informatics

MDPI

Article

Predicting Mental Health Problems in Gay Men in Peru Using Machine Learning and Deep Learning Models

Alejandro Aybar-Flores [1] and Elizabeth Espinoza-Portilla [2,*]

[1] Department of Engineering, Universidad del Pacífico, Lima 15072, Peru; a.aybarf@up.edu.pe
[2] Faculty of Health Sciences, School of Medicine, Universidad Continental, Lima 15046, Peru
* Correspondence: eespinozap@continental.edu.pe; Tel.: +51-213-276-0

Abstract

Mental health disparities among those who self-identify as gay men in Peru remain a pressing public health concern, yet predictive models for early identification remain limited. This research aims to (1) develop machine learning and deep learning models to predict mental health issues in those who self-identify as gay men, and (2) evaluate the influence of demographic, economic, health-related, behavioral and social factors using interpretability techniques to enhance understanding of the factors shaping mental health outcomes. A dataset of 2186 gay men from the First Virtual Survey for LGBTIQ+ People in Peru (2017) was analyzed, considering demographic, economic, health-related, behavioral, and social factors. Several classification models were developed and compared, including Logistic Regression, Artificial Neural Networks, Random Forest, Gradient Boosting Machines, eXtreme Gradient Boosting, and a One-dimensional Convolutional Neural Network (1D-CNN). Additionally, the Shapley values and Layer-wise Relevance Propagation (LRP) heatmaps methods were used to evaluate the influence of the studied variables on the prediction of mental health issues. The results revealed that the 1D-CNN model demonstrated the strongest performance, achieving the highest classification accuracy and discrimination capability. Explainability analyses underlined prior infectious diseases diagnosis, access to medical assistance, experiences of discrimination, age, and sexual identity expression as key predictors of mental health outcomes. These findings suggest that advanced predictive techniques can provide valuable insights for identifying at-risk individuals, informing targeted interventions, and improving access to mental health care. Future research should refine these models to enhance predictive accuracy, broaden applicability, and support the integration of artificial intelligence into public health strategies aimed at addressing the mental health needs of this population.

Keywords: mental health; gay men; machine learning; deep learning; Shapley values; LRP; Peru

check for updates

Academic Editor: Kamran Sedig

Received: 8 May 2025
Revised: 12 June 2025
Accepted: 26 June 2025
Published: 2 July 2025

Citation: Aybar-Flores, A.; Espinoza-Portilla, E. Predicting Mental Health Problems in Gay Men in Peru Using Machine Learning and Deep Learning Models. *Informatics* **2025**, *12*, 60. https://doi.org/10.3390/informatics12030060

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mental health refers to a person's emotional and psychological state, influencing their ability to cope with stress, maintain relationships, and carry out daily activities [1]. Mental health disorders can manifest through emotional distress, cognitive difficulties, and behavioral changes, often requiring professional intervention when they interfere with daily functioning [2]. External stressors such as trauma, chronic stress, and social adversity are commonly associated with the onset or exacerbation of such conditions [3].

A multifaceted understanding of mental health allows for more effective identification of risks, the development of targeted interventions, and improved treatment outcomes [4].

Social identity significantly shapes mental health outcomes, particularly by influencing exposure to stress and access to support. For sexual minorities, these psychosocial stressors are often magnified, increasing vulnerability and creating unique challenges that impact psychological outcomes [2,3,5]. Although many non-heterosexual individuals maintain overall well-being, studies consistently show elevated risks of psychological distress, depression, anxiety, and substance use in these populations. These disparities are largely explained by minority stress theory, which attributes mental health differences to stressors unique to stigmatized social communities—including enacted stigma (such as social rejection, discrimination and victimization) and institutional stigma (including legal exclusion and restricted access to healthcare) [5,6]. Across various sociocultural contexts where such stigma persists, mental health disparities may further be exacerbated [6].

In Peru, structural inequality and social exclusion remain significant barriers to mental health for LGBTIQ+ populations. Mental health issues have become an increasing public health concern in Peru, with anxiety and depression affecting nearly a third of the general population [7]. Suicide rates have also seen a steady rise, particularly among young people [7]. However, these challenges are markedly intensified among Lesbian, Gay, Bisexual, Transgender, Intersex, Queer, and non-binary, pansexual, and asexual (LGBTIQ+) individuals, who face unique psychosocial stressors that place them at a heightened risk for mental health disorders. According to the 2017 First Virtual Survey for LGBTIQ+ People in Peru, 54.4% reported having mental health problems, with anxiety and depression being the most prevalent conditions [8]. Discrimination, reported by 63% of respondents, acts as a chronic source of emotional distress, social isolation, and internalized stigma [7]. Additionally, the lack of inclusive and affirming mental health services severely limits access to care, further reinforcing health disparities [9].

Among sexual minorities, gay men in Peru face disproportionately high rates of psychological distress. Research conducted by Más Igualdad Perú reveals that nearly 30% of gay men report suicidal ideation, and a significant portion suffer from depression and anxiety disorders [9]. Experiences of family rejection, workplace discrimination, and social stigma exacerbate these outcomes. Alarmingly, harmful practices such as sexual orientation conversion efforts remain prevalent, with 36.2% of gay men reporting having been subjected to such interventions [9]. These experiences, compounded by fear of discrimination in clinical settings, not only cause profound emotional trauma but also often discourage individuals from seeking professional help.

The intersection of stigma and systemic healthcare shortcomings exacerbates these issues. Over half (58.4%) of gay men in Peru report having encountered prejudice or misinformation from mental health professionals, reflecting a systemic lack of cultural competence [9]. Additionally, the country's underfunding of mental health services, allocating only 0.1% of its Gross Domestic Product (GDP), limits the availability of specialized psychological care [9]. The combination of social exclusion and institutional neglect increases the risk of chronic mental health conditions and self-harm among vulnerable populations [7].

Given these challenges, there is an urgent need for more effective strategies to detect and manage mental health disorders in this population. The implementation of affirmative psychological care, community-based mental health programs, and targeted policy reforms could significantly improve outcomes for gay men and the broader LGBTIQ+ community [7]. Without immediate intervention, mental health disparities will continue to deepen, leaving vulnerable populations without the necessary support to ensure their well-being.

Currently, traditional statistical approaches in healthcare often offer only broad and linear assessments, limiting their capacity to capture the multifactorial nature of mental health problems [10]. These limitations underscore the need for more sophisticated analytical frameworks capable of integrating diverse types of data and uncovering complex associations. In this context, emerging developments in Machine Learning (ML) and Deep Learning (DL) have introduced new opportunities to enhance prediction, diagnosis, and decision-making in health research [10,11].

ML and DL methods have shown growing promise in the mental health domain. By learning from large and heterogeneous datasets, these techniques can support early detection of mental health risks, guide treatment strategies, and help identify at-risk individuals more effectively than conventional tools [12,13]. DL models in particular can model nonlinear relationships and capture hidden patterns across high-dimensional data—properties that are especially relevant in understanding the complexity of mental disorders [14]. These capabilities offer complementary insights to traditional analyses.

Recently, ML and DL have also been applied to better understand mental health outcomes among LGBTIQ+ populations [15]. The psychosocial stressors affecting these communities are often underrepresented in traditional datasets, making predictive modeling an appealing strategy to reveal underlying risk structures. Although this line of research remains relatively new, several studies have explored how AI can assist in identifying high-risk profiles and informing inclusive public health responses [15,16].

Nonetheless, considering the existing literature, only a limited number of mental health problems have been addressed across the few relevant ML/DL studies conducted to date. Although there is evidence of a higher prevalence of mental disorders among the LGBTIQ+ population compared with cisgender and heterosexual counterparts, a small number of studies have been conducted on these issues [15]. In light of this, to the best of our knowledge, no study in Peru has explored AI techniques to predict mental health problems among gay men. Therefore, this study addresses that gap by integrating predictive modeling and explainability methods to enhance the understanding of factors associated with mental distress in this population.

This study addresses that gap by leveraging data from the First Virtual Survey for LGBTIQ+ People in Peru to: (1) develop machine learning and deep learning models that predict mental health problems in individuals who self-identify as gay men, and (2) apply explainability techniques to evaluate the influence of demographic, economic, behavioral, and psychosocial factors. By uncovering hidden patterns in the data, our approach seeks to improve early identification of high-risk individuals and inform policies that more effectively address the specific mental health challenges faced by this population.

2. Literature Review

2.1. Sexual Stigma and Social Determinants of Mental Health

Sexual stigma, societal bias against non-heterosexual orientations, shapes mental health outcomes by exposing individuals to chronic stressors such as rejection, discrimination, and institutional heterosexism [5]. Although many sexual minorities maintain overall well-being, they face higher rates of depression, anxiety, and substance use when compared to heterosexual peers, a gap that narrows once discrimination is accounted for [5,17]. Beyond external prejudice, internalized negativity, often termed internalized homophobia or heterosexism, fosters low self-esteem, demoralization, and relationship instability, further elevating psychological risk [18,19].

These identity-related pressures intersect with wider social determinants. Economic hardship, limited education, and job insecurity correlate strongly with common mental disorders, with the poorest groups bearing the greatest burden [6]. In older adults, social isolation compounds these effects, while women often experience steeper socioeconomic gradients in mental health [20,21]. The relationship is bidirectional: mental illness can impair work and earning capacity, which, in turn, deepens vulnerability to distress and illness [6]. At the same time, community and institutional supports—such as social networks, inclusive policies, and accessible services—play a protective role, buffering the impact of adversity and sustaining well-being over the life course.

Crucially, access to mental healthcare remains uneven. Populations in rural or low-income settings confront shortages of services and financial barriers, leading to untreated conditions and worse outcomes [6]. Addressing these gaps requires not only expanding clinical capacity but also integrating mental health into primary care and community programs. Reducing disparities demands collaboration across sectors—health, education, housing, and social welfare—to tackle both immediate care needs and the upstream factors that shape psychological well-being throughout life [22,23].

2.2. Machine Learning and Deep Learning for Prediction of Mental Health

ML has been applied to multiple areas of psychological treatments and has demonstrated significant potential in predicting and diagnosing mental health conditions, including depression, anxiety, and schizophrenia, and related health outcomes [12]. Among the various ML methods, supervised learning is widely used for the prediction of mental illnesses. These techniques build predictive models to detect patterns linked to mental health disorders, enabling early risk assessment and tailored treatment strategies [12].

Researchers have applied supervised algorithms in mental health studies by evaluating behavioral and physiological data [12,24]. It is a growing trend in analyzing data from electronic health records (EHRs) to assist with diagnosing mental health conditions. Likewise, classification models analyze structured datasets, such as sleep patterns, speech characteristics, and social interaction frequencies, to measure depressive symptoms. Supervised methods interpret text responses from mental health assessments, identifying linguistic markers of anxiety and depression. Diagnostic models that integrate sensor-derived data with self-reported information enhance predictions of mental health deterioration [24]. Probabilistic models estimate relapse risk by monitoring routine behavioral changes. Mobile-health applications collect continuous behavioral data, enabling real-time adjustments in therapeutic interventions. These approaches support early detection by flagging deviations from established patterns, reinforcing preventive care [24].

In another perspective, within ML, DL has gained considerable attention for its mental illness diagnostic applications [12]. This set of methods extracts high-dimensional features from data, advancing data-driven healthcare solutions [13], extending beyond single-disorder prediction to identify comorbidities and interconnected conditions [14]. Complex neural architectures (CNNs, RNNs, transformers, among others) identify intricate relationships within neuroimaging, EEG, and textual data from patient records or social media [14]. Recent applications include automated detection of depression, anxiety, bipolar disorder, and schizophrenia. For instance, CNNs examine neuroimaging to detect structural and functional brain anomalies with high accuracy. RNNs and transformer-based language models (e.g., BERT) parse text to uncover disorder-specific linguistic signals. In this sense, the combination of DL and traditional ML models also shows encouraging accuracy results when trained on small datasets without overfitting, pointing to a productive intersection of both methodologies in similar environments [12].

Within this context, ML and DL have become valuable tools in public health informatics, particularly in studying mental health issues among the LGBTIQ+ population [15]. This community often faces challenges such as stigma and social exclusion, which traditional research methods may not fully capture. ML and DL approaches help identify hidden patterns in data and improve predictive modeling capabilities, with several studies exploring and forecasting mental health distress among the community [15].

These investigations have primarily examined depression, suicide, mood disorders, and minority stress, mostly relying on supervised learning techniques to make predictions in this research area. Models such as conventional logistic regressions, ensemble methods (Random Forest, Decision Trees, Gradient Boosting and its variants, among others) and support-vector machines appear to be the most commonly applied classification techniques to analyze demographic details, behavioral patterns, and historical mental health records to generate accurate risk estimates. This approach has proven effective in identifying individuals at high risk of developing mental health conditions, enabling timely intervention. ML and, particularly, DL models have also been employed to examine how minority stress affects mental health in the LGBTIQ+ community. Based on discrimination experiences, social support systems, and coping mechanisms, these models estimate the likelihood of mental health challenges resulting from minority stress [15]. Furthermore, integration of these algorithms into mobile-health platforms enables continuous monitoring and early detection, offering a comprehensive view of real-world mental health trajectories.

2.3. Advancements in One-Dimensional Convolutional Neural Networks Architectures for Mental Health Research

Research in mental health modeling has evolved from traditional ML to DL methods, making One-dimensional Convolutional Neural Networks (1D-CNNs) a distinct focus. 1D-CNNs have recently gained substantial traction as a model fitted for analyzing sequential and complex data. Originally designed for tasks like speech recognition, their use has expanded across domains involving temporal, signal-based or static data and makes them particularly suitable for real-world applications where data arises from physiological, behavioral, or acoustic sources. Compared to other DL architectures, 1D-CNNs offer a better balance between expressiveness and computational efficiency, enabling the extraction of local temporal patterns with fewer parameters and faster convergence times [25,26]. Their applicability spans a wide range of domains, but their role in health-related data modeling—especially for mental health—has been particularly promising.

Within this domain, 1D-CNNs have shown increasing value in modeling the dynamics of emotional and cognitive states [27]. Unlike traditional ML pipelines, 1D-CNNs can directly model temporal variations in physiological, acoustic, or behavioral signals that reflect psychological states. Awan et al. [27] proposed a hybrid deep learning architecture that combines 1D-CNNs with Vision Transformers to classify emotional states. Their findings highlight the suitability of 1D-CNNs as effective temporal feature extractors in affective computing tasks, outperforming more conventional classification techniques, underscoring the relevance of 1D-CNNs for modeling internal psychological states that manifest through physiological responses.

Further evidence of the relevance of 1D-CNNs in mental health research can be found in their application to depression detection [28]. 1D-CNNs have been successfully incorporated into systems that analyze speech patterns, social media activity, and biosignals to identify early markers of psychological distress. For instance, a BERT-based approach that integrates 1D-CNNs has been used for depression and suicide identification by extracting temporal features from user-generated content and emotional expressions. Similarly, speech emotion recognition systems utilizing 1D-CNNs have demonstrated strong performance in capturing subtle vocal cues linked to affective disorders.

