
Review

Machine Learning Applications in Mental Health and Substance Use Research Among the LGBTQ2S+ Population: Scoping Review

Anasua Kundu¹, MSc; Michael Chaiton^{1,2}, PhD; Rebecca Billington³, MSW; Daniel Grace², PhD; Rui Fu⁴, PhD; Carmen Logie^{3,5}, PhD; Bruce Baskerville^{6,7}, PhD; Christina Yager¹, MSW; Nicholas Mitsakakis^{2,8}, P Stat, PhD; Robert Schwartz^{1,2,4}, PhD

¹Centre for Addiction and Mental Health, Toronto, ON, Canada

²Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

³Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

⁴Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada

⁵Women's College Research Institute, Toronto, ON, Canada

⁶Canadian Institutes of Health Research, Government of Canada, Ottawa, ON, Canada

⁷School of Pharmacy, Faculty of Science, University of Waterloo, Kitchener, ON, Canada

⁸Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada

Corresponding Author:

Anasua Kundu, MSc

Centre for Addiction and Mental Health

1000 Queen Street West

Toronto, ON, M6J 1H4

Canada

Phone: 1 6476326493

Email: anasua.kundu@mail.utoronto.ca

Abstract

Background: A high risk of mental health or substance addiction issues among sexual and gender minority populations may have more nuanced characteristics that may not be easily discovered by traditional statistical methods.

Objective: This review aims to identify literature studies that used machine learning (ML) to investigate mental health or substance use concerns among the lesbian, gay, bisexual, transgender, queer or questioning, and two-spirit (LGBTQ2S+) population and direct future research in this field.

Methods: The MEDLINE, Embase, PubMed, CINAHL Plus, PsycINFO, IEEE Xplore, and Summon databases were searched from November to December 2020. We included original studies that used ML to explore mental health or substance use among the LGBTQ2S+ population and excluded studies of genomics and pharmacokinetics. Two independent reviewers reviewed all papers and extracted data on general study findings, model development, and discussion of the study findings.

Results: We included 11 studies in this review, of which 81% (9/11) were on mental health and 18% (2/11) were on substance use concerns. All studies were published within the last 2 years, and most were conducted in the United States. Among mutually nonexclusive population categories, sexual minority men were the most commonly studied subgroup (5/11, 45%), whereas sexual minority women were studied the least (2/11, 18%). Studies were categorized into 3 major domains: web content analysis (6/11, 54%), prediction modeling (4/11, 36%), and imaging studies (1/11, 9%).

Conclusions: ML is a promising tool for capturing and analyzing hidden data on mental health and substance use concerns among the LGBTQ2S+ population. In addition to conducting more research on sexual minority women, different mental health and substance use problems, as well as outcomes and future research should explore newer environments, data sources, and intersections with various social determinants of health.

(*JMIR Med Inform* 2021;9(11):e28962) doi: [10.2196/28962](https://doi.org/10.2196/28962)

KEYWORDS

sexual and gender minorities; mental health; mental disorders; substance-related disorders; machine learning

Introduction

Background

Members of the lesbian, gay, bisexual, transgender, queer or questioning, and two-spirit (LGBTQ2S+) population experience significant mental health disparities and are at a higher risk of substance use problems compared with their heterosexual and cisgender peers [1-5]. A meta-analysis of 25 studies revealed that lesbian, gay, and bisexual individuals had 2.47 times increased lifetime risk of attempting suicide, 1.5 times increased risk of depression and anxiety disorders, and 1.5 times increased risk of alcohol and other substance dependence over a 12-month period [2]. Recent statistics from the 2015 National Survey on Drug Use and Health in the United States reported that the sexual minority population have an increased likelihood of past year use of illicit drugs, marijuana, and opioids; current use of cigarettes and alcohol; and past year diagnosis of any mental illness compared with sexual majority groups [6]. Members of the LGBTQ2S+ population also use mental health services and substance use treatment more frequently than cisgender and heterosexual individuals [6,7].

There is a robust evidence base documenting sexual orientation and gender identity as social determinants of health, whereby members of the LGBTQ2S+ population experience stressors from stigma, social, and economic exclusion that contribute to increased mental health challenges and resultant coping strategies, including problematic substance use [8-10]. In addition, intersecting experiences of marginalization such as race, ethnicity, disability, and homelessness; lack of familial and peer support; various acts of bullying, harassment, and hate crimes; and experience of self-stigmatization, such as internalized homophobia, biphobia, and transphobia, contribute to further deterioration of mental health and substance use concerns [8,11-16].

With advances in technology, novel statistical methods, such as machine learning (ML), have emerged as promising means of analyzing a vast range of complex data in public health informatics [17,18]. ML uses computational power to identify or *mine* hidden data patterns and has been increasingly used for content analysis and as a predictive modeling technique [17]. These characteristics are particularly important for investigating mental health and substance use issues among the LGBTQ2S+ population, where social stigma and institutional barriers make sexual and gender identity disclosure difficult, rendering the data invisible [19-21].

There are 3 major types of ML, including (1) supervised learning, (2) unsupervised learning, and (3) semisupervised learning. Supervised learning aims to learn from labeled data to predict the class of unlabeled input data or outcome variables [22]. Unsupervised learning does not require an outcome variable, thereby allowing the algorithm to freely detect and recognize hidden patterns with minimal human interference [22,23]. Semisupervised learning learns from both labeled and unlabeled data, where it can use readily available unlabeled data to improve supervised learning tasks when the labeled data are scarce or expensive [24]. A more advanced form of ML, deep learning, has gained popularity in health research in recent years

and uses an artificial neural network model with multiple layers to hierarchically define and process data [25]. These ML methods provide the opportunity to understand data more thoroughly and effectively, as well as yield meaningful predictions beyond traditional statistical methods.

Several reviews, including 3 recent systematic reviews, have been conducted to summarize the application of ML in substance use and mental health issues [23,26-28]. These systematic reviews have reported ML applications in 54 articles on mental health, 87 articles on suicidal behavior, and 17 articles on addiction research and reported good performance in predicting human behavior [23,26,28]. However, most of these reviews and studies focused on broad categories and the general population or patient records.

Objectives

Although one scoping review has explored studies that predict population-specific health with ML [29], the study did not identify ML applications among the LGBTQ2S+ population. There is a substantial gap in the literature, with no existing review focused on ML studies examining mental health and substance use among the LGBTQ2S+ population. As a result, we conducted a scoping review to address these knowledge gaps with the aim of mapping the current status of ML studies, focusing on this field and identifying the research gap to facilitate future research. Regarding persistent mental health and problematic substance use concerns and disparities among the LGBTQ2S+ population, the findings from this review will provide useful insights to inform research and programs.

