). Full texts were excluded for the reasons that were noted and documented (Fig. 1). The data extraction and quality assessment for the included studies (D.C.T., L.N.A., M.M., R.F.L., Y.-H.Y.) were handled by the investigation team. Extracted data includes the following: study identifier, publication year, continent, study sample, study kind, intervention, participant count, demographics, smoking cessation outcome measure, assessed smoking cessation predictors, and machine learning technique.

2.5. Methodological quality assessment
The Mixed Method Appraisal Tool (MMAT; [9]) was used to evaluate the caliber of the studies that were included. The MMAT has 19 methodological quality criteria that are used to evaluate the quality of
qualitative, quantitative, and mixed methods investigations. To determine an overall quality score for each included study, the 19 quality criterion domains are each given a score on a Yes/No/Can't Tell scale.

3. Results
3.1. Search results
4,306 citations were found in the initial database search. Through citation searching, two more studies were found and added. Duplicates (2,025) were excluded, leaving 2,283 titles and/or abstracts for screening. A total of 161 studies were sought for retrieval, and 136 full-text studies, including the two extra studies found through reference list searches, were evaluated (Fig. 1). Due to study design (n = 61), outcomes (n = 35), patient population (n = 3), and other (n = 25; for instance, text mining of tweets and examination of electronic medical records), we removed 124 studies. The review comprised 12 papers in total.

3.2. Study characteristics and quality assessment
Table 1 summarizes the general caliber of the studies that were included. Additionally, Supplementary Table 2 contains the results of the scoring of each MMAT item. A study with an MMAT score of 0% (no quality criteria met) received no points, a study with a score of 20% (one quality criteria met), a study with a score of 40% (two quality criteria met), a study with a score of 60% (three quality criteria met), a study with a score of 80% (four quality criteria met), and a study with a score of 100% (five quality criteria met).

3.3. Demographics
The characteristics of the included studies are shown in Table 1 and Supplementary Table 2. There were 40,208 individuals in all of the research, with the sizes of the studies varying from 39 [10] to 14,443 [11]. The studies that were incorporated were released between 2006 and 2022. The majority of the included studies—six out of twelve—were carried out in the United States, with additional research being carried out in the Netherlands, Canada, South Korea, China, and New Zealand. Smoking cessation in the included studies was assessed using self-report measures (such as point prevalence abstinence, reports of relapse, and daily cigarette consumption) and/or biochemical validation (such as carbon monoxide measurements). In Supplementary Table 2, demographic details given by each study are listed. A few research focused on more niche subgroups of smokers, including pregnant women and those from socioeconomically disadvantaged backgrounds [12, 11]. The majority of studies used non-specific samples. Participants in half of the trials (6/12) reported their baseline cigarette consumption, which ranged from at least 100 cigarettes in their lifetime to an average of 21 cigarettes per day.

3.4. Machine learning classes and techniques
Supervised machine learning techniques were used in each of the 12 included research. Table 1 lists certain machine learning techniques and their associated analysis of each study. The supervised machine learning techniques in particular comprise logistic regression, random forest, classification trees, and regression trees. Seven studies (n = 7) reported metrics for sensitivity and specificity, five studies utilized area under the curve (AUC), one study included both positive and negative predictive value, and three studies did not publish any pertinent machine learning metrics (Table 1).
3.5. Smoking cessation outcomes

Studies assessed the success of smoking cessation by self-report, biochemical evaluation, or both. Results of the biochemical validation included measurements of carbon monoxide using expired air (18) and cotinine using saliva (27). Point-prevalence abstinence [21, 22], the number of cigarettes smoked [expired air; 18], Timeline Followback for use of patches and lozenges [24], re-lapse/lapse [16, 20], abstinence [17, 23], whether they were a current or former smoker [19], and reaching their daily substance use goal for the last seven days before discontinuing the intervention [26] were among the outcomes that were reported by participants themselves. It should be noted that only one study [expired air; 18] combined self-reported (cigarette smoking) and biochemical validation (expired carbon monoxide).