Complementary insights reinforce the importance of selecting architectures capable of modeling temporal changes in both linguistic features and physiological data [25]. Mental health states—especially those involving acute risk like suicidal ideation—have been encoded in sequences of linguistic or behavioral features that unfold over time. In such cases, convolutional architectures like 1D-CNNs, with their capacity for local pattern detection and hierarchical feature learning, were particularly well suited. Moreover, in the context of depression detection from biosignals, there has been an increasing use of deep convolutional models to analyze EEG and ECG data [25]. Although a range of models including DNNs and LSTMs are discussed, the growing shift toward architectures that can process temporal dependencies efficiently—such as 1D-CNNs—is evident. These models have shown promising results in detecting depressive symptoms from multimodal inputs without requiring extensive feature engineering, making them both practical and scalable for real-world deployment.

Altogether, these developments establish 1D-CNNs as a methodologically justified and empirically validated choice for applications in psychological and emotional health assessment. Their versatility in processing physiological, acoustic, and textual signals aligns directly with the demands of mental health modeling, where capturing dynamic, context-sensitive patterns is essential.

3. Materials and Methods

3.1. Study Data and Design

This study draws on data from the 2017 First Virtual Survey for LGBTIQ+ People in Peru, conducted nationwide—across urban and rural areas in all 24 departments and the Constitutional Province of Callao [8]. Carried out between 17 May and 17 August 2017, the survey sought to generate statistical information on factors affecting LGBTIQ+ individuals, with particular emphasis on mental health and its determinants. A total of 12,026 people aged 18 or older participated, all of whom either identified as LGBTIQ+ or whose gender identity or expression diverged from binary norms; 72% (8630) of respondents were aged 18–29. Given the vulnerability of sexual and gender minorities, Peruvian government agencies implemented rigorous ethical protocols to ensure confidentiality, anonymity, and respect for participants' identities. All data were fully anonymized, with no personally identifiable information collected, and responses were secured via encrypted user codes to safeguard privacy and minimize risk.

This study employed a non-probabilistic convenience sampling approach due to the absence of official demographic data on the size and distribution of Peru's LGBTIQ+ population. Methodologically, this strategy was justified by the exploratory nature of the survey and the lack of a reliable sampling frame; accordingly, the results apply only to participants and should not be generalized to the broader LGBTIQ+ community. The survey was disseminated via a broad array of digital platforms, including social media, institutional websites, and civil society organization networks, to reach diverse geographic and socioeconomic segments of the LGBTIQ+ population [8]. Self-administration likely increased participation among individuals who might otherwise have been reluctant to disclose sensitive personal information.

The survey instrument comprised 76 questions organized into thematic modules: geographic and housing data (3 items), respondent identification (7 items), and the primary research domain (66 items), covering sociodemographics (36 items); experiences of discrimination and violence (11 items); knowledge of LGBTI rights (3 items); civic participation (2 items); perceptions of LGBTIQ+ identities (4 items); and household information (5 items). Regarding feature selection, we selected 12 variables based on theoretical relevance [29], including one primary analytic variable, self-reported mental health problems, and 11 auxil-

iary explanatory variables spanning demographic, economic, health, behavioral, and social factors. The dataset also includes a unique identifier for each respondent, used solely for internal tracking and omitted from all statistical analyses. The detailed process of feature selection is provided in Section 3.3.1. Moreover, sample preprocessing included simplifying variable structures, eliminating overlaps across thematic sections, and excluding incomplete or inconsistent responses. First, we removed 13 participants who were missing the main outcome variable, resulting in a sample of 12,013. Next, we excluded 7064 cases with missing values in any of the study variables. Finally, to focus on the subgroup of gay men, we removed 2786 participants who did not identify as such, yielding a final sample of 2186 individuals with complete data. The thorough process of sample selection is provided in Section 3.3.2.

Variables were encoded as follows: categorical variables underwent label encoding, with ascending integers assigned to each category level, while numerical variables remained in their original form. Missing-data rates varied widely across variables, from 0% up to 97%, averaging 26.3% overall. In many cases, high levels of missingness resulted from the survey's skip patterns, whereby subsequent items were shown only if respondents answered certain preceding questions. Accordingly, these omitted values were treated contextually rather than as standard non-responses.

3.2. Machine and Deep Learning Models

ML methods, including Random Forest (RF), Artificial Neural Networks (ANN), Logistic Regression (LR), eXtreme Gradient Boosting (XGBoost), and Gradient Boosting Machine (GBM), as well as a DL One-Dimensional Convolutional Neural Network (1D-CNN), were used to predict mental health problems among self-identified gay men.

3.2.1. XGBoost

The eXtreme Gradient Boosting (XGBoost) algorithm is a decision tree-based ensemble method that builds models in a stage-wise manner using gradient descent to minimize a loss function [30]. By sequentially combining weak learners, it improves classification performance through error correction. Its scalability, regularization mechanisms, and ability to model complex feature interactions make it effective for structured and high-dimensional data, especially in settings prone to overfitting or noise [30]. Shrinkage and column subsampling further enhance its generalization capacity, balancing precision and efficiency in heterogeneous or imbalanced inputs.

3.2.2. Random Forest

Random Forest is an ensemble classifier made up of multiple decision trees trained on bootstrap samples and random feature subsets [31]. Final predictions are based on majority voting across trees. It is robust to overfitting, handles multicollinearity well, and performs reliably in high-dimensional spaces. Its capacity to model nonlinear relationships and assess feature importance is useful in datasets with heterogeneous variables or minimal preprocessing [31]. The combination of data and feature randomness stabilizes predictions, especially in the presence of noise or missing values.

3.2.3. Gradient Boosting Machines

Gradient Boosting Machines (GBM) iteratively combine weak learners, typically shallow trees, using an additive model optimized via gradient descent [32]. Each model fits the residuals of the previous one to gradually reduce loss. This allows fine-tuned adjustments to hard-to-classify cases and is especially useful in imbalanced datasets. GBM's flexibility and accuracy make it effective for capturing complex patterns in tabular data. Regularization parameters such as learning rate and tree depth help control model com-

plexity, while iterative updates refine local decision boundaries [32]. Despite its complexity, the method remains computationally tractable for moderately sized datasets due to shallow tree structures and parallel training capabilities.

3.2.4. Artificial Neural Networks

Artificial Neural Networks (ANNs) consist of interconnected layers of processing units that transform input data via weighted links and nonlinear activation functions [33]. This architecture supports modeling of complex, nonlinear relationships, especially when linear methods are insufficient. ANNs adapt well to high-dimensional, multi-variable data and learn latent representations useful in abstract pattern detection [33]. By tuning topology and training parameters, ANNs can capture both subtle trends and large-scale variation. While less interpretable than linear models, they offer scalability and efficient learning via batch-based optimization and GPU acceleration.

3.2.5. Logistic Regression

Logistic Regression is a statistical technique used for classifying binary or categorical data. It relies on the sigmoid function, which converts a linear combination of predictor variables into a probability [34]. It is valued for its interpretability and efficiency in estimating associations between variables and outcomes. Often used as a baseline for comparison with other techniques, it offers consistent performance and clear benchmarks for more complex models. Its simplicity makes it practical when transparency is important, and its probabilistic outputs allow for flexible decision thresholds and calibration [34].

3.2.6. One-Dimensional Convolutional Neural Network (1D-CNN)

One-dimensional Convolutional Neural Networks (1D-CNNs) are a class of deep learning models designed to process ordered, one-dimensional data using a sequence of convolutional, pooling, dropout, and dense layers [35]. Their structure enables automatic extraction of relevant patterns by applying filters that scan across the input to generate feature maps, followed by nonlinear transformations such as the ReLU activation function to improve learning efficiency and gradient flow. Their convolutional filters are designed to capture local dependencies by sliding across input sequences and detecting patterns that might not be evident through manual feature engineering. This makes them effective in identifying hierarchical representations directly from the raw data.

The architecture offers a favorable trade-off between representational power and computational cost, particularly in contexts where input features exhibit sequential, spatial, or structured relationships [35]. Pooling layers contribute to translational invariance and dimensionality reduction, while dropout introduces robustness through regularization. 1D-CNNs are also applicable to structured datasets when the ordering of features or proximity conveys meaningful correlations, justifying their inclusion in classification tasks involving dense, fixed-length inputs [35]. Their parameter sharing and local connectivity significantly reduce the number of computations compared to fully connected architectures, improving training efficiency on structured datasets.

3.3. Methodology

Figure 1 illustrates the methodology employed to achieve the study objectives.

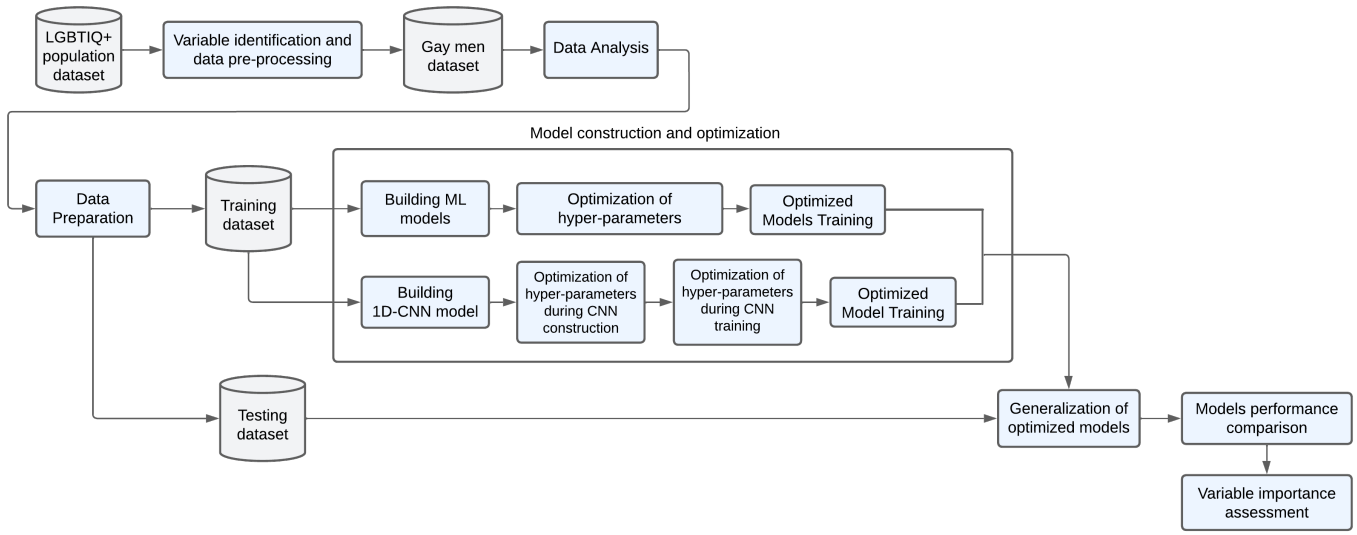


Figure 1. Methodology for mental health problems prediction.

3.3.1. Variable Identification and Selection

This study utilized self-reported data from a nationwide survey designed to investigate factors associated with mental health outcomes among Peru's LGBTIQ+ population. The original instrument comprised 76 items organized into thematic modules—sociodemographic characteristics; experiences of discrimination and violence; perceptions of LGBTIQ+ identities; civic participation; legal awareness; and household context. To prepare for predictive modeling, we applied a multi-step variable-selection procedure to isolate a focused set of predictors, as illustrated in Figure 2.

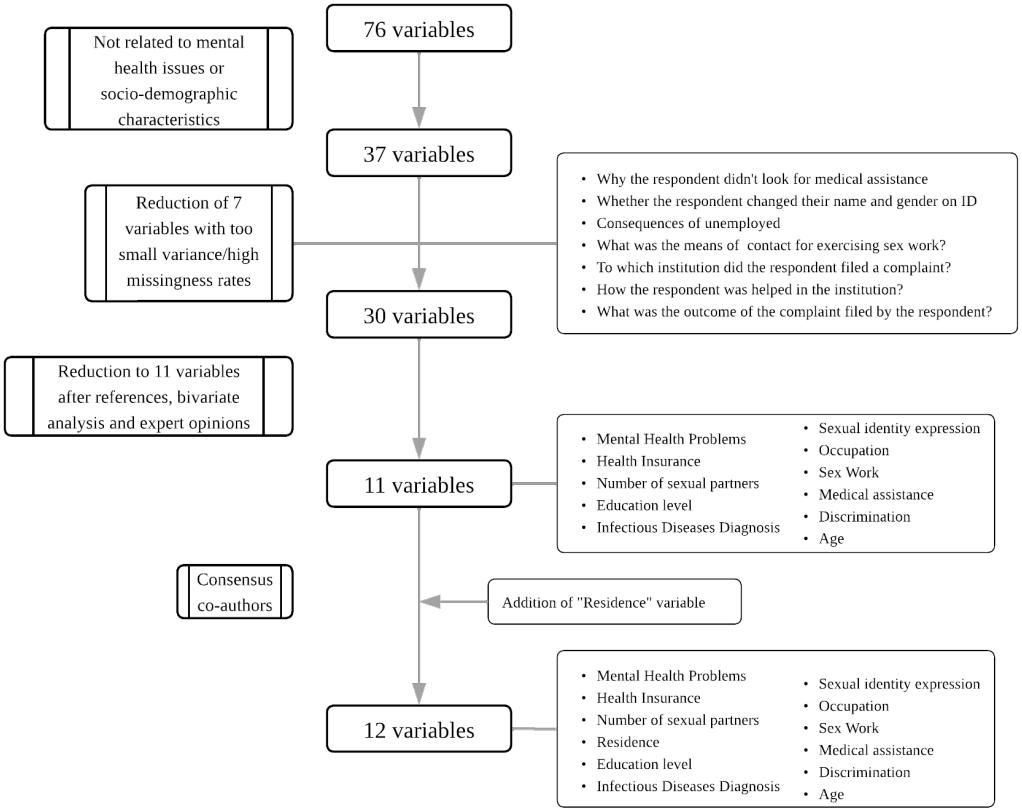


Figure 2. Flow diagram of variable selection within the study.

In the first step, survey items unrelated to mental health or sociodemographic characteristics were excluded on theoretical grounds, reducing the initial pool from 76 to 37 items. In the second step, we removed variables with very low variance or high rates of missing data, yielding a subset of 30 items. In the third step, the remaining variables were evaluated via bivariate analyses, consultation of the literature [29], and expert input to identify those with substantive associations with the dependent variable, current mental health status, resulting in 11 variables. Finally, by consensus of the co-authors, one additional variable was added, bringing the total to 12 variables that met the established criteria. Appendix A lists all selected variables with their operational definitions, as summarized in Table A1.

To assess participants' mental health, they were asked: "Do you have any physical or mental health conditions or illnesses lasting or expected to last 12 months or more that affect you and interfere with your normal activities?". One of the response options was "mental health condition". Based on this, the dependent variable, mental health problems, was measured dichotomously. Individuals who reported experiencing mental health issues were coded as "yes" (1), while those who did not were coded as "no" (0).

The explanatory variables included a range of demographic, economic/occupational, health-related, behavioral, and social factors.