Methods

Objectives and Methodology Framework

This review aims to conduct a comprehensive search of studies using ML to investigate mental health or substance use among LGBTQ2S+ communities and to determine the scope of future research. We used the following 5-stage methodological framework developed by Arksey and O'Malley [30]: (1) identifying specific research questions; (2) identifying relevant studies through a comprehensive search of different sources; (3) study selection by applying inclusion and exclusion criteria; (4) data charting using custom-made data extraction forms; and (5) collating, summarizing, and reporting the results. We also used an extension of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for scoping reviews [31] to present our findings, and the Joana Briggs Institute proposed methodology of scoping reviews [32] to narrate the implications for future research. The review protocol was registered on the Open Science Framework [33] on December 17, 2020, to facilitate transparency and reproducibility of the study.

Identifying Research Questions

Initially, we identified a broad set of preliminary questions for this scoping review:

- What is the volume of the literature that used machine learning analysis in the field of mental health and substance use among the LGBTQ2S+ population?

- What are the fields of mental health and substance use among the LGBTQ2S+ population that have been studied by machine learning?
- Which subgroups of the LGBTQ2S+ population have been investigated? Are there any specific subgroups that have been studied using machine learning analysis?
- What types of machine learning methods (eg, supervised, unsupervised, semisupervised, and deep learning) and algorithms (eg, decision trees, random forest, logistic regression, and penalized regression) have been used to study LGBTQ2S+ mental health and substance use?
- What are the real-world implications of these studies? Are there any knowledge gaps or untouched domains that should be addressed in future research?

Identifying Relevant Studies

To gather a large quantity of relevant literature, we followed previous review studies with similar objectives [27,29] and searched the following databases: MEDLINE (Ovid), Embase (Ovid), CINAHL Plus, APA PsycINFO (Ovid), PubMed, and IEEE Xplore. We also searched the Summon (ProQuest) database used by the University of Toronto Libraries, which searches across many other databases, journal packages, e-book collections, and other resources. Information technology databases such as IEEE Xplore were selected as a potential source of ML-related literature. Literature searches involved a combination of keywords (eg, *mental health, mental disease, mental health service, substance abuse, ML, sexual and gender minorities, LGBT, lesbian, gay, men who have sex with men, bisexual, queer, two-spirit, intersex, and transgender*) and medical subject headings, if applicable. A librarian was consulted regarding the keywords and search terms.

Two reviewers (AK and RB) conducted the database search from November 25 to December 13, 2020, and imported all citations to the Covidence web platform, where duplicate papers were removed automatically. The databases were searched from the date of inception of the databases to the year 2020, with no filter in place for publication year. The bibliography lists of the included studies and review papers were reviewed on December 13, 2020, to identify any potential studies. The full Embase search strategy, representing an example of the search query applied to all other databases, is presented in [Multimedia Appendix 1](#).

Study Selection

We included studies that used ML to investigate mental health or substance use behaviors of people within the LGBTQ2S+ population. Studies in which ML was used partially, but not for the main statistical analysis, were included in the review. We only included empirical investigations, thereby excluding editorials, opinion pieces, and reviews. We also excluded papers that used logistic regression analyses, not as a ML algorithm, but to determine LGBTQ2S+ identity status. In addition, studies in which full texts could not be retrieved with institutional

license, and studies of genomics, pharmacokinetics, and those that were not directly relevant to humans were excluded.

Two reviewers (AK and RB) independently screened each title and abstract based on the eligibility criteria and completed full-text screening of the remaining studies. Disagreements were resolved through discussions among the 3 reviewers (AK, RB, and MC) to yield a list of final included studies.

Data Charting

To facilitate data charting and reporting, individual reviewers (AK and RB) first reviewed all studies and extracted key phrases and concepts from each study. We based our data extraction items on features identified in a recent biomedical guideline for reporting ML studies [34]. Custom-made data extraction forms were developed from this guideline, which included major extraction categories such as general study characteristics (ie, author, year, country, target population, source of data, sample size, field of study, ML domains, ML methods, algorithms, and outcomes), key components of model development (ie, whether the studies discussed methods of feature selection, resampling, model performance metrics, and method of validation), and discussion of study findings (ie, importance ranking of features, intersectionality, and other procedures or features applied).

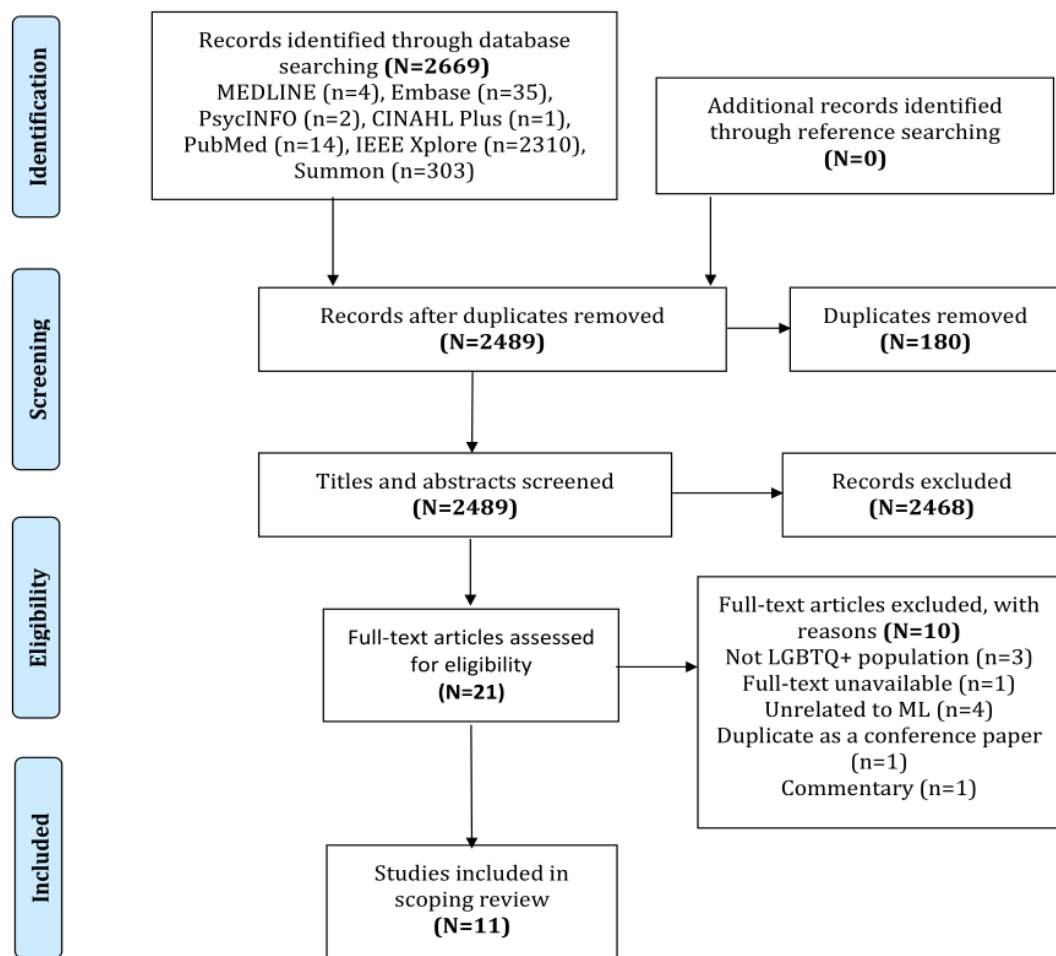
Collating, Summarizing and Reporting Results

We presented descriptive statistics for the extracted data sets by calculating the total number and percentage of all studies in each category. To provide a visual overview of the range of data, we presented a bar chart that showed the frequency analysis of studies according to the field of study and a pie chart that demonstrated the proportion of studies in the major domains of ML. We used a narrative synthesis approach [35] to describe the findings of the studies in the different ML domains and explored relationships in the data. Finally, we discussed research gaps to facilitate future research.