3.6. Predictors of smoking cessation evaluated

The 12 studies that were used made use of several smoking cessation predictors from seven different categories, including biomarkers, economic, environmental, and sociodemographic factors, engagement, neurocognitive factors, physical health-related factors, psychological factors, smoking severity and history, and other factors (Table 2). Exhaled carbon monoxide or neuroimaging (i.e., anatomical and functional imaging) were the predictors for biomarkers (n = 2 studies; [10, 15]). Economic, environmental, and sociodemographic factors (n = 9 studies; [11–19]) included elements including gender, race, household income, the number of smokers in the home, and the availability of cigarettes. User response rate and attendance predictors were included in the evaluation (n = 3 studies; [11, 16, 20]). Executive functioning (for example, delay discounting), memory, and attention were measured as part of neurocognitive (n = 1 study; [21]), which included them as predictors. Predictors for physical health included body mass index, alcohol consumption, sleep quality, and hypertension (n = 7 studies; [12-18]). Among the psychological predictors were motivation, self-efficacy, affect, and perceived stress (n = 9 studies; [11-15, 17-19, 21]). Indicators of nicotine dependency, such as the Fagerstrom Test for Cigarette dependency (formerly known as the Fagerstrom Test for Nicotine Dependence; [22]), age of smoking initiation, and quantity of intake were included in smoking severity and history (n = 8 studies; [11, 13, 15-19, 21]).
Table 1 Summary of studies evaluating smoking cessation outcomes with machine learning techniques.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Country</th>
<th>Study Methodology</th>
<th>Number of Participants</th>
<th>Smoking Cessation Method</th>
<th>Machine Learning Method</th>
<th>Validation Method</th>
<th>Did they test the predictions on an independent dataset?</th>
<th>What metrics were included?</th>
<th>Description of the outcome variable measured?</th>
<th>What were the outcomes measured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coughlin 2020</td>
<td>United States</td>
<td>Observational Cohort study - Cohort secondary data analysis</td>
<td>190</td>
<td>Yes</td>
<td>Supervised Classification and regression trees</td>
<td>Yes</td>
<td>Yes</td>
<td>Sensitivity; Specificity</td>
<td>Retreatment and 6-month follow-up</td>
<td>(continued on next page)</td>
</tr>
<tr>
<td>Doong and others 2020</td>
<td>Korea</td>
<td>Observational Other Public study - Conjoint data analysis</td>
<td>391</td>
<td>No</td>
<td>Supervised Logistic regression</td>
<td>No</td>
<td>Sensitivity; Specificity</td>
<td>Self-report point: Data collected from prevalence of smoking cessation reported in Korea</td>
<td>(continued on next page)</td>
<td></td>
</tr>
<tr>
<td>Fua 2022</td>
<td>Ontario, Canada</td>
<td>Observational Online study - Conjoint data analysis</td>
<td>884</td>
<td>No</td>
<td>Supervised Gradient boosting machines (GBM) model</td>
<td>Yes</td>
<td>Sensitivity; Specificity</td>
<td>Area under the ROC curve</td>
<td>Other: 'Incubation period about how many cigarettes were smoked'</td>
<td>(continued on next page)</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Study ID</th>
<th>What country did study design?</th>
<th>Study take place? (i.e., setting)</th>
<th>Where did the study take place?</th>
<th>What is the sample size of the study?</th>
<th>Does the study have an intervention?</th>
<th>Describe the machine learning methods used</th>
<th>Was there training/testing and/or cross-validation performed?</th>
<th>Did they test the predictions on an independently collected dataset?</th>
<th>What metrics were included?</th>
<th>Describe the smoking cessation outcome measured</th>
<th>What were the quality outcomes measured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hibbert 2022</td>
<td>United States</td>
<td>Experimental, in-lab study</td>
<td>Randomized control trial</td>
<td>Analyses were restricted to the 74 participants with a directly identified lapse (n = 52, 70.3%) or were verified as non-lapsed (n = 22, 29.7%) at the 4-week post-quit visit.</td>
<td>Yes</td>
<td>Supervised Cox proportional hazards regression and Penalized-er proportional hazards regression</td>
<td>Yes</td>
<td>No</td>
<td>Other Smoking status</td>
<td>Post quit of 4 weeks ** was self-reported at each EMA. Participants were asked if they had smoked any cigarette (even a puff), and how long ago they last smoked a cigarette, with seven response options ranging from 10 min ago to more than 6h ago. Temporally, data for each participant included all EMA observations up to the final measurement before the first lapse occurred. If the participant did not lapse, all of their observations were included and tabulated as censored.</td>
<td>Other: Self-Reports - TLFH for patches and mini-lozenges</td>
</tr>
<tr>
<td>Kim 2019</td>
<td>United States</td>
<td>Experimental, other</td>
<td>In-person counseling sessions, phone counseling, automated calls</td>
<td>021 (315 for recommended usual care: BOC - and 308 for adherence-optimized treatment AOT)</td>
<td>Yes</td>
<td>Supervised The Generalized, Unbiased, Interaction Detection and Estimation classification tree modeling program was used to identify predictors of patch use (in BOC and AOT) and mini-lozenges use in AOT, and to distinguish subgroups of smokers based on medication use.</td>
<td>No</td>
<td>No</td>
<td>Sensitivity</td>
<td>Patch use was assessed 4 weeks post-TQD (for both BOC and AOT) and 36 weeks post-TQD (for AOT). Mini-lozenges *</td>
<td>TQD = Target quit date</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Year</th>
<th>Country</th>
<th>Design</th>
<th>Participants</th>
<th>Data Collection</th>
<th>Machine Learning Methods</th>
<th>Model Evaluation</th>
<th>Outcomes</th>
<th>Smoking Cessation Outcome</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li 2021</td>
<td>China</td>
<td>Observational</td>
<td>Chart review</td>
<td>4875</td>
<td>No</td>
<td>Supervised</td>
<td>ANN, SVM, random forest, logistic regression, and decision trees</td>
<td>Area under the ROC curve</td>
<td>Sensitivity</td>
<td>Self-reports, point prevalence, abstainers</td>
</tr>
<tr>
<td>Mckinnon 2022</td>
<td>New Zealand</td>
<td>Observational</td>
<td>Cohort</td>
<td>14,443</td>
<td>Yes</td>
<td>Supervised</td>
<td>Various machine learning models, including decision trees, random forests, and neural networks</td>
<td>Area under the ROC curve</td>
<td>Sensitivity</td>
<td>Self-report, smoking cessation outcome</td>
</tr>
<tr>
<td>Payman 2006</td>
<td>United States</td>
<td>Observational</td>
<td>Cross-sectional</td>
<td>Current smokers = 7,421; Former smokers = 6,916</td>
<td>No</td>
<td>Supervised</td>
<td>Decision tree algorithms, logistic regression, and neural network models</td>
<td>Area under the ROC curve</td>
<td>Sensitivity</td>
<td>Self-report, current or former smoker</td>
</tr>
<tr>
<td>Ramos 2022</td>
<td>Netherlands</td>
<td>Observational</td>
<td>Online study</td>
<td>Cohort</td>
<td>52,889 participants</td>
<td>Yes, enrolled from January 2016 to December 2020</td>
<td>Supervised</td>
<td>Logistic Regression and Random Forest</td>
<td>Area under the ROC curve</td>
<td>Sensitivity</td>
</tr>
</tbody>
</table>