Demographic variables covered age (categorized into 18–24, 25–34, 35–44, and 45+), highest level of education attained, and place of residence, distinguishing between individuals living in Lima and those in other regions. Economic and occupational variables captured whether respondents had engaged in paid work in the past 12 months and whether they had ever participated in sex work. Health-related predictors included whether the respondent had health insurance, had sought medical care or used medical services in the past 12 months, and whether they had been diagnosed with, or currently had, an infectious disease (such as tuberculosis, HIV, or other sexually transmitted infections) within the past 12 months. Behavioral and social factors included the number of sexual partners in the past 12 months, whether individuals openly expressed their sexual identity, and whether they had experienced discrimination based on their sexual orientation or gender identity.

3.3.2. Study Sample Preprocessing

As the Peruvian LGBTIQ+ survey was organized into thematic sections addressing structural determinants, we carefully selected and organized the relevant study variables, outlined previously in Section 3.3.1, to ensure analytical consistency. Following that procedure, we conducted a data-cleaning and filtering process, as depicted in Figure 3. We simplified the dataset by removing redundancies and excluding cases with missing or inconsistent data for key variables, retaining only information relevant to our analysis and aligned with our inclusion criteria.

First, records lacking complete information on the primary outcome, self-reported current mental health issues, were removed, reducing the sample from 12,026 to 12,013 participants. Next, to preserve analytic robustness and mitigate bias from partial responses, we excluded 7064 participants missing data on any explanatory variable, yielding a final sample of 4949 participants with complete data across all selected study variables. Likewise, to align the sample with the study's target population, a final filtering step was applied: we retained only participants who self-identified as gay men, excluding 2763 individuals whose sexual orientation or gender identity did not meet this criterion. Thus, the resulting final study sample comprised 2186 self-identified gay men.

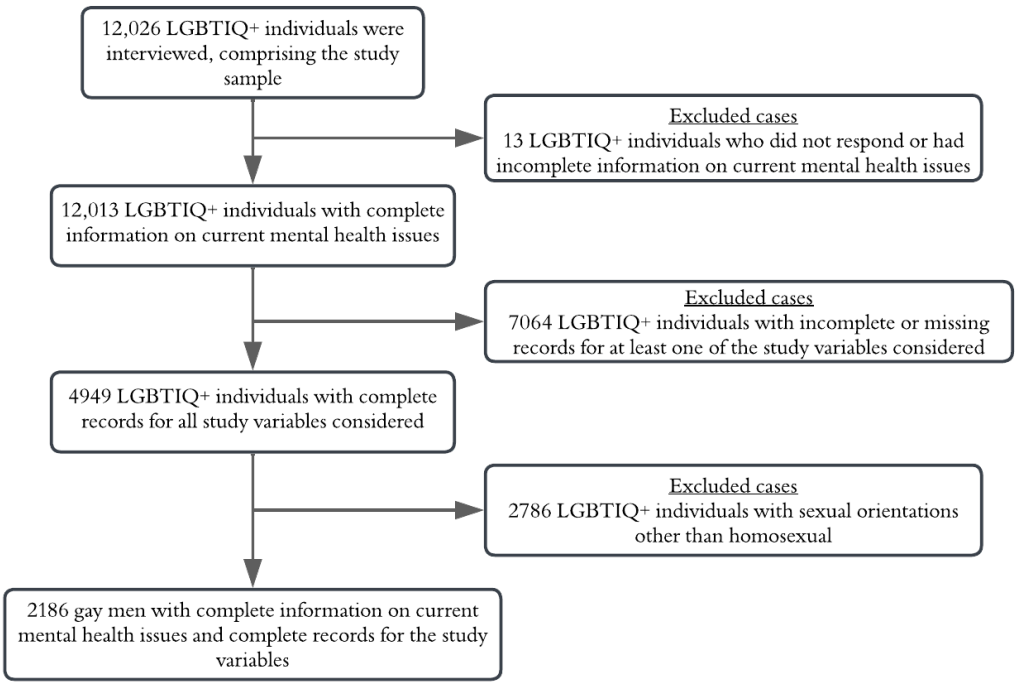


Figure 3. Flow diagram of sample selection within the study.

This filtering process was essential not only for maintaining statistical rigor but also for preventing misclassifications and controlling for potential confounders in our predictive models. Our inclusion and exclusion criteria were applied systematically, guided by theoretical considerations and previous empirical research [36]. Although complete-case analysis reduces sample size, it was necessary to preserve the internal validity of our models and to minimize bias arising from missing-data mechanisms.

3.3.3. Data Analysis

We conducted univariate and bivariate analyses to explore each features' relationship with mental health problems. Because all variables were categorical, we summarized univariate distributions using counts and proportions. For bivariate analyses, we applied two-tailed chi-square tests to assess each predictor's association with the dependent variable, treating $p < 0.01$ as statistically significant. To detect potential redundancies and assess predictor independence, we examined interactions and multicollinearity via Generalized Variance Inflation Factors (GVIFs) and computed a Cramer's V–based correlation matrix.

3.3.4. Data Preparation

To ensure robust model development and generalization, the dataset of 2186 observations was randomly split into training and testing subsets using an 80–20% ratio. Details of this split—designed to optimize both model training and evaluation while preserving sample representativeness—are provided in Appendix A and summarized in Table A2.

3.3.5. Models Construction and Optimization

In the initial phase, we constructed each ML model's architecture and set its core parameters. Once the models were built, we optimized their performance via hyperparameter tuning. Table 1 presents the hyperparameters evaluated for each model along with their tested values. Logistic Regression required no hyperparameter tuning, as it depends solely on a link function and has no additional parameters that affect predictive performance.

Table 1. Definition of hyperparameters for ML models.

Algorithm	Hyperparameter	Definition
LR	-	-
ANN	max_iter	Defines the maximum number of iterations the model will run before stopping
	hidden_layer_sizes	Specifies the number of neurons in the hidden layer of the network
	momentum	Controls the momentum factor used in the gradient descent update
	learning rate	Determines the step size for updating the model's weights during training
RF	n_estimators	Specifies the number of decision trees in the forest
	max_features	Defines the number of features considered for finding the best split
	max_depth	Sets the maximum depth allowed for each tree in the forest
	criterion	Determines the function used to evaluate the quality of a split
XGBoost	max_depth	Defines the maximum depth of individual estimators
	n_estimators	Specifies the number of boosting iterations to be performed
	learning_rate	Scales the contribution of each tree to the overall model
	min_split_loss	Regularization parameter that sets the minimum loss reduction required to allow a split
GBM	learning_rate	Controls the contribution of each tree to the final prediction
	max_depth	Sets the maximum depth of each tree
	subsample	Defines the fraction of training data used per boosting iteration
	n_estimators	Specifies the number of boosting iterations (trees)

Model training and validation were performed using a grid search algorithm combined with 10-fold cross-validation on the training subset. This approach involved iteratively fitting models with various hyperparameter combinations to identify the most effective configuration for each algorithm, thereby ensuring optimal performance.

Conversely, the DL approach involved building and training a 1D-CNN tailored to extract relevant features from the dataset. Model development was structured in two phases [37]: first, tuning network parameters associated with the model construction such as the number of filters, kernel size, and hidden layer configuration; and second, adjusting the training process parameters, including the number of epochs and batch size. This systematic optimization aimed to enhance feature extraction and generalization [37].

A baseline architecture was initially implemented, comprising convolutional, pooling, and dense layers, as illustrated in Figure 4, and subsequently refined through iterative tuning to boost classification performance.

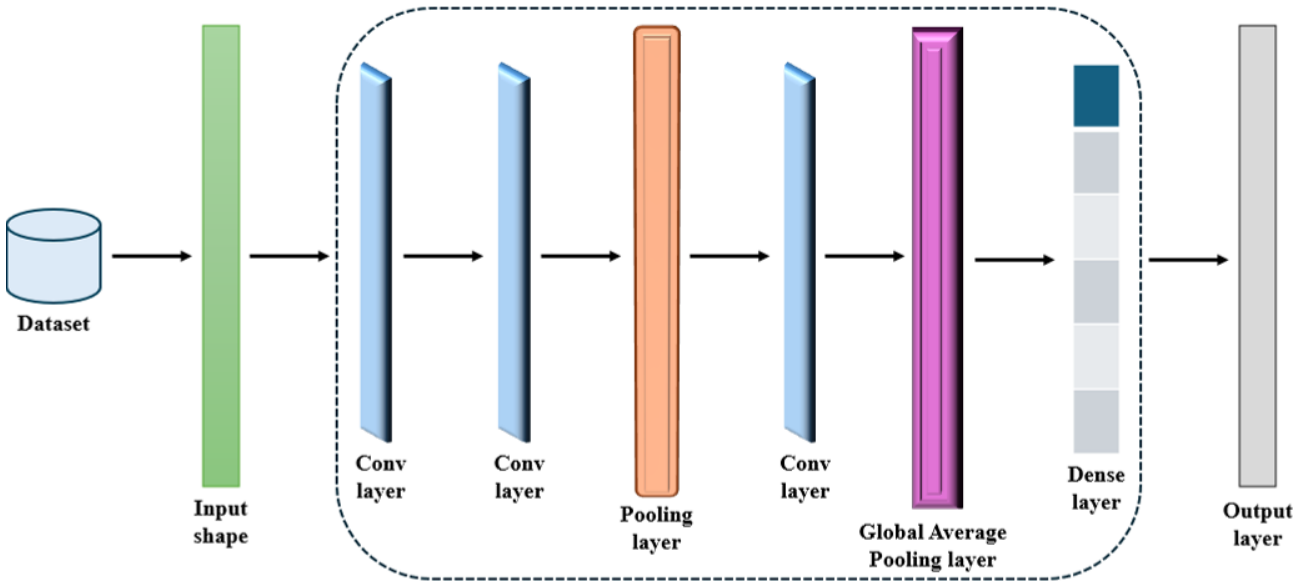


Figure 4. Proposed base architecture for the 1D-CNN construction and optimization. The dashed-line frame represents the internal layer configuration of the model.

Moreover, Table 2 presents the initial parameters used in the baseline model.

Table 2. 1D-CNN base parameters.

Parameter	Value
Optimizer	Adam
Batch size	32
Kernel size	2
Epochs	200
Loss	Binary Cross-Entropy
Convolutional and pooling layers activation function	ReLu
Fully connected layers activation function	Sigmoid
Learning rate	0.001

The construction phase involved systematically varying key hyperparameters using 10-fold cross-validation on the baseline model. First, the number of filters in the convolutional and dense layers was adjusted, testing values of 10, 25, 50, and 100, to balance feature extraction and model complexity. Subsequently, kernel sizes of 2, 3, 5, 7, and 11 were evaluated to optimize the trade-off between local pattern detection and computational efficiency. Finally, the overall architecture was improved by including batch normalization, dropout, and flatten layers, as well as reconfigurations of convolutional and dense layers aimed at improving generalization and classification accuracy. The final architecture resulted from iterative tuning of convolutional structures to balance accuracy and stability.

Subsequently, once the CNN architecture and structural parameters were fully optimized, training hyperparameters were tuned to improve convergence and efficiency. Different numbers of epochs, such as 50, 100, 150, 200 and 250, were tested to avoid underfitting and overfitting, while various batch sizes, including 32, 64, 96, 128 and 160, were evaluated to maximize learning efficiency and model robustness. Smaller batch sizes tended to improve stability and generalization, whereas larger ones had a noticeable impact on convergence dynamics. This process yielded a CNN model optimized in both architecture and training configuration for reliable performance evaluation.

3.3.6. Model Performance Comparison

Model performance was evaluated using metrics derived from the confusion matrix, which quantifies correct and incorrect classifications, as shown in Table 3.

Table 3. Confusion matrix for models' evaluation.

	Predicted Class	
True Class	**Negative**	**Positive**
Negative	True Negative (*TN*)	False Positive (*FP*)
Positive	False Negative (*FN*)	True Positive (*TP*)

We used standard classification metrics, namely, accuracy, precision, recall, specificity, F1-score, and AUROC, for model comparison. Their formulas, based on the elements of the confusion matri, are presented below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{5}$$

3.3.7. Feature Importance

To ensure consistent interpretation of feature importance in the best-performing model, we used SHAP (Shapley Additive exPlanations), which decomposes predictions into additive contributions from each input feature, quantified as Shapley values [38,39]. This enhances clarity on feature influence while enabling localized and model-agnostic comparisons [38,39].

To further improve explainability, particularly for deep learning models, SHAP was complemented with Layer-wise Relevance Propagation (LRP), a neural network-specific method that produces relevance heatmaps by tracing feature contributions across layers [40]. LRP supports both local and global interpretability by offering input features most critical to predicting True Positives and True Negatives, enhancing transparency in model decision-making, to capture feature relevance in a way that is structurally aligned with the internal mechanics of the neural network [40].

3.3.8. Statistical Analysis

Univariate and bivariate analyses were conducted in R (v4.4.3). Model development combined a 1D-CNN implemented with Keras and TensorFlow in Python (v3.13.2), and ML models were built using scikit-learn and XGBoost. Performance was compared using scikit-learn metrics, while SHAP and iNNvestigate libraries were used to assess feature importance.

4. Results

4.1. Univariate and Bivariate Analysis of Study Sample

Considering sociodemographic, economic, and health factors, Table 4 presents the univariate analysis of the explanatory variables among respondents with and without mental health problems, along with the results of the chi-square (χ^2) tests and GVIF values.

Table 4. Descriptive statistics of the study variables.

	Mental Health Problems			
	Yes (*n* = 847)	**No (*n* = 1339)**	**χ^2 Test**	**GVIF**
Characteristics	***n*/%**	***n*/%**	***p*-Value**	**Value**
Health Insurance				
Yes	576 (68.0%)	1036 (77.4%)	<0.01 ***	1.03
No	271 (32.0%)	303 (22.6%)		
Number of sexual partners				
No sexual partners	539 (63.6%)	748 (55.9%)	<0.01 ***	1.02
One sexual partner	297 (35.1%)	573 (42.8%)		
More than one sexual partner	11 (1.3%)	18 (1.3%)		
Residence				
Lima	573 (67.7%)	956 (71.4%)	<0.01 ***	1.02
Inland regions	274 (32.3%)	383 (28.6%)		

Table 4. *Cont.*

	Mental Health Problems			
	Yes (n = 847)	**No (n = 1339)**	**χ^2 Test**	**GVIF**
Characteristics	***n*/%**	***n*/%**	***p*-Value**	**Value**
Education level				
No education to incomplete elementary school	1 (0.1%)	4 (0.3%)		
Complete elementary school to incomplete high school	10 (1.2%)	9 (0.7%)	<0.01 ***	1.05
Complete high school to incomplete superior education	444 (52.4%)	567 (42.3%)		
Complete superior education to postgraduate education	392 (46.3%)	759 (56.7%)		
Infectious diseases diagnosis				
Yes	155 (18.3%)	712 (53.2%)	<0.01 ***	1.02
No	692 (81.7%)	627 (46.8%)		
Sexual identity expression				
Yes	335 (39.6%)	625 (46.7%)	<0.01 ***	1.01
No	512 (60.4%)	714 (53.3%)		
Occupation				
Yes	500 (59.0%)	911 (68.0%)	<0.01 ***	1.09
No	347 (41.0%)	428 (32.0%)		
Sex Work				
Yes	106 (12.5%)	195 (14.6%)	<0.01 ***	1.03
No	741 (87.5%)	1144 (85.4%)		
Medical assistance				
Yes	627 (74.0%)	1297 (96.9%)	<0.01 ***	1.02
No	220 (26.0%)	42 (3.1%)		
Discrimination				
Yes	720 (85.0%)	1039 (77.6%)	<0.01 ***	1.01
No	127 (15.0%)	300 (22.4%)		
Age				
18–24 years	399 (47.1%)	407 (30.4%)		
25–34 years	326 (38.5%)	610 (45.6%)	<0.01 ***	1.05
35–44 years	82 (9.7%)	218 (16.3%)		
More than 45 years	40 (4.7%)	104 (7.8%)		

*** highly significant values $p < 0.01$, GVIF: Generalized Variance Inflation Factors.

Health insurance coverage was lower among individuals with mental health problems (68.0%) compared to those without (77.4%). Similarly, higher education was more common in the group without mental health problems (56.7%) than in the group with mental health problems (46.3%). Health disparities were evident: a higher prevalence of infectious disease diagnoses occurred among individuals without mental health problems (53.2%), while medical assistance was more frequently reported in this group (96.9%) compared to those with mental health problems (74.0%). Experiences of discrimination were more common among individuals with mental health problems (85.0%) than among those without (77.6%). Employment rates were lower among individuals with mental health problems (59.0%), whereas those without such problems had higher employment levels (68.0%). Younger participants, especially those aged 18–24 years, reported mental health problems more frequently (47.1%) than older age groups.