Results

The initial search of databases yielded 2669 articles, of which 2489 were retrieved after removing duplicates. We also searched the reference lists of potentially eligible articles and previous reviews but could not identify any studies that matched our inclusion criteria. After title and abstract screening, 21 articles were selected for full-text screening. Of these, we excluded articles that did not meet the target population criteria of the LGBTQ2S+ population (3/21, 14%), full-texts could not be retrieved (1/21, 4%), unrelated to ML (4/21, 19%), duplicate article published in a conference proceeding (1/21, 4%), and a commentary (1/21, 4%). This resulted in 11 studies being included in the final review [36-46]. The detailed selection process of the articles is presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram ([Figure 1](#)).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram documenting study exclusion. LGBTQ+: lesbian, gay, bisexual, transgender, queer, or questioning; ML: machine learning.



Study Characteristics

All 11 included studies [36-46] were published within the last 2 years (Table 1). Most of the studies were carried out in the United States (7/11, 63%) [36,38,39,41-43,45]. Among the target population categories that were not mutually exclusive, sexual minority men (gay, men who have sex with men,

bisexual) were the most commonly studied (5/11, 45%) subgroups [37,40,42-44], followed by transgender (3/11, 27%) [39,45,46] and LGBTQ+ (3/11, 27%) [36,38,41] people at large, whereas sexual minority women (lesbian and bisexual) (2/11, 18%) [43,45] were the least commonly represented populations. None of the studies included two-spirit persons as their target population (Table 1).

Table 1. Summary statistics of included studies (N=11) [36-46].^a

Characteristics	Number of studies, n (%)
Countries	
United States	7 (63)
China	2 (18)
Sweden	1 (9)
Australia	1 (9)
Years published	
2019	5 (45)
2020	6 (54)
Field of study	
Mental health (n=9)	
Suicide or self-injury	2 (18)
Depression	2 (18)
Mood or affect processes	3 (27)
Minority stress	1 (9)
Gender incongruence	1 (9)
Substance use (n=2)	
Tobacco	1 (9)
Poppers or alkyl nitrites	1 (9)
Target population^b	
Sexual minorities: male (gay, MSM ^c , bisexual)	5 (45)
Sexual minorities: female (lesbian, bisexual)	2 (18)
Transgender or gender minorities	3 (27)
LGBT/LGBTQ+ ^d	3 (27)
Domains of ML^e	
Web content analysis	6 (55)
Prediction modeling	4 (36)
Imaging study	1 (9)
Type of ML	
Supervised	9 (82)
Unsupervised	3 (27)
Deep	1 (9)
ML algorithms	
LDA ^f	3 (27)
RF ^g	2 (18)
SVM ^h	2 (18)
CNN ⁱ	1 (9)
MLP ^j	1 (9)
NB ^k	1 (9)
Penalized regression (LASSO ^l , elastic net regularized regression, ridge regression)	2 (18)
Logistic regression	1 (9)

Characteristics	Number of studies, n (%)
Boosting (XGBoost ^m , AdaBoost ⁿ , GBM ^o)	3 (27)
Classification tree	2 (18)
Feature selection	
Yes	7 (64)
No	4 (36)
Discussed model performance	
Used performance metrics	9 (82)
Didn't use performance metrics	1 (9)
Didn't discuss performance	1 (9)
Method of validation	
Hold-out	2 (18)
Cross-validation	7 (64)
External validation	2 (18)
Unspecified	4 (36)

^aMultiple response options were possible for some study characteristics.

^bCategories are not mutually exclusive.

^cMSM: men who have sex with men.

^dLGBT/LGBTQ+: lesbian, gay, bisexual, and transgender/lesbian, gay, bisexual, transgender, queer, or questioning.

^eML: machine learning.

^fLDA: latent Dirichlet allocation.

^gRF: random forest.

^hSVM: support vector machine.

ⁱCNN: convolutional neural network.

^jMLP: multilayered perceptron.

^kNB: Naive Bayes.

^lLASSO: least absolute shrinkage and selection operator.

^mXGBoost: eXtreme Gradient Boosting.

ⁿAdaBoost: Adaptive Boosting.

^oGBM: Generalized Boosted Model.

Most of the studies focused on mental health (9/11, 82%) [36-42,45,46], and only 18% (2/11) studies [43,44] focused on substance use concerns. Most studies examined several mental health issues, such as depression, suicide, mood or affect processes, minority stress, and gender incongruence [36-42,45,46], whereas other studies that focused on substance use only examined tobacco and poppers or alkyl nitrites use [43,44]. No study looked into mental health issues and substance use concerns among the LGBTQ2S+ population simultaneously (Table 1).

The studies were categorized into 3 major ML domains: web content analysis, prediction modeling, and imaging study. Over half of the studies (6/11, 55%) were identified as web content analysis [36-41], and 36% (4/11) were identified as prediction modeling [42-45]; 1 study (9%) was identified as an imaging study [46] (Table 1).

The most commonly used class of ML methods was supervised (9/11, 82%) [37-39,41-46], followed by unsupervised (3/11, 27%) [36,37,40] and deep learning (1/11, 9%; Table 1) [41]. The most frequently used ML algorithms were latent Dirichlet

allocation (3/11, 27%) and boosting (3/11, 27%), followed by random forest, support vector machines, penalized regression (ie, least absolute shrinkage and selection operator, elastic net regularized regression, and ridge regression), classification tree, logistic regression, naive Bayes, multilayered perceptron, and convolutional neural network (Table 1).

Approximately two-thirds (7/11, 64%) of the studies [37,38,42-46] discussed their methods of feature selection, among which the median number of features used was 19. Most of the studies used cross-validation methods (7/11, 64%) [37-39,41,44-46], especially 10-fold cross-validation. Furthermore, 18% (2/11) of the articles used the hold out method [39,41], 18% (2/11) used external validation [37,41], and 36% (4/11) articles [36,40,42,43] did not report how they validated their method. Most studies (9/11, 82%) [36-39,41-43,45,46] used at least one performance metric (eg, area under ROC curve, precision-recall, or F1 score) to discuss model performance. However, the remaining studies either did not use any performance metric [44] or did not discuss any model performance [40] (Table 1).