Notes: Further details on the study design and methods are available in the full text of the article.
Table 2 Smoking cessation predictors.

<table>
<thead>
<tr>
<th>Study</th>
<th>Variables</th>
<th>Economic, Environmental, &amp; Sociodemographics</th>
<th>Engagement</th>
<th>Neurocognitive</th>
<th>Physical Health - Related</th>
<th>Smoking Severity &amp; History</th>
<th>Best Predictor Identified (by machine learning technique)</th>
<th>Evidence of Predictive Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cagilin 2020</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>- Delay discounting</td>
<td>- FTND</td>
<td>Delay discounting - The first spit of each trial, low discounting on average correctly classified 74.6% of participants</td>
</tr>
<tr>
<td>Dungare 2020</td>
<td>N/A</td>
<td>Age</td>
<td>Gender</td>
<td>Education</td>
<td>Household Income</td>
<td>Marital Status</td>
<td>Number of daily smokers at home</td>
<td>Occupation</td>
</tr>
<tr>
<td>Fu 2022</td>
<td>N/A</td>
<td>Age</td>
<td>Sex</td>
<td>Education</td>
<td>Employment status</td>
<td>Marital status</td>
<td>Chronic physical health status</td>
<td>History of or current diagnosis of asthma, chronic pain, other conditions</td>
</tr>
<tr>
<td>Robert 2021</td>
<td>N/A</td>
<td>Current physical location</td>
<td>Interacting with smokers</td>
<td>Smoking base</td>
<td>Cigarettes are available to use</td>
<td>N/A</td>
<td>Purchase alcohol consumption</td>
<td>Cigarette craving and urge</td>
</tr>
</tbody>
</table>