All variables showed *p*-values below 0.01, indicating statistically significant associations with the dependent variable and confirming that each independent variable was meaningfully related to the outcome. Additionally, multicollinearity was assessed using Generalized Variance Inflation Factors (GVIFs), with all values close to 1, indicating low shared variance among predictors and supporting the interpretability and reliability.

To further assess potential dependencies, a correlation matrix using Cramer's V was computed, as shown in Figure 5. The strongest associations were observed between Age

and Occupation (0.32) and between Occupation and Education level (0.30). All other pairwise correlations were lower, indicating limited redundancy. Together with the low GVIF values, these results suggest acceptable independence among the predictors.

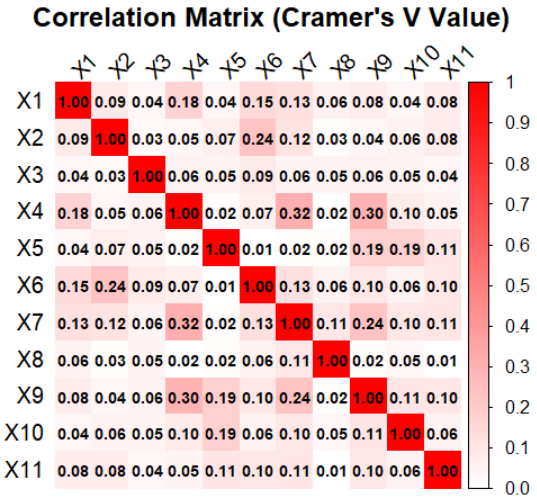


Figure 5. Correlation matrix based on Cramer's V values of independent variables. Labels represent the following factors: X1 = Health Insurance, X2 = Infectious Disease Diagnosis, X3 = Sexual Identity Expression, X4 = Occupation, X5 = Sex Work, X6 = Medical Assistance, X7 = Age, X8 = Discrimination, X9 = Education Level, X10 = Number of Sexual Partners, X11 = Residence.

4.2. Development and Optimization of Models

4.2.1. Optimization of ML Models

Table 5 summarizes the hyperparameter options tested for each ML model, the selected optimal values, and the best validation accuracies obtained after tuning and cross-validation, following the process described in Section 3.3.5. Subsequently, these models with the specified parameters were applied to the testing subset to identify the best approach for classifying mental health problems.

Table 5. Hyperparameter tuning and selection of optimal values for the ML models.

Model	Hyper-Parameter	Tested Values	Optimal Values	Best Validation Accuracy
LR	-	-	-	70.77%
ANN	Number of neurons	2–26	**16**	71.22%
	Max iterations	250–500–750–1000	**500**	
	Learning rate	0.1–0.5–0.01	**0.01**	
	Momentum	0.4–0.9	**0.7**	
RF	Number of estimators	100–200	**160**	72.14%
	Max features	sqrt-log2	**log2**	
	Max depth	4–12	**8**	
	Criterion	Gini–Entropy	**Entropy**	
GBM	Learning rate	0.01–0.025–0.05–0.075–0.1	**0.1**	72.77%
	Max depth	2–10	**4**	
	Subsample	0.5–1.0	**0.9**	
	Number of estimators	10–100	**10**	
XGBoost	Max depth	2–10	**6**	72.48%
	Number of estimators	20–220	**100**	
	Learning rate	0.1–0.01–0.05–0.075	**0.01**	
	Min split loss	0–0.2–0.4–0.6–0.8–1.0	**1.0**	

The bold values in the table body represent the optimal values selected after the cross-validation and grid search process for each model.

4.2.2. Optimization During CNN Construction

Initially, a simple 1D-CNN base architecture was proposed, consisting of two convolutional layers, a max-pooling layer, another convolutional layer, a global average pooling layer, and a dense layer, as depicted in Figure 4, to evaluate performance across hyperparameter settings. The first hyperparameter optimized was the number of filters per layer. Table 6 lists the tested configurations for convolutional and dense layers to assess their impact on network performance.

Table 6. Proposed models for assessment of filter values variations.

Model	Filter Values
Model 1	Filter: 75-50-25 (Conv1D) & 25-1 (Dense)
Model 2	Filter: 50-25-10 (Conv1D) & 10-1 (Dense)
Model 3	Filter: 30-15-5 (Conv1D) & 10-1 (Dense)
Model 4	Filter: 20-10-5 (Conv1D) & 5-1 (Dense)
Model 5	Filter: 100-50-25 (Conv1D) & 20-1 (Dense)
Model 6	Filter: 120-100-120 (Conv1D) & 40-1 (Dense)
Model 7	Filter: 150-100-50 (Conv1D) & 25-1 (Dense)
Model 8	Filter: 20-10-5 (Conv1D) & 10-1 (Dense)

Table 7 presents the evaluation results of the proposed models across different filter value variations.

Table 7. 1D-CNN evaluation accuracy for each proposed filter variant.

Iteration N°	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
1	72.60%	73.29%	70.09%	71.23%	71.92%	70.78%	71.23%	71.00%
2	71.23%	73.29%	68.49%	71.23%	71.23%	70.09%	70.55%	70.09%
3	71.92%	73.06%	70.55%	70.78%	71.46%	69.41%	69.86%	70.55%
4	71.92%	74.20%	71.00%	71.00%	71.23%	70.09%	69.86%	69.41%
5	72.60%	75.11%	71.23%	71.23%	70.55%	69.86%	70.78%	72.15%
6	71.92%	73.14%	71.92%	71.00%	71.00%	69.41%	70.09%	71.00%
7	72.60%	74.05%	72.60%	70.55%	71.46%	71.00%	71.92%	71.92%
8	72.29%	73.06%	70.09%	71.92%	71.23%	71.23%	72.15%	71.70%
9	73.00%	73.00%	70.55%	70.55%	70.78%	70.78%	72.29%	70.78%
10	71.23%	72.60%	71.92%	72.60%	71.92%	70.78%	70.55%	73.06%
Avg. acc.	72.13%	**73.48%**	70.84%	71.21%	71.28%	70.34%	70.93%	71.17%
Std. dev.	±0.0060	**±0.0075**	±0.0117	±0.0063	±0.0044	±0.0066	±0.0093	±0.0107

Avg. acc.: Average accuracy, Std. dev.: Standard Deviation. Bold values indicate the best filter values based on the performance (highest accuracy and lowest std. dev.) achieved by Model 2.

Model 2 achieved the highest average accuracy (73.48%), balancing effective feature extraction and stability. In contrast, Model 6 showed lower performance (70.34%), possibly due to noise or overfitting, while Model 3's smaller filter setup resulted in reduced accuracy (70.84%), indicating insufficient pattern capture. Based on these results, we selected a 1D-CNN architecture balancing accuracy and stability: three convolutional layers with 50, 25, and 10 filters, followed by a dense layer with 10 neurons and an output neuron.

Next, kernel size was optimized to assess its effect on feature extraction. Table 8 shows the tested kernel sizes and their corresponding accuracy values with standard deviations.

Table 8. Evaluation accuracy of 1D-CNN with different kernel sizes.

Iteration N°	Kernel = 2	Kernel = 3	Kernel = 5	Kernel = 7	Kernel = 11
1	71.92%	73.06%	70.78%	71.00%	69.41%
2	74.43%	71.92%	73.29%	71.46%	70.55%
3	72.60%	74.89%	73.29%	71.92%	70.32%
4	73.52%	72.60%	71.92%	69.18%	70.09%
5	72.83%	74.20%	72.15%	70.09%	69.18%
6	72.60%	74.43%	70.55%	71.00%	71.00%
7	73.14%	73.14%	72.60%	71.46%	70.09%
8	73.06%	73.29%	73.14%	69.41%	70.75%
9	73.52%	72.60%	72.60%	71.92%	69.18%
10	73.00%	73.29%	72.52%	70.09%	69.00%
Avg. acc.	73.06%	**73.34%**	72.28%	70.75%	69.96%
Std. dev.	±0.0067	**±0.0092**	±0.0097	±0.0100	±0.0072

Avg. acc.: Average accuracy, Std. dev.: Standard Deviation. The values in bold indicate the optimal kernel configuration based on the highest average accuracy and lowest standard deviation observed across the evaluated iterations.

A kernel size of 3 yielded the highest average accuracy (73.34%), balancing feature extraction and model stability. In contrast, a kernel size of 11 showed the lowest accuracy (69.96%), likely due to poor capture of local dependencies. Thus, kernel size 3 was selected as optimal for accuracy and consistent performance.

Finally, the 1D-CNN structure influences performance. Table 9 summarizes the proposed model variations with their layer compositions and hyperparameters.

Table 9. Operational definition of the proposed architectures and hyperparameter values.

Architecture	Description	Hyper-Parameters Values
Architecture 1 (A1)	2 successive convolutional layers followed by a max-pooling layer, a convolutional layer, a global average pooling layer and a dense layer	Filters: 50-25-10 (conv), 10 (dense). Kernel: 3.
Architecture 2 (A2)	2 additional consecutive convolutional layers and a max-pooling layer were added to A1	Filters: 50-25-50-25-10 (conv), 10 (dense). Kernel: 3.
Architecture 3 (A3)	2 additional consecutive dense layers were added to A1	Filters: 50-25-10 (conv), 20-15-10 (dense). Kernel: 3.
Architecture 4 (A4)	2 additional consecutive convolutional layers, a max-pooling layer and 2 additional consecutive dense layers were added to A1	Filters: 50-25-50-25-10 (conv), 20-15-10 (dense). Kernel: 3.
Architecture 5 (A5)	A dropout layer was added to A1 after the dense layer	Filters: 50-25-10 (conv), 10 (dense). Kernel: 3. Dropout value: 0.1
Architecture 6 (A6)	The dropout value was increased to 0.2 for the A5	Filters: 50-25-10 (conv), 10 (dense). Kernel: 3. Dropout value: 0.2
Architecture 7 (A7)	Batch normalization layers were added after each convolutional and dense layers for the A6	Filters: 50-25-10 (conv), 10 (dense). Kernel: 3. Dropout value: 0.2
Architecture 8 (A8)	A flatten layer was added after the global average pooling layer for the A6	Filters: 50-25-10 (conv), 10 (dense). Kernel: 3. Dropout value: 0.2
Architecture 9 (A9)	A flatten layer was added after the global average pooling layer for the A7	Filters: 50-25-10 (conv), 10 (dense). Kernel: 3. Dropout value: 0.2

Table 10 shows the evaluation results for the proposed model structure variations.

Table 10. 1D-CNN evaluation accuracy for each proposed architecture.

Iteration	A1	A2	A3	A4	A5	A6	A7	A8	A9
1	73.29%	69.86%	72.60%	72.37%	72.37%	74.66%	73.97%	72.15%	72.37%
2	73.29%	73.74%	71.92%	72.15%	72.60%	73.97%	74.66%	72.15%	73.97%
3	73.06%	70.09%	72.15%	69.86%	72.60%	73.97%	73.97%	72.60%	73.06%
4	74.20%	68.72%	71.69%	70.32%	72.60%	73.52%	74.66%	73.52%	72.15%
5	75.11%	72.38%	72.83%	71.46%	76.03%	74.20%	73.97%	72.83%	73.52%
6	73.14%	69.86%	72.60%	72.60%	72.37%	74.20%	73.52%	72.37%	73.06%
7	74.05%	72.15%	71.92%	71.46%	72.15%	73.52%	73.97%	73.06%	72.15%
8	73.06%	73.29%	72.83%	70.09%	73.97%	72.97%	74.66%	72.37%	72.37%
9	73.00%	71.92%	71.46%	72.15%	72.60%	74.15%	74.20%	73.52%	73.74%
10	72.60%	73.74%	72.15%	71.92%	71.92%	74.66%	74.66%	73.66%	73.97%
Avg. acc.	73.48%	71.58%	72.22%	71.44%	72.92%	73.98%	**74.22%**	72.82%	73.04%
Std. dev.	±0.0075	±0.0182	±0.0048	±0.0100	±0.0122	±0.0053	**±0.0041**	±0.0059	±0.0074

Avg. acc.: Average accuracy, Std. dev.: Standard Deviation. Values in bold indicate the best-performing architecture based on the evaluated iterations, considering both accuracy and standard deviation.

Performance variations across architectures indicate that structural choices impact learning capacity and generalization. Architecture A7 achieved the highest accuracy (74.22%), highlighting the benefit of batch normalization for performance and stability. In contrast, A4, with more convolutional and dense layers, had lower accuracy (71.44%), suggesting that added complexity may harm consistency. Thus, Architecture A7, combining batch normalization with convolutional and dense layers, was selected as optimal for balancing accuracy and stability.

4.2.3. Optimization During CNN Training

Building on the optimized model, defined by the best filter numbers, kernel size, and layer configuration, this section presents results on training-related hyperparameter tuning.

Batch size, which strongly affects training dynamics, was tested at various values. Table 11 shows the corresponding accuracies and standard deviations.

Table 11. Evaluation accuracy output of 1D-CNN with different sizes of batches.

Iteration N°	Batch = 32	Batch = 64	Batch = 96	Batch = 128	Batch = 160
1	72.83%	71.69%	73.52%	72.83%	72.60%
2	73.74%	72.15%	72.83%	70.32%	72.60%
3	73.29%	72.60%	73.06%	73.52%	71.23%
4	73.97%	73.74%	72.37%	72.83%	71.92%
5	73.06%	73.74%	73.06%	70.78%	73.52%
6	73.52%	71.23%	72.83%	71.23%	71.92%
7	73.74%	73.29%	73.29%	71.23%	72.23%
8	72.97%	72.37%	73.14%	72.83%	71.69%
9	72.97%	71.92%	72.37%	72.60%	72.83%
10	73.14%	73.06%	73.06%	72.15%	73.06%
Avg. acc.	**73.32%**	72.58%	72.95%	72.03%	72.36%
Std. dev.	**±0.0039**	±0.0086	±0.0037	±0.0107	±0.0069

Avg. acc.: Average accuracy, Std. dev.: Standard Deviation. Values in bold indicate the optimal batch size obtained from the evaluated iterations, considering both accuracy and standard deviation.