Machine Learning Domains

[Multimedia Appendix 2](#) summarizes the characteristics of the final 11 included studies [36-46] and [Multimedia Appendix 3](#) [36-46] presents the ML methodology used in the studies.

The 54% (6/11) studies [36-41] in the web content analysis domain obtained their data from social media sources such as Twitter, Blued, Tumblr, Reddit, and LGBT Chat and Forums. The volume of data used ranged from 12,000 to 41 million web posts. Half of the studies used their data to analyze the mood or affect processes of the users related to their sexual and gender identities [39-41] ([Multimedia Appendix 2](#)).

Among the 4 studies in the prediction modeling domain, 50% (2/4) of the studies analyzed data on adult participants [42,44] and 50% (2/4) on adolescents [43,45]. Only 1 study used a public health data set of 28,811 participants [43]; other studies used either cross-sectional or cohort data from longitudinal studies [42,44,45]. Half of the studies focused on mental health (depression and suicide) [42,45] and half on substance use behavior (cigarette, e-cigarette, and poppers use) [43,44] ([Multimedia Appendix 2](#)). Of the 4 studies, only 25% (1/4) study [45] ranked their feature importance, and 50% (2/4) studies [42,45] examined intersectionalities ([Multimedia Appendix 3](#)). One study investigated the intersection of income and other social and environmental stressors with racial or ethnic disparities and its impact on depressive symptomology among men who have sex with men [42], whereas the other focused on the intersection between various social and behavioral determinants of health (self-image, race, education, socioeconomic status, family support, friends, stigma, discrimination, etc) as risk factors of self-injurious behaviors among sexual and gender minority women [45].

One imaging trial study used clinical and functional magnetic resonance imaging data of 25 transgender adults to identify the relationship between pretherapy functional brain connectivity and posthormone therapy body congruence [46]. All 4 studies [42-45] of the prediction modeling domain and 1 imaging study [46] used the supervised method of ML, whereas studies in the web content analysis domain [36-41] used supervised (4/11, 36%), unsupervised (3/11, 27%), and deep learning (1/11, 9%) methods ([Multimedia Appendix 3](#)).

Discussion

Principal Findings

Our results show that the application of ML to assess mental health and substance use behavior among the LGBTQ2S+ population is still new in health research, compared with the increasing use of ML techniques in other health research domains. Although there is continued criminalization and lack of LGBTQ2S+ rights protection in 67 United Nations member states at the end of 2020 [47], there appears to be an increasing acceptance of sexual and gender minority people in diverse contexts such as in North American countries and Western Europe [48]. However, very few of the included studies were conducted outside the United States ([Table 1](#)).

Only a few mental health problems were addressed across the few relevant ML studies conducted to date ([Table 1](#)). Although

there is evidence of a higher prevalence of anxiety disorders, posttraumatic stress disorder, and various mood disorders (eg, mania and persistent depressive disorder) among the LGBTQ2S+ population compared with cisgender and heterosexual counterparts [4], no studies have been conducted on these issues. Compared with mental health issues, substance use problems among the LGBTQ2S+ population were almost untouched. Moreover, both of the included substance use related studies predicted the present use of substances [43,44], and no studies have examined future substance use, cessation, or substance use treatment-seeking behavior.

Underlying factors behind the low number of ML studies on mental health and substance use issues among the LGBTQ2S+ population may be sex and gender identity-related data invisibility and social and institutional bias [21,49]. Electronic health records have been used as a common and promising data source for ML techniques to predict population health in other research areas [27,29]. However, binary representation of sex and gender (ie, man or woman) in the electronic health records system makes some data unavailable for analysis by ML, which can underrepresent the actual problem [21,50,51]. Adopting inclusive gender, sex, and sexual orientation (GSSO) information practices, collecting sexual and gender diversity, has the potential to ensure data justice, alleviate unintentional bias, and reduce health inequity [49]. A good example of inclusive GSSO information practice could be the proposed equity stratifiers by the Canadian Institute of Health Information [52]. However, other potential data sources of ML applications, such as social media, cross-sectional survey data, longitudinal cohort, and administrative data sets were used in the included studies ([Multimedia Appendix 2](#)).

Most studies were in the web content analysis domain, indicating social media to be a potentially useful epidemiological resource for collecting data on LGBTQ2S+ people and analyzing the data using ML ([Multimedia Appendix 2](#)). We observed that unsupervised ML has also been applied in these studies with data drawn from social media [36,37,40], thus holding the potential to support qualitative research by handling large textual data sets with its computational power. This is particularly useful in LGBTQ2S+ health research, given the stigma-related and structural barriers toward identity disclosure that may inhibit data collection through other methodologies [50,51,53,54]. The use of ML in these studies has shown potential for automated identification of at-risk individuals for crisis suicide prevention and intervention [36], depressive emotions [37], minority stressors [38], negative emotions [40], and mental health signals [41] among the LGBTQ2S+ community. In addition, the sequence of transgender identity disclosure identified in a study by Haimson et al [39] may guide resource allocation and provide support through gender transition. However, self-reported mental health problems on social media might not reflect clinical diagnoses or symptomologies.

Although there is evidence of the influence of intersections of various social and behavioral determinants of health on the increased prevalence of mental health and substance use concerns among the LGBTQ2S+ population [11-16], only 2 studies examined the intersection of sexual and gender identity with ethno-racial identities, and several social, economic, and

behavioral factors (ie, income, social stigma, discrimination, and family support), and their impact on depression and self-injurious behaviors [42,45]. No such studies in our review explored intersectionality in the field of substance use. Identifying these intersections by leveraging ML techniques would have practical implications by determining risk and protective factors as well as informing strategies for promoting mental well-being and substance use prevention and intervention with and for LGBTQ2S+ people. In the context of various techniques used in intersectional research, both qualitative and quantitative, and recent trends in mixed methods research [55], ML can be a very useful tool for processing vast quantities of data, data mining and clustering, and classifying attribute relationships [56,57]. Apart from the partial dependency-based measures, newer techniques and methods [58,59] in ML have emerged for analyzing interaction effects and are more suitable for assessing intersectionality.

Following the current guidelines for reporting ML studies in biomedical research [34], we documented a range of explanatory findings seen in the included studies and found that most studies mentioned their performance metrics, method of feature selection, and method of validation of their model (Table 1 and Multimedia Appendix 3). However, only 27% (3/11) studies [37,38,45] adopted the approach of approximating a relative importance score of individual features that reflected their overall contributions to the model (Multimedia Appendix 3). The implications of providing an importance score to features are particularly valuable for predictive modeling studies, where the most important predictors are targeted for future strategy adoption. Another notable finding was about half (n=2) [42,43] of the predictive modeling studies did not report any method of validation, and none of them conducted external validation of the resulting model on a different population (Multimedia Appendix 3). Validation is an important aspect of the predictive modeling process, which increases the reproducibility and generalizability of the model [60]. Hence, future studies in this domain should follow existing guidelines to validate their models [34]. Moreover, half of the predictive modeling studies had small sample sizes (<1000) (Multimedia Appendix 2). Small data sets can affect the model performance [61]. Using large population-based data sets for future research can overcome this problem and fully leverage the benefits of ML.