(continued on next page)
3.7. Machine learning identified predictors

Table 2 is a list of the factors that each included study found to have the best predictive usefulness. Overall, delay discounting, the reduction in reward value as a function of delay to receipt [21], having people who smoke every day at home [16], attending smoking cessation education [16], having more positive experiences (measured by the Vaping Experiences Score) [17], higher perceived odds of smoking today [14], and an increase in risky behavior were the predictors with the highest predictive utility found in the 12 studies. (i.e., cognitive improvement, primary dependence, and taste/sensory properties) [18], parent ethnicity [11], smoking intervals [11], boredom [11], male sex [11], consumption when the intervention first started [20], engagement [20], irritability [12], cigarette availability [12], exposure to smokers [12], smoking restrictions [12], recent alcohol consumption [12] All seven of the aforementioned categories were included in the predictors this review found. A priori decisions, model reduction, and variable selection were all methods used in the included studies to evaluate which predictors to include (Supplementary Table 1). The included studies calculated the importance of these measures, odds ratios (probability that an individual will observe an outcome given an exposure compared to its absence [24]), univariate cox proportional hazards regression models (probability that an individual will experience a given event centered around a defined point in time [25], and Shapley additive exPlanations (SHAP) importance plots (quantify the prediction of an ins) as well as other methods. We acknowledge that some of the used algorithms have more obvious meanings in the feature space while others take a more "black box" approach, and that determining the relevance of a given feature is not an easy task. See [27] for a summary of these techniques.

4. Discussion

This is the first systematic review that we are aware of that assesses the use of machine learning techniques to forecast the success of smoking cessation. We examined the many kinds and subcategories of machine learning techniques employed in smoking cessation prediction. Then, we summarized the seven categories of currently used smoking cessation predictors and emphasized those with the most predictive power. We concluded by summarizing the smoking cessation outcome measures. Approaches to machine learning offer strong tools that can provide answers from a new paradigm. Many see machine learning as a mystifying phenomenon that excels in prediction but presents difficulties in interpretation. Some methods, such Support Vector Machines and Artificial Neural Networks, might support this viewpoint. However, certain machine learning techniques (such as classification and regression trees and decision trees) are based on trees and offer a very interpretable model. Several psychological factors, such as mood and self-efficacy, were found to be significant predictors of cessation by Medina et al. [11]. Additionally, Couglin et al. [21] examined a variety of psychological and neurocognitive parameters as predictors of smoking cessation in cognitive behavioral therapy (CBT) treatment. They did this by using classification and regression trees. The strongest predictor, delay discounting, which is the decrease in reward value as a function of delay to receipt, correctly identified post-treatment smoking for 80% of subjects. These findings show how tree-based machine learning might support therapeutic smoking cessation programs. Many of the most powerful predictors found using machine learning techniques are in agreement with those found using non-machine learning techniques earlier in the literature. For instance, systematic reviews and meta-analyses that looked at factors that influence smoking cessation have identified factors related to economics, the environment, and sociodemography (such as age, ethnicity, socioeconomic status, and environmental cigarette
exposure); factors related to physical health (such as chronic conditions, alcohol consumption); and factors related to smoking severity and history (such as dependence measures, duration of smoking, and age of smoking initiation). Although predictors related to economic, environmental, and sociodemographic factors, physical health-related factors, smoking severity and history, and can help identify who may benefit from additional interventions, they are less informative on how to improve clinical interventions without understanding underlying mechanisms of behavior change. Neurocognitive [30] and psychological predictors of smoking cessation outcomes [24, 26, 29] have also been established in extant literature, despite the fact that several studies identified in this review focus on the predictive power of underlying psychological and neurocognitive mechanisms of behavior change. In the cognitive-behavioral model of relapse prevention, for instance, affective states and cognitive processes like self-efficacy and motivation have long been identified as drivers of relapse in substance use disorders [31–34]. Delay discounting is also strongly associated with smoking outcomes, differentiating smokers from controls [35-40], forecasting the intensity of use [36, 41–43], and predicting treatment outcomes [44, 45]. While several biomarkers for quitting smoking (such as cotinine and carbon monoxide) [22] have been discovered using non-machine learning techniques, machine learning has helped to uncover a brand-new biomarker in fMRI resting-state activity [10]. Two further studies that were thoroughly reviewed but ultimately disregarded were:...
4.1. Limitations and potential future directions