A batch size of 32 achieved the highest average accuracy (73.32%), indicating that smaller batches promote stable learning and effective feature capture. Conversely, a batch size of 128 showed the lowest accuracy (72.03%), suggesting larger batches may reduce generalization. Therefore, batch size 32 was selected to balance accuracy and training stability.

Next, the number of training epochs—key for learning and generalization—was evaluated. Table 12 shows tested epoch values with their accuracy and standard deviation.

Table 12. Evaluation accuracy of 1D-CNN with different epochs.

Iteration N°	Epochs = 50	Epochs = 100	Epochs = 150	Epochs = 200	Epochs = 250
1	73.29%	72.60%	74.43%	72.60%	71.92%
2	75.11%	74.89%	72.83%	72.83%	72.60%
3	71.23%	73.06%	72.37%	73.52%	73.52%
4	73.52%	73.74%	73.06%	74.43%	72.83%
5	73.74%	73.97%	73.74%	73.97%	72.37%
6	73.74%	74.14%	73.06%	73.52%	72.14%
7	74.14%	73.52%	72.83%	74.14%	72.83%
8	71.92%	73.06%	74.14%	72.83%	72.60%
9	72.60%	74.89%	72.52%	72.60%	71.92%
10	73.60%	72.83%	73.37%	73.14%	73.60%
Avg. acc.	73.29%	**73.67%**	73.24%	73.36%	72.63%
Std. dev.	± 0.0111	**±0.0081**	±0.0068	±0.0066	±0.0059

Avg. acc.: Average accuracy, Std. dev.: Standard Deviation. Values in bold indicate the optimal number of epochs obtained from the evaluated iterations, considering both accuracy and standard deviation.

The model trained with 100 epochs achieved the highest average accuracy (73.67%), indicating stable learning and effective feature capture. Training for 250 epochs resulted in lower accuracy (72.63%), suggesting overfitting. Thus, 100 epochs was selected to balance accuracy, stability, and computational cost.

The optimized architecture, as depicted in Figure 6 and described in Table 13, includes three convolutional layers with 50, 25, and 10 filters, all with kernel size 3. Batch normalization follows each convolutional layer for stability. A max-pooling layer after the second convolution reduces dimensionality, followed by global average pooling. Dense layers have 10 neurons, batch normalization, and 0.2 dropout to reduce overfitting. Finally, a single sigmoid neuron produces the output.

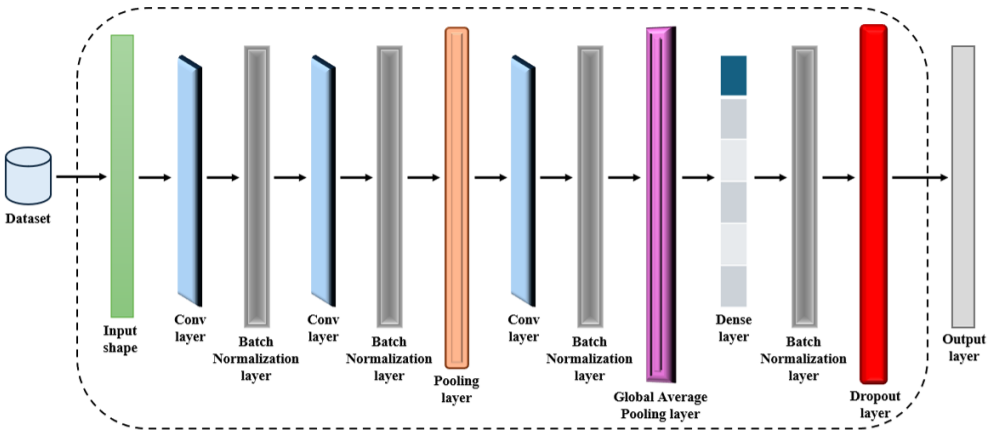


Figure 6. Optimized structure for the 1D-CNN model for predicting mental health problems. The dashed-line frame represents the internal layer configuration of the model.

Table 13. 1D-CNN architecture parameters.

Layer Name	Output Shape	Number of Parameters
Conv1D	(None, 11, 50)	200
BatchNormalization	(None, 11, 50)	200
Conv1D	(None, 11, 25)	3775
BatchNormalization	(None, 11, 25)	100
MaxPooling1D	(None, 5, 25)	0
Conv1D	(None, 5, 10)	760
BatchNormalization	(None, 5, 10)	40
GlobalAveragePooling1D	(None, 10)	0
Dense	(None, 10)	110
BatchNormalization	(None, 10)	40
Dropout	(None, 10)	0
Dense	(None, 1)	11

Figure 7 shows the accuracy and loss curves during the model's 100-epoch training. Training accuracy steadily increased and stabilized around 75%, while loss decreased consistently to around 50%, indicating effective learning with no signs of overfitting.

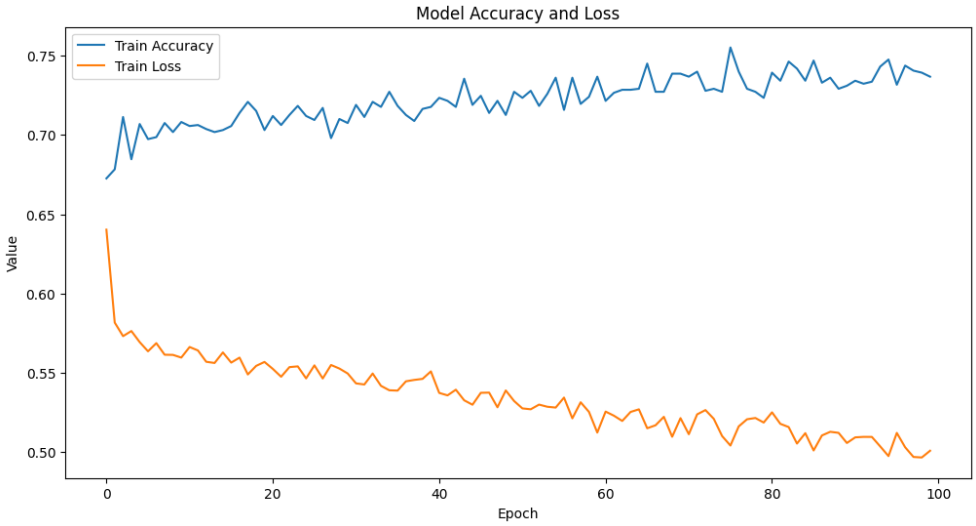


Figure 7. Accuracy and loss plot for the optimized 1D-CNN model.

4.3. Comparison of Classifiers for the Prediction

The results of the training and generalization processes for both ML models and the 1D-CNN are summarized in Tables 14 and 15, respectively. Additional evaluation metrics are provided in Supplementary Material S1 to support the robustness of these findings.

As shown in Table 14, training accuracy ranged from 70.77% to 74.71%, with the 1D-CNN achieving the highest accuracy. The 1D-CNN also demonstrated superior balance between recall (55.07%) and specificity (87.62%). It yielded the highest F1-score (0.63), outperforming GMB (0.58) and LR (0.57), indicating better overall balance between precision and recall. These results suggest that the 1D-CNN model delivers the strongest predictive performance among the evaluated algorithms.

Table 14. Performance comparison of ML and DL models in the training subset.

Metric/Model	1D-CNN	XGBoost	LR	ANN	RF	GBM
Accuracy (%)	**74.71**	72.48	70.77	71.22	72.14	72.77
Precision (%)	74.51	78.95	67.54	**79.93**	75.52	73.48
Recall (%)	**55.07**	41.30	48.91	35.47	42.77	47.74
Specificity (%)	87.62	92.82	84.85	**94.26**	91.06	88.90
F1 Score	**0.63**	0.54	0.57	0.49	0.55	0.58
AUC	**0.84**	0.79	0.77	0.79	0.78	0.83

Values in bold represent the best obtained results according to each goodness-of-fit metric.

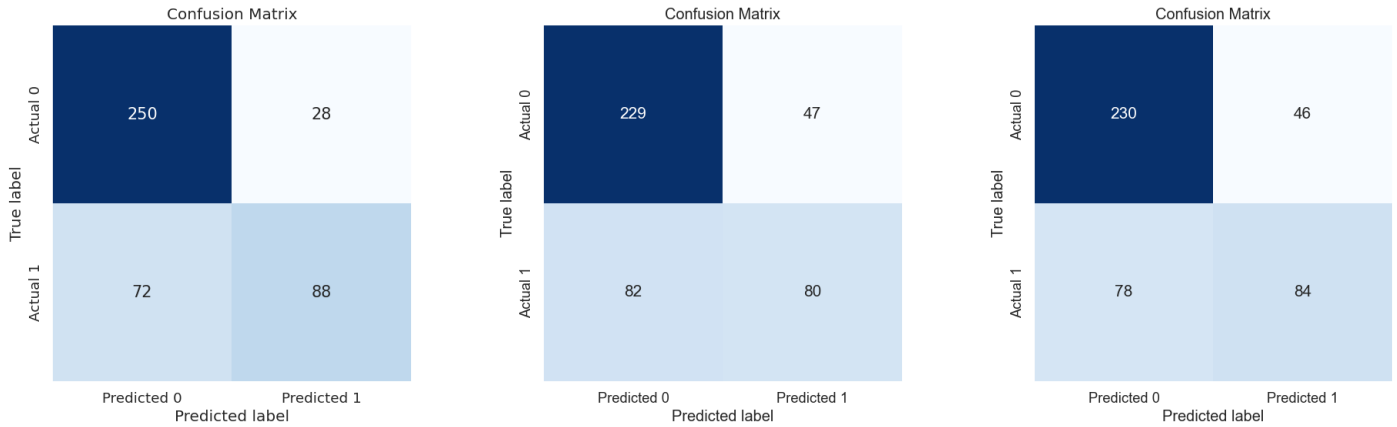
Table 15 shows that testing accuracy ranged from 70.55% to 77.17%, with 1D-CNN achieving the highest accuracy, consistently outperforming all ML models. It also demonstrated the highest specificity (89.93%) and precision (75.86%), indicating strong performance in correctly identifying negative and positive cases, respectively. Although its recall (55.00%) was moderate, 1D-CNN attained the highest F1-score (0.64), reflecting the best balance between precision and recall among the models. These results further support the superior predictive capacity of the 1D-CNN for mental health distress.

Table 15. Performance comparison of ML and DL models in testing subset.

Metric/Model	1D-CNN	XGBoost	LR	ANN	RF	GBM
Accuracy (%)	**77.17**	70.55	71.69	70.78	71.46	72.15
Precision (%)	**75.86**	62.99	64.62	64.41	69.07	68.18
Recall (%)	**55.00**	49.38	51.85	46.91	41.36	46.30
Specificity (%)	**89.93**	82.97	83.33	84.78	89.13	87.32
F1 Score	**0.64**	0.55	0.57	0.54	0.52	0.55
AUC	**0.78**	0.74	0.75	0.75	0.75	0.75

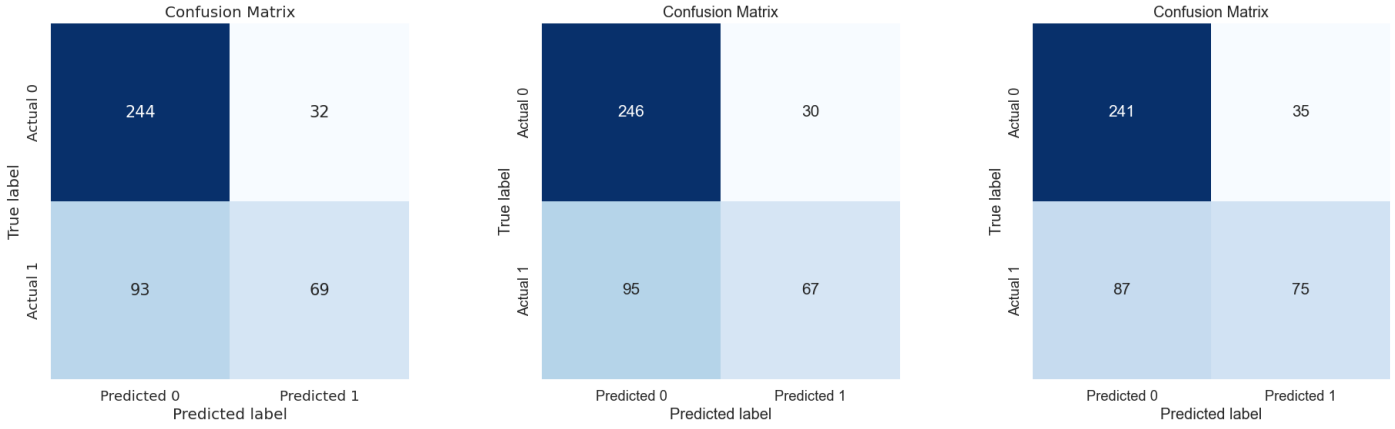
Values in bold represent the best obtained results according to each goodness-of-fit metric.

Moreover, confusion matrices for all models on the testing set are presented in Figure 8.



(**a**) (**b**) (**c**)

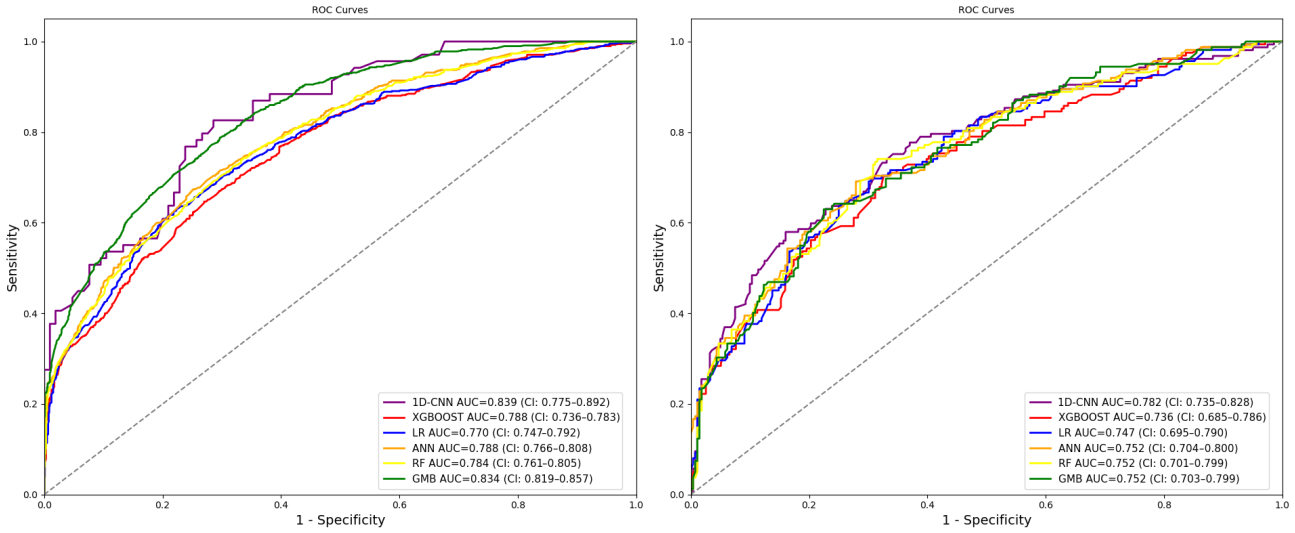
Figure 8. *Cont.*



(**d**) (**e**) (**f**)

Figure 8. Testing confusion matrices: (**a**) 1D-CNN, (**b**) XGBoost, (**c**) LR, (**d**) ANN, (**e**) RF and (**f**) GBM.