Compared with the other 2 domains, there was a significant gap in ML research using imaging data (ie, functional magnetic resonance imaging or electroencephalography) to examine mental health and substance use among the LGBTQ2S+ population (Table 1). Although a single identified imaging study [46] predicted cross-sex hormonal therapy responsiveness in the transgender population, which is useful for guiding and selecting candidates for therapy, the sample size was small, limiting the generalizability of the findings.

Future Research Directions

We detected significant research gaps in ML applications for mental health and substance use research among the LGBTQ2S+ population. First, future research should investigate other mental health issues (ie, anxiety disorders and mood disorders) and substance use behavior and problems (ie, alcohol, opioids, and

illicit drugs) among the LGBTQ2S+ population. Second, the potential of ML applications in predicting substance use related outcomes (ie, cessation, overdose events, routes of administration, driving impairments, and other adverse reactions), mental health service access, and mental health-related outcomes (ie, disabilities, symptom management, suicide and suicide attempts, economic burden, and health care costs) should be explored.

Third, further research is needed on sexual minority women. The small number of studies included (Table 1) did not allow exploration of shared and different health needs and priorities between and within the LGBTQ2S+ population. Fourth, as the legal and societal context in which the LGBTQ2S+ population lives differ significantly between countries [48], more research should be conducted in countries outside the United States. Fifth, specific research initiatives targeted at investigating the intersection of sexual and gender minority identity with other social determinants of health (ie, race, ethnicity, citizenship, socioeconomic status, and housing condition) are necessary to better understand their potential for fostering risk and resilience regarding mental health and substance use. Finally, different data sources should be used in ML studies. Large-population-level administrative data sets should be used for prediction modeling studies for the accurate application of ML models. In addition, with the advancement of technology, the digitalization of health care, and where LGBTQ2S+ status is captured in electronic health records, these health records can be a potential data resource for ML studies with real-world clinical implications for LGBTQ2S+ people.

Strength and Limitations

To the best of our knowledge, our review is the first of its kind to explore the use of ML applications in examining mental health and substance use among LGBTQ2S+ populations. We adopted a comprehensive search strategy, including searching various multidisciplinary peer-reviewed databases to identify relevant articles as much as possible. The findings of our review need to be interpreted with consideration of one key limitation. Owing to the small number of studies, highly heterogeneous characteristics of the included studies, and inconsistent reporting of model development and validation, we could not perform a critical appraisal of the studies and therefore could not comment significantly on the overall performance of the ML techniques. However, we followed the approaches of previous scoping reviews with similar objectives [27,29] and were interested in understanding the general topics or areas being investigated by ML in the field of mental health and substance use among the LGBTQ2S+ population (ie, most commonly used data sources, study countries, and study populations) and identifying research gaps to inform future research.

As more studies are published on this research topic in the future, a systematic review with critical appraisal of relevant literatures should be conducted as the next step in research. Researchers are attempting to expand established reporting guidelines to include items that accommodate ML studies, such as the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement specific for M [62], the Artificial Intelligence extension for Consolidated

Standards of Reporting Trials [63], and Artificial Intelligence extension for Standard Protocol Items: Recommendations for Interventional Trials [63] guidelines. Once developed, these guidelines can be used as critical appraisal tools for studies that adopt ML-based data analysis. There is also an opportunity to incorporate fairness and equity considerations in the development of appraisal tools for ML studies. Preliminary research has already developed mathematical metrics to measure the fairness of a ML algorithm, and if intersectionalities are met in the models [64].

Conclusions

Although there is an exponential growth of ML applications in other health research sectors, few studies have used these techniques in the field of mental health and substance use among the LGBTQ2S+ population. In addition to undertaking more research, future researchers should focus on applying ML algorithms with considerations for real-world implications through public health interventions and adopting policies that aim to improve health equity.

Acknowledgments

The authors would like to thank Elena Springall, a librarian at the Gerstein Science Information Centre, University of Toronto, for her support in reviewing the database search strategies. The study was funded by the Canadian Institutes of Health Research, grant number 1000993. The funder had no role in the study design, collection, analysis, or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

Authors' Contributions

MC contributed to the study design and obtained funding and supervision. AK and RB conducted the database search, article screening, and data extraction. AK conducted the data analysis and primary drafting of the manuscript. All authors, AK, MC, RB, DG, RF, CHL, BB, CY, NM, and RS, contributed to the conceptualization, drafting, review, and approval of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Embase search query.

[\[DOCX File , 13 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Summary of studies using machine learning analysis in mental health and substance use among lesbian, gay, bisexual, transgender, queer or questioning, and two-spirit population (N=11).

[\[DOCX File , 17 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Summary of characteristics of machine learning methods used (N=11).

[\[DOCX File , 18 KB-Multimedia Appendix 3\]](#)

References

1. Marshal MP, Friedman MS, Stall R, King KM, Miles J, Gold MA, et al. Sexual orientation and adolescent substance use: a meta-analysis and methodological review. *Addiction* 2008 Apr;103(4):546-556 [\[FREE Full text\]](#) [doi: [10.1111/j.1360-0443.2008.02149.x](https://doi.org/10.1111/j.1360-0443.2008.02149.x)] [Medline: [18339100](https://pubmed.ncbi.nlm.nih.gov/18339100/)]
2. King M, Semlyen J, Tai SS, Killaspy H, Osborn D, Popelyuk D, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. *BMC Psychiatry* 2008 Aug 18;8:70 [\[FREE Full text\]](#) [doi: [10.1186/1471-244X-8-70](https://doi.org/10.1186/1471-244X-8-70)] [Medline: [18706118](https://pubmed.ncbi.nlm.nih.gov/18706118/)]
3. Marshal MP, Dietz LJ, Friedman MS, Stall R, Smith HA, McGinley J, et al. Suicidality and depression disparities between sexual minority and heterosexual youth: a meta-analytic review. *J Adolesc Health* 2011 Aug;49(2):115-123 [\[FREE Full text\]](#) [doi: [10.1016/j.jadohealth.2011.02.005](https://doi.org/10.1016/j.jadohealth.2011.02.005)] [Medline: [21783042](https://pubmed.ncbi.nlm.nih.gov/21783042/)]
4. Institute of Medicine. *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding*. Washington (DC): National Academies Press (US); 2011.
5. National survey on LGBTQ youth mental health. The Trevor Project. 2019. URL: <https://www.thetrevorproject.org/wp-content/uploads/2019/06/The-Trevor-Project-National-Survey-Results-2019.pdf> [accessed 2021-10-19]
6. Medley G, Lipari R, Bose J, Cribb D, Kroutil L. Sexual orientation and estimates of adult substance use and mental health: results from the 2015 national survey on drug use and health. *National Survey on Drug Use and Health*. 2016. URL: <https://www.samhsa.gov/data/2k16/national-survey-on-drug-use-and-health-2015>