The sparse inclusion of the neurobehavioral mechanisms underlying smoking-related decision-making is a significant study area constraint. According to the conflicting neurobehavioral decision systems (CNDS) theory, an imbalance between the executive and impulsive decision-making processes contributes to addiction [53, 54]. Individuals with higher delay discounting rates show hyper-activation in impulsive brain systems (such limbic system parts) and hypoactivation of executive regions (like prefrontal cortex) [55]. Delay discounting has been established as an indication of CNDS balance. Delay discounting and possibly other neurocognitive decision-making variables could therefore be included in machine learning and smoking cessation research in the future, which may lead to the discovery of new treatment targets (c.f., [30]). The most effective smoking intervention treatment, according to research, changed delay discounting through rate-dependent effects, meaning that people with the highest baseline delay discounting rates experienced the biggest changes in discounting after treatment [56]. Interventions that lessen discounting simultaneously result in decreased cigarette demand [57, 58]. Delay discounting could therefore be usefully included into machine learning models for a more specialized treatment strategy. The apparent (1) reliance on self-reported measures, both as explanatory variables (e.g., FTND) and outcomes of smoking cessation (e.g., cigarettes per day), is another limitation in this field. The robustness of using machine learning would be improved by including clinical diagnoses and biochemical evidence of smoking cessation (e.g., CO levels). (2) Overfitting may occur when machine learning techniques are used to smaller, homogeneous data sets. By using big data sets (as done in [13]) and evaluating models using a different data set (as done in Coughlin et al., 2020 [21]), this restriction may be theoretically solved. (3) Supervised machine learning was used in every study mentioned here. Alternative approaches, such unsupervised and reinforcement learning, might provide more information on the prediction of smoking cessation. (4) Because two studies failed to identify the most reliable indicator of smoking cessation, they cannot be used to guide clinical interventions. By completely disclosing the predictors utilized and their individual predictive capacities, this barrier can be reduced.

We are aware of a number of potential limitations in the approach used for this review in addition to the general constraints of the field mentioned above. While we acknowledged that the MeSH subheaders for machine learning may not be comprehensive, we relied on them to identify search phrases for smoking cessation and machine learning. For instance, some writers do not consider logistic regression to be a machine-learning strategy. We operationalized machine learning as a paper annotated with the relevant MeSH subheader in this review. Therefore, our search was potentially limited because we only considered studies that were found using specific search terms (see Section 2.1 for search terms), while we omitted papers that used comparable methodology. The MeSH subheaders for tobacco usage have similar constraints. Second, although finding two references for this review in a review of relevant publications, a formal snow-ball search process was not used.

5. Conclusions

In conclusion, machine learning technologies are only partially able to demonstrate therapeutic effects and have not yet significantly improved the smoking cessation paradigm. The topic is still in its infancy, and there remains a crucial knowledge gap on the brain mechanisms underlying decision-making and behavior change. When examining neural-based decision-making processes, using decision-tree-based machine learning techniques may offer the most illuminating and comparable models to enhance therapeutic results.
References


50. AH Alsharif, N. Philip, Classifying and predicting instances for smoking cessation management system (Smoke mind), 2016 International Conference on Engineering & MIS (ICEMIS), IEEE, 2016, doi:10.1109/icemis.2016.7745360.