Alternatively, Figure 9 presents the AUROC curves for each algorithm, offering a comparative view of their predictive performance.



(**a**) (**b**)

Figure 9. ROC curves of ML and DL models for training subset (**a**) and testing subset (**b**).

In the training set, as illustrated in Figure 9a, the 1D-CNN model achieved the highest AUROC at 0.839 (95% CI: 0.775–0.892), followed closely by GMB at 0.834 (95% CI: 0.819–0.857). XGBoost, ANN, and RF exhibited comparable AUROCs ranging from 0.784 to 0.788. In the testing set, as shown in Figure 9b, 1D-CNN again outperformed the other models with an AUROC of 0.782 (95% CI: 0.735–0.828), though the differences with ANN, RF, and GMB were narrower. These findings highlight 1D-CNN's robust discrimination capability and overall superior predictive performance.

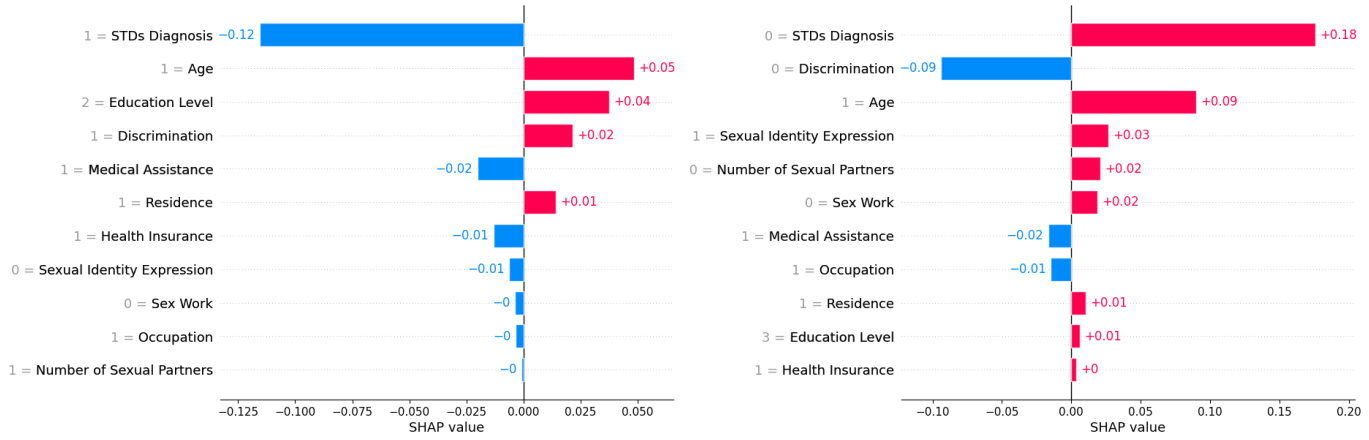
The consistent superiority of the 1D-CNN across multiple evaluation metrics suggests that this architecture is particularly well suited to capturing complex patterns in the data that may underlie mental health distress. Its ability to maintain high performance in both training and testing contexts indicates not only strong internal learning but also robust generalization to unseen cases. This is especially relevant given the heterogeneity typically present in mental health data, where subtle, distributed signals may not be easily captured by traditional models. The results, therefore, point to the potential of

convolutional approaches to support more reliable mental health screening tools, offering a valuable direction for future applications in digital health settings.

4.4. Feature Importance

To assess the local and global relative effects of study features, SHAP values were computed for the 1D-CNN model, which achieved the best overall performance. Local SHAP explanations for one true positive and one true negative case are shown in Figure 10, illustrating how specific features influenced individual predictions. Global SHAP results are presented in Figure 11a, summarizing feature importance, and Figure 11b, showing the direction and distribution of effects across the test set. Feature importance analyses for the ML models, which performed below the 1D-CNN, are provided in Supplementary Material S2, offering additional comparison of the results and allowing for a more detailed validation of variable influence and interactions across models.

Figure 10a shows a SHAP local feature importance plot for a *TP* case where the model correctly predicted mental health problems. Positive contributions from Age, Education Level, and Discrimination dominated the prediction, while a negative effect of STDs Diagnosis counterbalanced but did not reverse the outcome. Minor effects from healthcare and residence factors complemented the profile. The model associates older age, higher education, and discrimination exposure with increased mental health vulnerability in this case. Conversely, Figure 10b presents a local importance plot for a *TN* case where the model correctly predicted absence of mental health problems. The prediction was mainly driven by the absence of Discrimination, supported by negative contributions from Medical Assistance, Occupation, and STDs Diagnosis. Minor positive effects from Age, Sexual Identity Expression, and sexual behavior variables were present but outweighed by protective factors. The model interprets this profile—marked by low discrimination, limited healthcare reliance, and absence of STDs—as linked to lower mental health vulnerability.

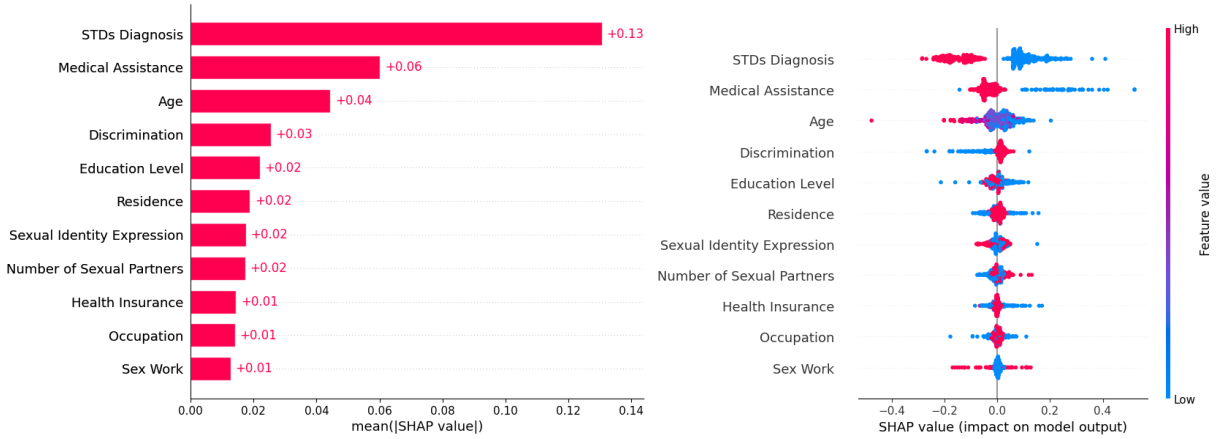


(**a**) (**b**)

Figure 10. Local importance plots for: (**a**) *TP* and (**b**) *TN* cases in the testing set.

Figure 11a presents the mean absolute SHAP values, reflecting each feature's overall importance in the prediction. Infectious disease diagnosis emerged as the most influential variable, followed by medical assistance and age—highlighting the central role of health and demographic factors. Sexual identity expression and discrimination also stood out as meaningful predictors, emphasizing behavioral and social dimensions. Other variables, including education, residence, sex work, and occupation, contributed to a lesser extent. Figure 11b shows the distribution of SHAP values, revealing nuanced effects of specific features on model predictions. For instance, younger age, absence of medical assistance, and no infectious disease diagnosis were associated with a higher predicted likelihood of

mental health issues, whereas prior medical attention or infectious disease diagnoses linked to lower risk, possibly reflecting healthcare access benefits. Sexual identity expression and discrimination contributed positively to risk predictions, while education, health insurance, and occupation showed minor and mixed effects.



(**a**) (**b**)

Figure 11. Global SHAP bar plot (**a**) and SHAP beeswarm plot (**b**) of 1D-CNN model.

Importantly, the 1D-CNN managed to capture interactions between the features, beyond their individual effects, due to its ability to model complex and nonlinear dependencies. The combined effect of discrimination and lack of medical assistance was linked to a notably higher risk prediction for mental health problems. Similarly, age modulated the impact of sexual identity expression, with younger individuals being more sensitive to social factors. Additionally, interactions between infectious disease diagnosis and health insurance likely influenced predictions, with limited coverage increasing risk for those with health issues.

To further examine the effect and complexity of the interactions among features that influence mental health problems, LRP heatmaps were generated from the trained 1D-CNN model to interpret both local and global predictions. Relevance scores reflect each feature's impact—positive values (red) support the prediction, while negative values (blue) oppose it. Figure 12a,b display local LRP explanations for individual *TP* and *TN* cases, respectively, whereas Figure 12c,d show global explanations summarizing relevance patterns across all *TP* and *TN* samples in the study.

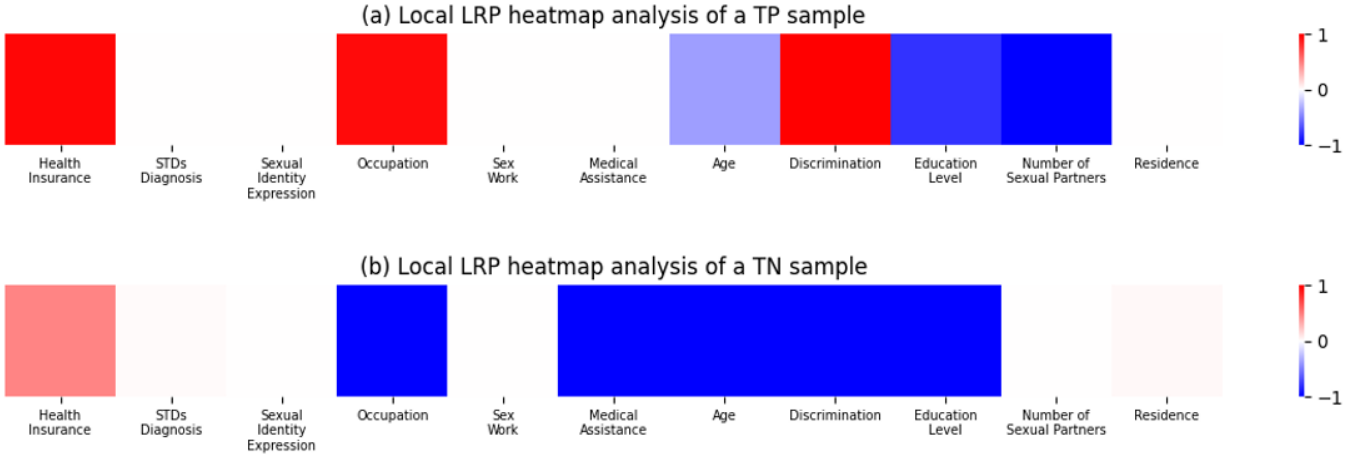


Figure 12. *Cont.*

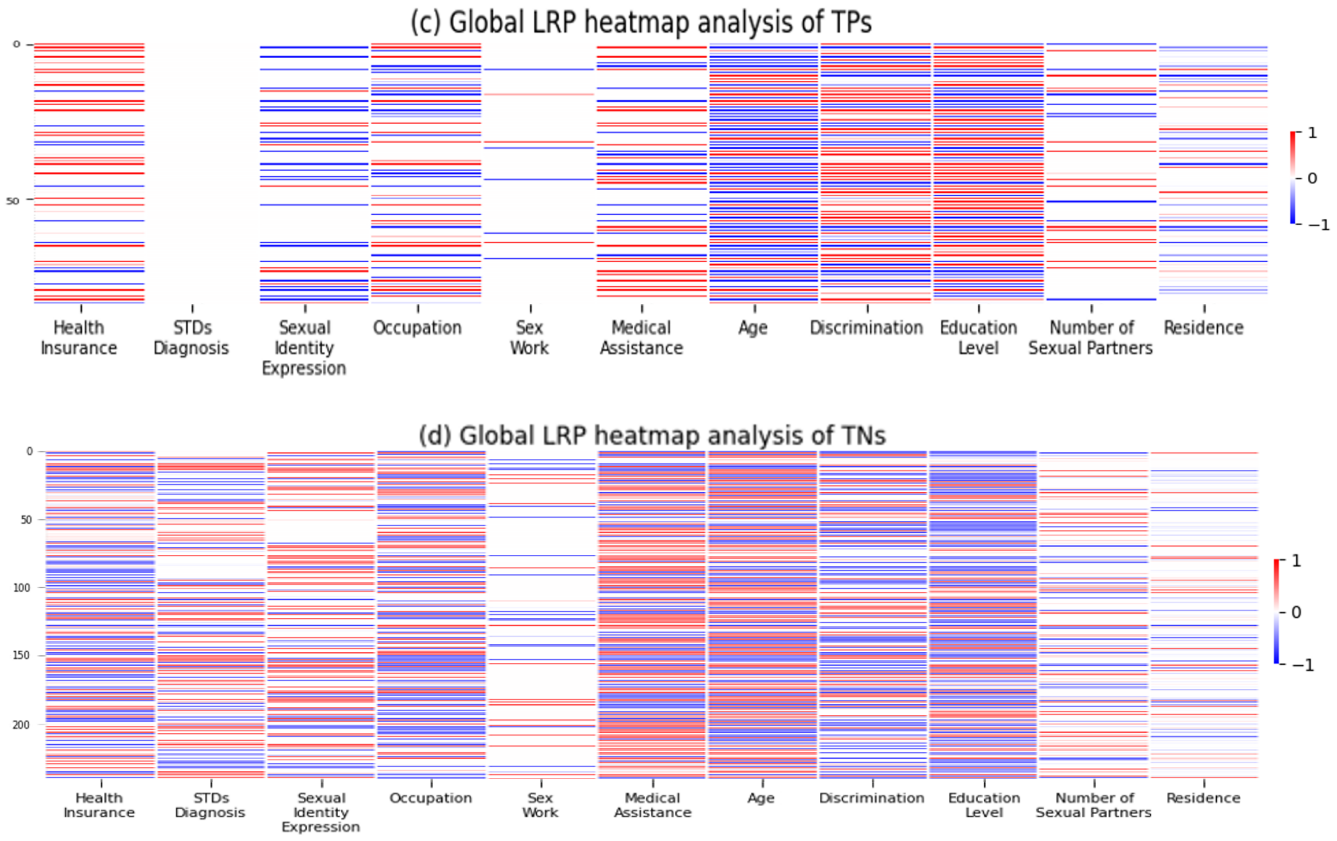


Figure 12. Visualizing LRP heatmaps for local (individual (*TP* (**a**) and *TN* (**b**))) and global (all (*TP* (**c**) and *TN* (**d**))) samples in the Mental Health Problems testing set.