- [/www.samhsa.gov/data/sites/default/files/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015.htm](http://www.samhsa.gov/data/sites/default/files/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015.htm) [accessed 2021-10-19]
7. Abramovich A, de Oliveira C, Kiran T, Iwajomo T, Ross LE, Kurdyak P. Assessment of health conditions and health service use among transgender patients in Canada. *JAMA Netw Open* 2020 Aug 03;3(8):e2015036 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.15036](https://doi.org/10.1001/jamanetworkopen.2020.15036)] [Medline: [32857149](https://pubmed.ncbi.nlm.nih.gov/32857149/)]
 8. Wilson C, Cariola L. LGBTQI+ youth and mental health: a systematic review of qualitative research. *Adolescent Res Rev* 2019 May 21;5(2):187-211 [FREE Full text] [doi: [10.1007/s40894-019-00118-w](https://doi.org/10.1007/s40894-019-00118-w)]
 9. Logie C. The case for the World Health Organization's commission on the social determinants of health to address sexual orientation. *Am J Public Health* 2012 Jul;102(7):1243-1246. [doi: [10.2105/AJPH.2011.300599](https://doi.org/10.2105/AJPH.2011.300599)] [Medline: [22594723](https://pubmed.ncbi.nlm.nih.gov/22594723/)]
 10. Pega F, Veale JF. The case for the World Health Organization's Commission on Social Determinants of Health to address gender identity. *Am J Public Health* 2015 Mar;105(3):58-62. [doi: [10.2105/AJPH.2014.302373](https://doi.org/10.2105/AJPH.2014.302373)] [Medline: [25602894](https://pubmed.ncbi.nlm.nih.gov/25602894/)]
 11. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychol Bull* 2003 Sep;129(5):674-697 [FREE Full text] [doi: [10.1037/0033-2909.129.5.674](https://doi.org/10.1037/0033-2909.129.5.674)] [Medline: [12956539](https://pubmed.ncbi.nlm.nih.gov/12956539/)]
 12. Ryan C, Huebner D, Diaz RM, Sanchez J. Family rejection as a predictor of negative health outcomes in white and Latino lesbian, gay, and bisexual young adults. *Pediatrics* 2009 Jan;123(1):346-352. [doi: [10.1542/peds.2007-3524](https://doi.org/10.1542/peds.2007-3524)] [Medline: [19117902](https://pubmed.ncbi.nlm.nih.gov/19117902/)]
 13. Burns MN, Ryan DT, Garofalo R, Newcomb ME, Mustanski B. Mental health disorders in young urban sexual minority men. *J Adolesc Health* 2015 Jan;56(1):52-58 [FREE Full text] [doi: [10.1016/j.jadohealth.2014.07.018](https://doi.org/10.1016/j.jadohealth.2014.07.018)] [Medline: [25294230](https://pubmed.ncbi.nlm.nih.gov/25294230/)]
 14. Duncan DT, Hatzenbuehler ML. Lesbian, gay, bisexual, and transgender hate crimes and suicidality among a population-based sample of sexual-minority adolescents in Boston. *Am J Public Health* 2014 Feb;104(2):272-278. [doi: [10.2105/AJPH.2013.301424](https://doi.org/10.2105/AJPH.2013.301424)] [Medline: [24328619](https://pubmed.ncbi.nlm.nih.gov/24328619/)]
 15. Kosciw J, Greytak E, Palmer N, Boesen M. The 2013 National School Climate Survey: the experiences of lesbian, gay, bisexual and transgender youth in our nation's schools. GLSEN. 2014. URL: <https://www.glsen.org/research/2013-national-school-climate-survey> [accessed 2021-10-19]
 16. Choi S, Wilson B, Shelton J, Gates G. Serving our youth 2015: the needs and experiences of lesbian, gay, bisexual, transgender, and questioning youth experiencing homelessness. The Williams Institute with True Colors Fund. 2015. URL: <https://scholarship.org/uc/item/1pd9886n> [accessed 2021-10-19]
 17. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015 Jul 17;349(6245):255-260. [doi: [10.1126/science.aaa8415](https://doi.org/10.1126/science.aaa8415)] [Medline: [26185243](https://pubmed.ncbi.nlm.nih.gov/26185243/)]
 18. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. *Biomed Inform Insights* 2016;8:1-10 [FREE Full text] [doi: [10.4137/BII.S31559](https://doi.org/10.4137/BII.S31559)] [Medline: [26843812](https://pubmed.ncbi.nlm.nih.gov/26843812/)]
 19. Let's discuss stigma and discrimination around mental health and substance use problems. Canadian Mental Health Association, British Columbia Division. 2014. URL: <https://www.heretohelp.bc.ca/sites/default/files/stigma-and-discrimination-around-mental-health-and-substance-use-problems.pdf> [accessed 2021-10-19]
 20. Committee on the Science of Changing Behavioral Health Social Norms, Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine. Ending Discrimination Against People with Mental and Substance Use Disorders: The Evidence for Stigma Change. Washington (DC): National Academies Press (US); 2016:1-170.
 21. Ruberg B, Ruelos S. Data for queer lives: how LGBTQ gender and sexuality identities challenge norms of demographics. *Big Data Soc* 2020 Jun 18;7(1):205395172093328. [doi: [10.1177/2053951720933286](https://doi.org/10.1177/2053951720933286)]
 22. Naqa IE, Murphy MJ. What is machine learning? In: Naqa IE, Li R, Murphy MJ, editors. *Machine Learning in Radiation Oncology*. Cham: Springer; 2015:3-11.
 23. Mak KK, Lee K, Park C. Applications of machine learning in addiction studies: a systematic review. *Psychiatry Res* 2019 May;275:53-60. [doi: [10.1016/j.psychres.2019.03.001](https://doi.org/10.1016/j.psychres.2019.03.001)] [Medline: [30878857](https://pubmed.ncbi.nlm.nih.gov/30878857/)]
 24. Zhu X, Goldberg AB. Introduction to semi-supervised learning. *Synth Lect Artif Intell Mach Learn* 2009 Jan;3(1):1-130 [FREE Full text] [doi: [10.2200/s00196ed1v01y200906aim006](https://doi.org/10.2200/s00196ed1v01y200906aim006)]
 25. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015 May 28;521(7553):436-444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)] [Medline: [26017442](https://pubmed.ncbi.nlm.nih.gov/26017442/)]
 26. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans Comput-Hum Interact* 2020 Oct 05;27(5):1-53 [FREE Full text] [doi: [10.1145/3398069](https://doi.org/10.1145/3398069)]
 27. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 2019 Jul;49(9):1426-1448. [doi: [10.1017/S0033291719000151](https://doi.org/10.1017/S0033291719000151)] [Medline: [30744717](https://pubmed.ncbi.nlm.nih.gov/30744717/)]
 28. Bernert R, Hilberg A, Melia R, Kim J, Shah N, Abnoui F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *Int J Environ Res Public Health* 2020 Aug 15;17(16):5929 [FREE Full text] [doi: [10.3390/ijerph17165929](https://doi.org/10.3390/ijerph17165929)] [Medline: [32824149](https://pubmed.ncbi.nlm.nih.gov/32824149/)]

29. Morgenstern JD, Buajitti E, O'Neill M, Piggott T, Goel V, Fridman D, et al. Predicting population health with machine learning: a scoping review. *BMJ Open* 2020 Oct 27;10(10):e037860 [FREE Full text] [doi: [10.1136/bmjopen-2020-037860](https://doi.org/10.1136/bmjopen-2020-037860)] [Medline: [33109649](https://pubmed.ncbi.nlm.nih.gov/33109649/)]
30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol Theory Pract* 2005 Feb;8(1):19-32 [FREE Full text] [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
31. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018 Oct 02;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
32. Peters M, Godfrey C, McInerney P, Soares C, Khalil H, Parker D. The Joanna Briggs Institute Reviewers' Manual 2015: methodology for JBI scoping reviews. Joanna Briggs Institute. 2015. URL: <https://nursing.lsuhsoc.edu/JBI/docs/ReviewersManuals/Scoping-.pdf> [accessed 2021-10-19]
33. Kundu A, Billington R, Chaiton M. Machine learning applications in mental health and substance use research among LGBTQ2S+ population: protocol for a scoping review. *Open Sci Framework* 2020:A. [doi: [10.17605/OSF.IO/TMPV3](https://doi.org/10.17605/OSF.IO/TMPV3)]
34. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016 Dec 16;18(12):e323 [FREE Full text] [doi: [10.2196/jmir.5870](https://doi.org/10.2196/jmir.5870)] [Medline: [27986644](https://pubmed.ncbi.nlm.nih.gov/27986644/)]
35. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme - Version 1. Peninsula Medical School, Universities of Exeter and Plymouth. 2006. URL: <https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf> [accessed 2021-10-19]
36. Liang C, Abbott D, Hong Y, Madadi M, White A. Clustering help-seeking behaviors in LGBT online communities: a prospective trial. In: Meiselwitz G, editor. *Social Computing and Social Media. Design, Human Behavior and Analytics*. Cham: Springer; 2019:345-355.
37. Li Y, Cai M, Qin S, Lu X. Depressive emotion detection and behavior analysis of men who have sex with men social media. *Front Psychiatry* 2020;11:830 [FREE Full text] [doi: [10.3389/fpsy.2020.00830](https://doi.org/10.3389/fpsy.2020.00830)] [Medline: [32922323](https://pubmed.ncbi.nlm.nih.gov/32922323/)]
38. Saha K, Kim SC, Reddy MD, Carter AJ, Sharma E, Haimson OL, et al. The language of LGBTQ+ minority stress experiences on social media. *Proc ACM Hum Comput Interact* 2019 Nov;3(CSCW):89 [FREE Full text] [doi: [10.1145/3361108](https://doi.org/10.1145/3361108)] [Medline: [32935081](https://pubmed.ncbi.nlm.nih.gov/32935081/)]
39. Haimson OL, Veinot TC. Coming out to doctors, coming out to "Everyone": understanding the average sequence of transgender identity disclosures using social media data. *Transgend Health* 2020;5(3):158-165 [FREE Full text] [doi: [10.1089/trgh.2019.0045](https://doi.org/10.1089/trgh.2019.0045)] [Medline: [32923666](https://pubmed.ncbi.nlm.nih.gov/32923666/)]
40. Huang G, Cai M, Lu X. Inferring opinions and behavioral characteristics of gay men with large scale multilingual text from blues. *Int J Environ Res Public Health* 2019 Sep 26;16(19):3597 [FREE Full text] [doi: [10.3390/ijerph16193597](https://doi.org/10.3390/ijerph16193597)] [Medline: [31561423](https://pubmed.ncbi.nlm.nih.gov/31561423/)]
41. Zhao Y, Guo Y, He X, Wu Y, Yang X, Prospero M, et al. Assessing mental health signals among sexual and gender minorities using Twitter data. *Health Informatics J* 2020 Jun;26(2):765-786 [FREE Full text] [doi: [10.1177/1460458219839621](https://doi.org/10.1177/1460458219839621)] [Medline: [30969146](https://pubmed.ncbi.nlm.nih.gov/30969146/)]
42. Barrett B, Abraham A, Dean L, Plankey M, Friedman M, Jacobson L, et al. Social inequalities contribute to racial/ethnic disparities in depressive symptomology among men who have sex with men. *Soc Psychiatry Psychiatr Epidemiol* 2021 Feb;56(2):259-272 [FREE Full text] [doi: [10.1007/s00127-020-01940-7](https://doi.org/10.1007/s00127-020-01940-7)] [Medline: [32780176](https://pubmed.ncbi.nlm.nih.gov/32780176/)]
43. Azagba S, Latham K, Shan L. Cigarette smoking, e-cigarette use, and sexual identity among high school students in the USA. *Eur J Pediatr* 2019 Sep;178(9):1343-1351. [doi: [10.1007/s00431-019-03420-w](https://doi.org/10.1007/s00431-019-03420-w)] [Medline: [31292730](https://pubmed.ncbi.nlm.nih.gov/31292730/)]
44. Demant D, Oviedo-Trespalacios O. Harmless? A hierarchical analysis of poppers use correlates among young gay and bisexual men. *Drug Alcohol Rev* 2019 Jul;38(5):465-472. [doi: [10.1111/dar.12958](https://doi.org/10.1111/dar.12958)] [Medline: [31209963](https://pubmed.ncbi.nlm.nih.gov/31209963/)]
45. Smith DM, Wang SB, Carter ML, Fox KR, Hooley JM. Longitudinal predictors of self-injurious thoughts and behaviors in sexual and gender minority adolescents. *J Abnorm Psychol* 2020 Jan;129(1):114-121. [doi: [10.1037/abn0000483](https://doi.org/10.1037/abn0000483)] [Medline: [31657599](https://pubmed.ncbi.nlm.nih.gov/31657599/)]
46. Moody T, Feusner J, Reggente N, Vanhoecke J, Holmberg M, Manzouri A, et al. Predicting outcomes of cross-sex hormone therapy in transgender individuals with gender incongruence based on pre-therapy resting-state brain connectivity. *Neuroimage Clin* 2021;29:102517 [FREE Full text] [doi: [10.1016/j.nicl.2020.102517](https://doi.org/10.1016/j.nicl.2020.102517)] [Medline: [33340976](https://pubmed.ncbi.nlm.nih.gov/33340976/)]
47. Mendos L, Botha K, Lelis R, Tan D, de la Peña E, Savelev I, et al. State-sponsored homophobia : global legislation overview update. ILGA, Geneva. 2020. URL: https://ilga.org/downloads/ILGA_World_State_Sponsored_Homophobia_report_global_legislation_overview_update_December_2020.pdf [accessed 2021-10-19]
48. Poushter J, Kent N. The global divide on homosexuality persists, but increasing acceptance in many countries over past two decades. Pew Research Center. 2020. URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2020/06/PG_2020.06.25_Global-Views-Homosexuality_FINAL.pdf [accessed 2021-10-19]