Figure 12a shows an LRP heatmap for a *TP* case where the model correctly predicted mental health problems. This prediction was driven by higher levels of Health Insurance, Occupation, and Discrimination, alongside lower Education and fewer Sexual Partners. Age contributed moderately due to its relatively low value. These patterns suggest the model interprets secure health coverage, employment, discrimination exposure, limited education, fewer sexual partners, and middle age as a profile linked to increased mental health vulnerability. Similarly, Figure 12b shows an LRP heatmap for a *TN* case where the model correctly predicted no mental health problems. Lower levels of Occupation, Medical Assistance, Age, Discrimination, and Education strongly contributed, while Health Insurance and STD diagnosis had moderate positive relevance. This indicates the model associates limited discrimination, reduced healthcare use, lower education, younger age, combined with health insurance access and STD diagnosis, with lower mental health vulnerability.

Globally, Figure 12c displays the LRP heatmap for all correctly classified *TP* cases. Consistent positive relevance of Health Insurance, Medical Assistance, and Discrimination highlights their key role in predicting mental health issues. Age and Education show mixed relevance reflecting individual variability, while Sexual Identity Expression and Occupation have more case-dependent contributions. Figure 12d shows the global heatmap for *TN* cases. Features such as Age, Education Level, Sexual Identity Expression, and Discrimination exhibit strong negative relevance, indicating that lower values support predictions of no mental health problems. Similarly, Medical Assistance, Occupation, Health Insurance, and Infectious Disease Diagnosis show recurring negative relevance, consistent with profiles linked to better mental health outcomes in this sample.

These results, along with the earlier SHAP-based findings, provide a comprehensive understanding of the model's behavior. The SHAP analysis indicated the role of infectious

disease diagnosis, medical assistance, and age as the most influential predictors, followed by sexual identity expression and discrimination—emphasizing the combined influence of health, demographic, and social factors. These results align with the class-level relevance patterns observed in the LRP heatmaps, revealing the model's complex and nonlinear behavior. Understanding which factors drive *TP* and *TN* predictions can guide more effective policy design, targeted outreach, and risk assessment. Identifying socio-behavioral variables that signal mental health risk helps prioritize screenings and interventions, while recognizing features linked to *TNs* supports reinforcing protective factors.

5. Discussion

5.1. Summary of Findings

This study evaluated ML and DL models to predict mental health problems among individuals who self-identify as gay men in Peru. A dataset of 2186 participants from the 2017 First Virtual Survey for LGBTIQ+ People in Peru was used to examine how demographic, economic, health-related, behavioral, and social factors influence mental health outcomes, and to develop predictive models optimized for performance and generalizability. In total, 38.7% of participants reported experiencing mental health problems in the past 12 months, suggesting that this population experiences significant psychological distress, likely influenced by social and structural factors. Conversely, among the models evaluated, the 1D-CNN achieved the best overall results, outperfoming Logistic Regression, Artificial Neural Networks, and ensemble methods, such as Random Forest, XGBoost and GBM, across training and testing phases, demonstrating improved accuracy and discrimination compared to conventional ML approaches. Finally, explainability analyses based on SHAP and LRP revealed that infectious disease diagnosis, access to medical assistance, and age were the most influential predictors, followed by sexual identity expression and experiences of discrimination. Together, these variables illustrate the combined influence of health, demographic, and social dimensions on mental health vulnerability, and underscore the model's capacity to capture complex, nonlinear interactions among them. These findings highlight the value of predictive modeling in recognizing patterns associated with mental health concerns and improving early identification strategies for this population.

5.2. Mental Health Determinants and Predictive Modeling

The findings of this study reaffirm the significant mental health burden among self-identified gay men in Peru, aligning with broader research on LGBTIQ+ populations. National survey data reported that 23.8% of LGBTIQ+ respondents experienced mental health problems in the past year, most commonly anxiety and depression [2]. Within this population, bisexual and non-binary individuals have shown higher levels of distress [1], while open expression of identity appears protective, reducing mental health problems by 17% [4]. Conversely, discrimination and violence remain prevalent, reported by 69.1% of participants—particularly transgender individuals—and are strongly associated with poor mental health outcomes [41]. Among gay men, discrimination increased the prevalence of mental health issues by 72% [2], illustrating the compounded effect of stigma and marginalization.

Health-related vulnerabilities also emerged as relevant predictors. MSM face elevated rates of infectious diseases such as syphilis, gonorrhea, and HIV, which are frequently linked to emotional distress and suicidal ideation [3,42,43]. These physical health challenges are further aggravated by structural barriers: individuals lacking access to medical care or health insurance are more likely to report anxiety and depression [1,29], while discriminatory attitudes from healthcare providers exacerbate mental health disparities [4,41]. Repeated exposure to exclusion in healthcare, education, and employment settings has been

linked to cumulative psychological harm [3,4]. Moreover, age and identity expression interact with mental health vulnerability: younger MSM are at greater risk of emotional distress and suicidality [44,45], whereas older individuals may benefit from greater resilience [41]. Strong identification with one's sexual orientation can intensify exposure to stigma [17], while concealment is associated with higher rates of mental health problems [4]. Notably, non-binary individuals report significantly higher levels of anxiety and depression compared to their cisgender counterparts [3].

Furthermore, predictive modeling offers valuable knowledge. Our DL) model—specifically, a 1D-CNN—outperformed traditional ML classifiers, which already achieved robust results. Over the past decade, research in mental health diagnosis has gradually shifted focus toward the development of deep learning algorithms. Nonetheless, machine learning methods have played, and still play, a significant role in this domain [12]. Previous studies have reported 75–90% accuracy in predicting depression and anxiety using ML [46], and up to 94% accuracy for psychiatric conditions using CNNs and ANNs [24]. ML has also been successfully applied to behavioral data from smartphones, identifying depressive symptoms with 83% accuracy based on indicators such as irregular sleep and reduced activity [47]. Similarly, suicide risk prediction models that integrate clinical and behavioral variables have surpassed 80% accuracy [48]. The strength of DL, and particularly 1D-CNNs, lies in their automatic feature learning and scalability to large, sequential datasets without requiring manual feature engineering [49,50]. These models are also computationally efficient [51,52], handle local and long-range dependencies [53], and are especially well suited for vulnerable populations with complex health profiles [54]. Our findings support the applicability of 1D-CNNs for mental health assessment in LGBTIQ+ subgroups and underscore the importance of combining sociostructural insights with predictive modeling in public health research.

6. Conclusions and Recommendations

This study explored the use of predictive models to assess mental health status among individuals who self-identified as gay men in Peru, identify those at elevated risk based on variables commonly included in health surveys, and examine the influence of social and health-related factors on mental health outcomes in this population.

Machine learning and deep learning approaches, particularly the 1D-CNN model, revealed patterns associated with psychological distress among the study population. The application of explainability techniques, such as SHAP values and LRP heatmaps, underscored the relevance of social and health determinants in predicting mental health risks within marginalized communities. From a technical standpoint, these models demonstrated their capacity to process complex relationships in large-scale datasets, enabling risk stratification based on multiple interacting predictors.

6.1. Strengths and Limitations

As a key strength, this study is among the first to apply and compare multiple machine learning and deep learning models to predict mental health problems among individuals who self-identified as gay men in Peru. Through the analysis of a previously unexplored dataset with relevant sociodemographic and psychological information, it provides a structured evaluation of predictive approaches and offers a valuable baseline for future research and public health initiatives focused on sexual minority populations in similar sociocultural contexts.

Nonetheless, the study has several important limitations. First, the data were collected in 2017 through the First Virtual Survey for LGBTIQ+ People in Peru, which remains the only nationwide survey specifically addressing this population to date. The absence of more

recent or longitudinal data limits the study's capacity to capture dynamic sociopolitical, cultural, or public health changes that may have affected mental health risk factors in the intervening years. As such, the findings should be interpreted with caution regarding their temporal generalizability, and causal relationships between predictors and mental health outcomes cannot be inferred due to the cross-sectional nature of the data.

Second, the survey employed a non-probabilistic sampling strategy and relied on self-reported data collected online, which may have influenced the composition of the sample. Individuals with internet access and stronger ties to LGBTIQ+ networks are likely to be overrepresented, limiting the representativeness of the findings, particularly regarding more marginalized or digitally excluded groups.

Third, although the sample size and feature selection were adequate to support the application of machine learning and deep learning models and inference from the models' results, the original instrument lacked clinically and socially nuanced variables—such as psychiatric history, coping strategies, or detailed exposure to minority stressors—that could enhance model interpretability and predictive accuracy. The absence of such information constrains the models' capacity to fully capture the complexity of mental health outcomes in this population.

Fourth, while the 1D-CNN model demonstrated superior performance, its evaluation was limited to internal cross-validation within a single dataset. External validation with independent and updated datasets is essential to assess the model's robustness, fairness, and generalizability across different subgroups and settings. Moreover, the practical implementation of such models may face barriers related to computational demands and ethical concerns, including the risk of reinforcing stigma or excluding vulnerable individuals.

Lastly, while this study focuses specifically on individuals who self-identify as gay men in Peru, several identified risk patterns and protective factors, such as socioeconomic adversity, infectious disease diagnosis, access to medical care, experiences of discrimination, and concealment of sexual identity, are also commonly reported across LGBTIQ+ populations in other low- and middle-income countries. Thus, the findings may generalize to similar sociocultural contexts marked by structural stigma, healthcare barriers, and limited access to mental health services. Nevertheless, generalization should be approached with caution, as differences in structural conditions, cultural norms, and policy environments may significantly influence mental health trajectories, the heterogeneity within other sexual minority groups, and model performance. Thus, future studies should examine the applicability of these frameworks using context-specific data from diverse vulnerable populations to ensure context-sensitive interpretations.

6.2. Study Implications and Future Scope

While the primary dataset used in this study was collected in 2017, more recent national reports indicate that the structural barriers and mental health disparities it captured remain largely unchanged [7]. Persistent gaps in the availability and accessibility of mental health services for LGBTIQ+ individuals in Peru continue to be documented [7]. Despite increased public attention to mental health in recent years, the specific needs of sexual minorities—particularly gay men—remain insufficiently addressed in both research and policy. National evidence shows that this group continues to report elevated psychological distress, limited access to supportive environments, and exposure to health services that often lack training in affirmative practices [7]. Internationally, similar structural and clinical limitations have been reported in comparable sociopolitical contexts, where health systems have yet to fully integrate inclusive and non-pathologizing approaches [55].

In this context, the present study offers an efficient application of machine and deep learning models to predict mental health risk among gay men in Peru—a population for which data-driven research remains scarce. The findings contribute to strengthening local research capacity and provide a replicable framework for similar low- and middle-income settings where traditional statistical methods have predominated. In addition to its methodological contributions, the study produces interpretable outputs that can inform targeted interventions. By identifying key predictors linked to mental health outcomes, the analysis supports the development of more responsive programs and underscores the potential of computational methods in public mental health research within underrepresented populations.

The development of predictive models based on variables commonly collected in health surveys suggests that these approaches could be integrated into routine screening processes to identify individuals at higher risk of mental health problems. In Peru, where mental health issues are frequently under-diagnosed and services are limited, such models could support early detection efforts and enhance resource allocation within already overburdened mental health systems. Nevertheless, because the data were collected in 2017, future work should prioritize the use or collection of more recent datasets to ensure that predictive models reflect current social conditions and determinants. Societal shifts in attitudes, policy, and healthcare access may have altered risk patterns, making periodic model retraining and temporal validation essential.

While additional validation is needed, the integration of predictive modeling into Peru's public health system could improve the targeting of interventions—particularly for vulnerable groups such as gay men, who continue to face systemic and social barriers to care. Early identification of high-risk profiles could enable the implementation of preventive strategies, reduce the burden of untreated mental health conditions, and contribute to narrowing health disparities. Furthermore, the modeling approach and interpretability techniques employed here may be generalizable to similar populations in other low- and middle-income countries, or to groups that face comparable structural vulnerabilities and barriers to healthcare. The study also lays the groundwork for future research on the use of artificial intelligence in mental health, suggesting that the incorporation of a broader range of clinical, psychosocial, and contextual variables could enhance classification performance. A richer understanding of mental health risks could, in turn, support the development of culturally sensitive and population-specific interventions.

Longitudinal studies will also be critical to assess whether models retain predictive validity over time and can adapt to evolving epidemiological patterns, particularly within the shifting landscape of public health in Peru. By demonstrating the potential of deep learning, the study emphasizes the broader role of AI in complementing existing assessment tools and public health strategies. With further development, predictive models could be embedded in national mental health programs, facilitating more precise interventions, improving early diagnosis, and expanding service access for marginalized groups. From a policy standpoint, these findings underscore the need to invest in digital health infrastructure, strengthen data collection systems, and promote the integration of AI-driven frameworks in mental health surveillance and planning. Predictive analytics can inform more equitable, evidence-based decision-making, helping to prioritize high-risk populations and optimize the use of scarce mental health resources in Peru.

Finally, exploring lightweight or low-resource model variants could facilitate the deployment of predictive tools in real-world healthcare settings with limited computational infrastructure. Future research may also benefit from integrating multimodal data sources and hybrid model architectures that could provide deeper knowledge of mental health determinants, offering more robust tools for public health monitoring and decision-making.

Incorporating recurrent architectures such as LSTM or GRU networks could enhance the detection of temporal patterns in mental health data. In parallel, adaptive or heuristic-based hyperparameter optimization strategies may further refine classification performance. Continued exploration of artificial intelligence techniques is strongly encouraged given their promising applications for strengthening public health responses, informing mental health policy, and ultimately improving outcomes for vulnerable populations in Peru [56].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/informatics12030060/s1, Supplementary Material S1: Evaluation Metrics for ML and DL techniques. Supplementary Material S2: Feature importance for ML techniques.

Author Contributions: Conceptualization: A.A.-F.; methodology: A.A.-F.; validation: A.A.-F.; formal analysis: A.A.-F. and E.E.-P.; investigation: A.A.-F. and E.E.-P.; resources: A.A.-F. and E.E.-P.; data curation: A.A.-F.; writing—original draft preparation: A.A.-F. and E.E.-P.; writing—review and editing: A.A.-F. and E.E.-P.; visualization: A.A.-F.; supervision: A.A.-F.; project administration: A.A.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are available on the request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LGBTIQ+	Lesbian, Gay, Bisexual, Transgender, Intersex, Queer, and Non-binary, Pansexual, and Asexual
ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence
CNN	Convolutional Neural Networks
RNN	Recurrent Neural Networks
EEG	Electroencephalograms
NLP	Natural Language Processing
BERT	Bidirectional Encoder Representations from Transformers
SHAP	Shapley Additive exPlanations
MSM	Men Who Have Sex With Other Men
LSTM	Long Short-term Memory
GRU	Gated Recurrent Unit

Appendix A

Table A1. Operational definition of the factors used in the study.