49. Davison K, Queen R, Lau F, Antonio M. Culturally competent gender, sex, and sexual orientation information practices and electronic health records: rapid review. *JMIR Med Inform* 2021 Feb 11;9(2):e25467 [FREE Full text] [doi: [10.2196/25467](https://doi.org/10.2196/25467)] [Medline: [33455901](https://pubmed.ncbi.nlm.nih.gov/33455901/)]
50. Sokkary N, Awad H, Paulo D. Frequency of sexual orientation and gender identity documentation after electronic medical record modification. *J Pediatr Adolesc Gynecol* 2021 Jun;34(3):324-327 [FREE Full text] [doi: [10.1016/j.jpag.2020.12.009](https://doi.org/10.1016/j.jpag.2020.12.009)] [Medline: [33333261](https://pubmed.ncbi.nlm.nih.gov/33333261/)]
51. Lau F, Antonio M, Davison K, Queen R, Bryski K. An environmental scan of sex and gender in electronic health records: analysis of public information sources. *J Med Internet Res* 2020 Nov 11;22(11):e20050 [FREE Full text] [doi: [10.2196/20050](https://doi.org/10.2196/20050)] [Medline: [33174858](https://pubmed.ncbi.nlm.nih.gov/33174858/)]
52. Canadian Institute for Health Information. In Pursuit of Health Equity: Defining Stratifiers for Measuring Health Inequality - A Focus on Age, Sex, Gender, Income, Education and Geographic Location. Ottawa, ON: CIHI; 2018.
53. Owen-Smith AA, Woodyatt C, Sineath RC, Hunkeler EM, Barnwell LT, Graham A, et al. Perceptions of barriers to and facilitators of participation in health research among transgender people. *Transgend Health* 2016;1(1):187-196 [FREE Full text] [doi: [10.1089/trgh.2016.0023](https://doi.org/10.1089/trgh.2016.0023)] [Medline: [28861532](https://pubmed.ncbi.nlm.nih.gov/28861532/)]
54. Lucassen M, Fleming T, Merry S. Tips for research recruitment: the views of sexual minority youth. *J LGBT Youth* 2017 Jan 13;14(1):16-30 [FREE Full text] [doi: [10.1080/19361653.2016.1256246](https://doi.org/10.1080/19361653.2016.1256246)]
55. Hankivsky O, Grace D. Understanding and emphasizing difference and intersectionality in multimethod and mixed methods research. In: Hesse-Biber SN, Johnson RB, editors. *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*. Oxford, United Kingdom: Oxford University Press; 2015.
56. Pastrana JL, Reigal RE, Morales-Sánchez V, Morillo-Baro JP, de Mier RJ, Alves J, et al. Data mining in the mixed methods: application to the study of the psychological profiles of athletes. *Front Psychol* 2019;10:2675 [FREE Full text] [doi: [10.3389/fpsyg.2019.02675](https://doi.org/10.3389/fpsyg.2019.02675)] [Medline: [31866896](https://pubmed.ncbi.nlm.nih.gov/31866896/)]
57. Leavitt A. Human-centered data science: mixed methods and intersecting evidence, inference, and scalability. Annenberg School for Communication & Journalism, University of Southern California. 2016. URL: https://cscw2016hcds.files.wordpress.com/2015/10/29_alexleavitt.pdf [accessed 2021-10-19]
58. Schiltz F, Masci C, Agasisti T, Horn D. Using regression tree ensembles to model interaction effects: a graphical approach. *Appl Econ* 2018 Jul 05;50(58):6341-6354 [FREE Full text] [doi: [10.1080/00036846.2018.1489520](https://doi.org/10.1080/00036846.2018.1489520)]
59. Oh S. Feature interaction in terms of prediction performance. *Appl Sci* 2019 Nov 29;9(23):5191 [FREE Full text] [doi: [10.3390/app9235191](https://doi.org/10.3390/app9235191)]
60. Han K, Song K, Choi BW. How to develop, validate, and compare clinical prediction models involving radiological parameters: study design and statistical methods. *Korean J Radiol* 2016;17(3):339-350 [FREE Full text] [doi: [10.3348/kjr.2016.17.3.339](https://doi.org/10.3348/kjr.2016.17.3.339)] [Medline: [27134523](https://pubmed.ncbi.nlm.nih.gov/27134523/)]
61. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014 Dec 22;14:137 [FREE Full text] [doi: [10.1186/1471-2288-14-137](https://doi.org/10.1186/1471-2288-14-137)] [Medline: [25532820](https://pubmed.ncbi.nlm.nih.gov/25532820/)]
62. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. *Lancet* 2019 Apr;393(10181):1577-1579. [doi: [10.1016/s0140-6736\(19\)30037-6](https://doi.org/10.1016/s0140-6736(19)30037-6)]
63. CONSORT-AISPIRIT-AI Steering Group. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med* 2019 Oct 24;25(10):1467-1468. [doi: [10.1038/s41591-019-0603-3](https://doi.org/10.1038/s41591-019-0603-3)] [Medline: [31551578](https://pubmed.ncbi.nlm.nih.gov/31551578/)]
64. Foulds J, Islam R, Keya K, Pan S. An intersectional definition of fairness. arXiv. 2019 Sep 10. URL: [http://jfoulds.informationsystems.umbc.edu/papers/2020/Foulds%20\(2020\)%20-%20An%20Intersectional%20Definition%20of%20Fairness%20\(ICDE\).pdf](http://jfoulds.informationsystems.umbc.edu/papers/2020/Foulds%20(2020)%20-%20An%20Intersectional%20Definition%20of%20Fairness%20(ICDE).pdf) [accessed 2019-10-19]

Abbreviations

LGBTQ2S+: lesbian, gay, bisexual, transgender, queer or questioning, and two-spirit

ML: machine learning

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by C Lovis; submitted 20.03.21; peer-reviewed by M Cai, K Davison; comments to author 10.07.21; revised version received 02.09.21; accepted 03.10.21; published 11.11.21

Please cite as:

Kundu A, Chaiton M, Billington R, Grace D, Fu R, Logie C, Baskerville B, Yager C, Mitsakakis N, Schwartz R

Machine Learning Applications in Mental Health and Substance Use Research Among the LGBTQ2S+ Population: Scoping Review
JMIR Med Inform 2021;9(11):e28962

URL: <https://medinform.jmir.org/2021/11/e28962>

doi: [10.2196/28962](https://doi.org/10.2196/28962)

PMID:

©Anasua Kundu, Michael Chaiton, Rebecca Billington, Daniel Grace, Rui Fu, Carmen Logie, Bruce Baskerville, Christina Yager, Nicholas Mitsakakis, Robert Schwartz. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 11.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.