Variable	Definition	Type	Categories
Mental Health Problems	In the last 12 months, the individual has experienced mental health issues.	Dependent	1 = Yes 0 = No
Health Insurance	Whether the respondent currently has health insurance coverage or not.	Predictor	1 = Yes 0 = No
Number of sexual partners	Indicates the extent of sexual activity within the past 12 months of the interviewee.	Predictor	0 = No sexual partners 1 = One sexual partner 2 = More than one sexual partner
Residence	Refers to the location of the respondent's primary residence.	Predictor	1 = Lima 0 = Inland regions
Education level	The highest level of education that the respondent has.	Predictor	0 = No education to incomplete elementary school 1 = Complete elementary school to incomplete high school 2 = Complete high school to incomplete superior education 3 = Complete superior education to postgraduate education
Infectious Diseases Diagnosis (STDs diagnosis)	In the last 12 months, the individual has been diagnosed with infectious diseases such as tuberculosis, STIs and HIV.	Predictor	1 = Yes 0 = No
Sexual identity expression	Whether the respondent openly expresses their sexual orientation or gender identity.	Predictor	1 = Yes 0 = No
Occupation	Indicates whether the respondent has engaged in any form of remunerated work within the past 12 months.	Predictor	1 = Yes 0 = No
Sex Work	Whether the respondent has engaged in any form of commercial sexual activity for financial gain.	Predictor	1 = Yes 0 = No
Medical assistance	Whether the interviewee utilized medical services for health-related issues or not in the last 12 months.	Predictor	1 = Yes 0 = No
Discrimination	Whether the respondent has experienced discrimination due to their sexual identity in the past.	Predictor	1 = Yes 0 = No
Age	Age of the respondent	Predictor	1 = 18–24 years 2 = 25–34 years 3 = 35–44 years 4 = More than 45 years

Table A2. Dataset distribution for the study.

Dependent Variable	Instances	Sample Size	Class 0	Class 1
Mental Health Problems	Training set (80%)	1748	1063	685
	Test set (20%)	438	276	162

References

1. Castillo, A.; Cornejo, D. Factores Asociados al Autoreporte de Depresión y Ansiedad en los Últimos Doce Meses en Personas LGTBI vía una Encuesta Virtual en Perú, 2017. Bachelor's Dissertation, Universidad Peruana de Ciencias Aplicadas, Lima, Peru, 2020.
2. Soriano-Moreno, D.; Saldaña-Cabanillas, D.; Vasquez-Yeng, L.; Valencia-Huamani, J.; Alave-Rosas J.; Soriano, A. Discrimination and mental health in the minority sexual population: Cross-sectional analysis of the first peruvian virtual survey. *PLoS ONE* **2022**, *17*, E0268755. [CrossRef] [PubMed]
3. Ponce, D. Factores Asociados a Problemas de Salud Mental. 2022. Análisis de la Primera Encuesta Virtual para Personas LGBTI, Perú. Bachelor's Dissertation, Universidad Ricardo Palma, Lima, Peru, 2017.
4. Castaneda, J.; Poma, N.; Mougenot, B.; Herrera-Añazco, P. Association between the Expression of Sexual Orientation and/or Gender Identity and Mental Health Perceptions in the Peruvian LGBTI Population. *Int. J. Environ. Res. Public Health* **2023**, *20*, 5655. [CrossRef] [PubMed]
5. Herek, G.; Garnets, L. Sexual orientation and mental health. *Annu. Rev. Clin. Psychol.* **2007**, *3*, 353–375. [CrossRef]
6. Alegría, M.; NeMoyer, A.; Falgàs, I.; Wang, Y.; Alvarez, K. Social Determinants of Mental Health: Where We Are and Where We Need to Go. *Curr. Psychiatry Rep.* **2018**, *20*, 95. [CrossRef] [PubMed]
7. Más Igualdad Perú. II Estudio de Salud Mental LGBTIQ+—Informe Final. 2024. Available online: https://www.masigualdad.pe/_files/ugd/4aec54_c8d6e0ecec9c43fab85c79aaa04bbbd1.pdf (accessed on 10 January 2025).
8. INEI (National Institute of Statistics and Informatics). Primera Encuesta Virtual para Personas LGBTI 2017: Principales Resultados. 2017. Available online: https://www.inei.gob.pe/media/MenuRecursivo/boletines/lgbti.pdf (accessed on 10 January 2025).
9. Más Igualdad Perú. Salud Mental de Personas LGBTQ+ en Perú—Informe Final. 2021. Available online: https://www.masigualdad.pe/_files/ugd/4aec54_d267bbb3a8564e1980f90ccd15281c39.pdf (accessed on 11 January 2025).
10. Roth, C.; Papassotiropoulos, A.; Brühl, A.; Lang, U.; Huber, C. Psychiatry in the Digital Age: A Blessing or a Curse? *Int. J. Environ. Res. Public Health* **2021**, *18*, 8302. [CrossRef]
11. Graham, S.; Depp, C.; Lee, E.; Nebeker, C.; Tu, X.; Kim, H.; Jeste, D. Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. *Curr. Psychiatry Rep.* **2019**, *21*, 116. [CrossRef]
12. Iyortsuun, N.; Kim, S.; Jhon, M.; Yang, H.; Pant, S. A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. *Healthcare* **2023**, *11*, 285. [CrossRef]
13. Huang, Z. Prediction of Mental Problem Based on Deep Learning Models. *Appl. Comput. Eng.* **2025**, *138*, 56–63. [CrossRef]
14. Madububambachu, U.; Ukpebor, A.; Ihezue, U. Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review. *Clin. Pract. Epidemiol. Ment. Health* **2024**, *20*, e17450179315688. [CrossRef]
15. Kundu, N.; 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. Inf.* **2021**, *9*, e28962. [CrossRef]
16. Chapagain, S. Predictive Insights into LGBTQ+ Minority Stress: A Transductive Exploration of Social Media Discourse. *arXiv* **2024**, arXiv:2411.13534v1.
17. Malik M.; Iqbal, S.; Noman, M.; Sarfraz, Z.; Sarfraz, A.; Mustafa, S. Mental Health Disparities Among Homosexual Men and Minorities: A Systematic Review. *Am. J. Men's Health* **2023**, *17*, 15579883231176646. [CrossRef]
18. Frost, D.; Meyer, I.; Lin, A.; Wilson, B.; Lightfoot, M.; Russell, S.; Hammack, P. Social Change and the Health of Sexual Minority Individuals: Do the Effects of Minority Stress and Community Connectedness Vary by Age Cohort? *Arch. Sex. Behav.* **2022**, *51*, 2299–2316. [CrossRef] [PubMed]
19. Herek, G.; Gillis, J.; Cogan, J. Internalized Stigma Among Sexual Minority Adults: Insights From a Social Psychological Perspective. *J. Couns. Psychol.* **2009**, *56*, 32–43. [CrossRef]
20. World Health Organization; Calouste Gulbenkian Foundation. *Social Determinants of Mental Health*; World Health Organization: Geneva, Switzerland, 2014. Available online: https://iris.who.int/bitstream/handle/10665/112828/9789241506809_eng.pdf (accessed on 21 January 2025).
21. Kirkbride, J.; Anglin, D.; Colman, I.; Dykxhoorn, J.; Jones, P.; Patalay, P.; Pitman, A.; Soneson, E.; Steare, T.; Wright, T.; Griffiths, S. The social determinants of mental health and disorder: Evidence, prevention and recommendations. *World Psychiatry Off. J. World Psychiatr. Assoc.* **2024**, *23*, 58–90. [CrossRef] [PubMed]

22. Zhang Y.; Chen, S.; Akintunde, T.; Okagbue, E.; Isangha, S.; Musa, T. Life course and mental health: A thematic and systematic review. *Front. Psychol.* **2024**, *15*, 1329079. [CrossRef]
23. Shim, R.; Compton, M. Addressing the Social Determinants of Mental Health: If Not Now, When? If Not Us, Who? *Psychiatr. Serv.* **2018**, *69*, 2018000. [CrossRef]
24. Burrichter, K. The use of machine learning algorithms to predict mental health outcomes based on behavioral data collected through digital devices. *Arch. Clin. Psychiatry* **2022**, *49*, 122–129.
25. Zawad, R.; Haque, Y.; Kaiser, M.; Mahmud, M.; Chen, T. Computational Intelligence in Depression Detection. In *Artificial Intelligence in Healthcare—AI in Mental Health*; Chen, T., Carter, J., Mahmud, M., Khuman, A., Eds.; Springer Nature: Singapore, 2022; pp. 1–19. [CrossRef]
26. Saleh, A.; Xian, L. Stress Classification using Deep Learning with 1D Convolutional Neural Networks. *Knowl. Eng. Data Sci.* **2021**, *4*, 145–152. [CrossRef]
27. Awan, A.; Taj, I.; Khalid, S.; Usman, S.; Imran, A.; Akram, M. Advancing Emotional Health Assessments: A Hybrid Deep Learning Approach Using Physiological Signals for Robust Emotion Recognition. *IEEE Access* **2024**, *12*, 141890–141904. [CrossRef]
28. Ige, A.; Sibiya, M. State-of-the-Art in 1D Convolutional Neural Networks: A Survey. *IEEE Access* **2024**, *12*, 144082–144105. [CrossRef]
29. Romani, L.; Ladera-Porta, K.; Quiñones-Laveriano, D.; Rios-Garcia, W.; Juarez-Ubillus, A.; Vilchez-Cornejo, J. Factors associated with the non-use of health services in LGBTI people from Peru. *Rev. Peru. Med. Exp. Salud Publ.* **2021**, *38*, 240–247. [CrossRef] [PubMed]
30. Noorunnahar, M.; Chowdhury, A.; Mila, F. A tree-based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *PLoS ONE* **2023**, *18*, e0283452. [CrossRef]
31. Kulkarni, V.; Sinha, P. Random Forest Classifiers: A Survey and Future Research Directions. *Int. J. Adv. Comput.* **2013**, *36*, 1144–1153.
32. Touzani, S.; Granderson, J.; Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* **2018**, *158*, 1533–1543. [CrossRef]
33. Bonilla, M.; Olmeda, I.; Puertas, R. Modelos paramétricos y no paramétricos en problemas de credit scoring. *Rev. Esp. Financ. Contab.* **2003**, *32*, 833–869.
34. Hosmer, D.; Lemeshow, S.; Sturdivant, R. Introduction to the Logistic Regression Model. In *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; pp. 1–35.
35. Qazi, E.; Almorjan, A.; Zia, T. A One-Dimensional Convolutional Neural Network (1D-CNN) Based Deep Learning System for Network Intrusion Detection. *Appl. Sci.* **2022**, *12*, 7986. [CrossRef]
36. Hernández-Vásquez, A.; Chacón-Torrico, H. Manipulación, análisis y visualización de datos de la encuesta demográfica y de salud familiar con el programa R. *Rev. Peru. Med. Exp. Salud Publ.* **2019**, *36*, 128–133. [CrossRef]
37. Lateef, R.; Abbas, A. Tuning the Hyperparameters of the 1D CNN Model to Improve the Performance of Human Activity Recognition. *Eng. Technol. J.* **2022**, *40*, 547–554. [CrossRef]
38. Lundberg, S.; Lee, S. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017. [CrossRef]
39. Sahakyan, M.; Aung, Z.; Rahwan, T. Explainable artificial intelligence for tabular data: A survey. *IEEE Access* **2021**, *9*, 135392–135422. [CrossRef]
40. Ullah, I.; Rios, A.; Gala, V.; Mckeever, S. Explaining Deep Learning Models for Tabular Data Using Layer-Wise Relevance Propagation. *Appl. Sci.* **2022**, *12*, 136. [CrossRef]
41. Guerra, M. Análisis Exploratorio Sobre la Violencia y/o Discriminación Reportados por Personas LGBTI en el Perú, 2017. Bachelor's Dissertation, Universidad Peruana de Ciencias Aplicadas, Lima, Peru, 2023.
42. Trust for America's Health. Addressing the Social Determinants of Health Inequities Among Gay Men and Other Men Who Have Sex with Men in the United States. 2014. Available online: https://www.tfah.org/wp-content/uploads/archive/assets/files/TFAH-2014-MSM-Report-final.pdf (accessed on 6 February 2023).
43. Lamontagne, E.; Leroy, V.; Yakusik, A.; Parker, W.; Howell, S.; Ventelou, B. Assessment and determinants of depression and anxiety on a global sample of sexual and gender diverse people at high risk of HIV: A public health approach. *BMC Public Health* **2024**, *24*, 215. [CrossRef]
44. Zhang, X.; Zhou, Y.; Zhang, K. Social capital, perceived stress, and mental health of men who have sex with men in China: A cross-sectional study. *Front. Psychol.* **2023**, *14*, 1134198. [CrossRef]
45. Bränström, R.; Hughes, T.; Pachankis, J. Global LGBTQ Mental Health. In *Global LGBTQ Health*; Springer: Berlin, Germany, 2024; pp. 45–78.
46. Radhika, C.; Shraddha, N.; Vaishnavi, P.; Shirisha, K. Prediction of Mental Health Instability using Machine Learning and Deep Learning Algorithms. *J. Comput. Sci. Appl.* **2023**, *15*, 47–58.

47. Laijawala, V.; Aachaliya, A.; Jatta, H.; Pinjarkar, V. Mental Health Prediction using Data Mining: A Systematic Review. In Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST), Bahir Dar, Ethiopia, 4–6 November 2022.
48. Madineni, S. Mental Health Survey Analysis & Prediction Using Deep Learning Algorithms. In Proceedings of the Symposium of Student Scholars, Kennesaw, GA, USA, 18–21 April 2023. Available online: https://digitalcommons.kennesaw.edu/undergradsymposiumksu/spring2023/presentations/113 (accessed on 15 April 2025)
49. Taye, M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* **2023**, *12*, 91. [CrossRef]
50. Karima, S.; Ouassila, H. An overview of machine learning and deep learning. *Alger. J. Sci.* **2021**, *14*, 139–143.
51. Singh, K.; Mahajan, A.; Mansotra, V. 1D-CNN based Model for Classification and Analysis of Network Attacks. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 604–613. [CrossRef]
52. Xiao, D.; Chen, Y.; Li, D. One-Dimensional Deep Learning Architecture for Fast Fluorescence Lifetime Imaging. *IEEE J. Sel. Top. Quantum Electron.* **2021**, *27*, 7000210. [CrossRef]
53. Odeniyi, O.; Oyinloye, O.; Thompson, A. Fraud Detection Using Multilayer Perceptron and Convolutional Neural Network. *Int. J. Adv. Secur.* **2021**, *14*, 1–11.
54. Zhou, P. Capacity estimation of retired lithium-ion batteries using random charging segments from massive real-world data. *Cell Rep. Phys. Sci.* **2021**, *6*, 102444. [CrossRef]
55. See, J. What it Means to Suffer in Silence. Challenges to Mental Health Access among LGBT People, 2019. Galen Centre for Health and Social Policy. Available online: https://galencentre.org/wp-content/uploads/2019/04/PFA02_2019_Challenges-to-Mental-Health-Access-Among-LGBT-People.pdf (accessed on 5 June 2025).
56. Zuiderwijk, A.; Chen, Y.; Salem, F. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Gov. Inf. Q.* **2021**, *38*, 101577. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.