
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202412.00079

Machine Learning Applications for Enhancing Predictive Power in Non-suicidal Self-injury: A Systematic Review

Authors: Gao Baixue, Xie Yunlong, Luo Junlong, He Wen, He Wen

Date: 2024-12-04T00:00:00+00:00

Abstract

Non-suicidal self-injury (NSSI) is a major public health problem characterized by high stigmatization, high complexity, and high heterogeneity. Traditional NSSI research has limited measurement and analytical methods, resulting in low predictive power of identified risk factors. In recent years, machine learning has been gradually applied to the analysis and modeling of NSSI, and through improving the predictive power of NSSI research tools, increasing the complexity and precision of predictive models, and distinguishing NSSI categories and subtypes, has raised overall predictive performance to a moderate level. Future research should integrate traditional NSSI theories and research methods to make screening criteria more stringent, expand the integration of non-questionnaire NSSI data with deep learning and unsupervised learning, and increase model replicability and comparability according to the principle of ‘subtyping first, then transfer’, to further improve predictive performance.

Full Text

Application of Machine Learning to Improve Predictive Performance in Non-Suicidal Self-Injury: A Systematic Review

GAO Baixue¹, **XIE** Yunlong¹, **LUO** Junlong^{1, 2}, **HE** Wen^{1, 2}

¹College of Psychology, Shanghai Normal University, Shanghai 200234, China

²Lab for Educational Big Data and Policymaking, Ministry of Education, Shanghai Normal University, Shanghai 200234, China

Abstract

Non-suicidal self-injury (NSSI) represents a significant public health challenge characterized by profound stigma, high complexity, and substantial heterogeneity. Traditional measurement and analytical approaches in NSSI research have been limited, yielding low predictive power for identified risk factors. In recent years, machine learning has been progressively applied to NSSI analysis and modeling, elevating overall predictive performance to moderate levels by enhancing the predictive capacity of research instruments, increasing model complexity and precision, and distinguishing NSSI categories and subtypes. Future research should integrate traditional NSSI theories and methodologies to establish more rigorous screening criteria, expand the combination of non-questionnaire NSSI data with deep learning and unsupervised learning, and enhance model replicability and comparability based on the principle of “subtyping first, then transfer learning,” thereby further improving predictive performance.

Keywords: machine learning, non-suicidal self-injury, predictive power, application

Non-suicidal self-injury (NSSI) refers to the deliberate alteration or infliction of damage to one’s body tissue without intent to die (Jiang et al., 2011). NSSI constitutes a highly prevalent psychological disorder that exists widely across populations, with adolescents being the most severely affected group. The detection rate of self-injury among Chinese adolescents ranges from 24.9% to 29% (Qu et al., 2023; Tang et al., 2018). NSSI carries severe destructive consequences, causing direct physical harm and serving as a crucial risk indicator for suicide—indeed, it predicts suicidal behavior more strongly than a history of suicide attempts (Ribeiro et al., 2016). NSSI frequently co-occurs with other mental disorders such as depressive disorders, anxiety disorders, borderline personality disorder, and substance addiction (Nitkowski & Petermann, 2010; Sun et al., 2020). The treatment costs and productivity losses associated with NSSI impose substantial financial burdens on individuals and governments (Kinchin et al., 2017). Therefore, improving the predictive power of NSSI research is essential for accurate and timely intervention, holding significant value for both individuals and society.

The stigmatization of NSSI substantially compromises research predictive power. Recent reviews and meta-analyses of NSSI research reveal that questionnaire surveys have become the dominant methodology (De Luca et al., 2023; Peel-Wainwright et al., 2021; Xiao et al., 2022). However, the profound stigma surrounding NSSI often impedes questionnaire administration or yields socially desirable responses. This stigma encompasses misunderstanding, prejudice, and discrimination from patients themselves, others, and mass media, leading NSSI individuals to experience guilt, shame, and embarrassment (Fu et al., 2020; Staniland et al., 2021), which causes them to conceal their behavior during surveys (Farajzadeh & Sadeghzadeh, 2023). Moreover, since adolescent NSSI prevalence far exceeds that of other populations, adolescents are frequently the

preferred research subjects. Large-scale data collection from this group requires consent not only from the adolescents themselves but also from teachers and parents. NSSI exhibits clear contagion effects among adolescents—awareness of friends' self-injury may trigger imitation (Syed et al., 2020)—which raises parental and teacher vigilance and refusal. Experimental NSSI research requires simulating self-injury pain through tasks involving electric shock, pressure pain, cold pressor, and heat pressor procedures, necessitating careful ethical consideration (Yan, 2015). Consequently, current NSSI research lacks effective measurement tools.

Simultaneously, the inherent high complexity of NSSI renders existing research methods inadequate for comprehensive and systematic understanding. NSSI influencing factors span multiple domains: biological and genetic aspects involve FKBP5 genes, serotonin, opioid systems, dopamine systems, and abnormalities in brain regions related to emotion, control, pain, and reward (Bai et al., 2023; Deng et al., 2022); individual trait aspects involve emotion regulation deficits, personality disorders, and low self-esteem; and social factors involve parent-child relationships, family education, and interpersonal stress (Hu et al., 2023; Jiang et al., 2022; Steinhoff et al., 2021). Single influencing factors hold limited significance; integrating multiple variable types across biological, social, and psychological domains better predicts NSSI occurrence and development (Huang et al., 2013). Theoretical models of NSSI emergence and development remain inconsistent, requiring consideration of as many factors as possible across different contexts and individuals (Yin et al., 2022), which substantially increases the difficulty of identification, prediction, and analysis. Traditional analytical methods impose limited variable selection, resulting in only a few NSSI predictors maintaining statistical significance after analysis and causing substantial information loss (Zhong et al., 2024). They may also impose linear relationships on more complex associations, preventing established predictive models from accurately reflecting real-world situations (Burke et al., 2019). Meta-analyses of NSSI research reveal that, amid a lack of protective factor studies (Yang et al., 2023), risk factors exhibit extremely weak predictive power (Fox et al., 2015).

High heterogeneity in NSSI research constitutes another important cause of low predictive power, arising from three distinct levels. First, NSSI co-occurs with multiple mental disorders—over 80% of self-injuring patients in general hospitals have mental disorders such as depression, anxiety, and alcohol use disorder (He et al., 2023; Sun et al., 2020)—meaning that NSSI with identical presentations may have different risk factors or different comorbidity patterns with various mental disorders (He et al., 2023). Second, NSSI itself presents different subtypes, with frequency, methods, severity, and functions varying across individuals (Baer et al., 2020; Daukantaite et al., 2021; Radziwiłłowicz & Lewandowska, 2017). Predictive factors show high heterogeneity across age, gender, and other groups, with effects sometimes operating in completely opposite directions (Akbari et al., 2024; Haregu et al., 2023). Existing subtyping methods primarily rely on single-dimensional features, neglecting complex combinations and interactions among different characteristics (Wang et al., 2024). Third, inconsistent

NSSI definitions and measurement methods may yield different predictive factors across samples and methods (Rahman et al., 2021). Therefore, addressing NSSI heterogeneity is crucial for improving predictive power.

In summary, limited measurement and analytical approaches, combined with high stigma, complexity, and heterogeneity of NSSI itself, have resulted in constrained predictive power in research. To enhance NSSI predictive capacity, researchers must actively seek alternative measurement tools while improving questionnaire efficacy. In NSSI analysis and modeling, more predictive factors should be incorporated to identify critical variables and establish more complex, accurate models. Simultaneously, effectively distinguishing highly heterogeneous NSSI categories and subtypes represents a key strategy for improving predictive power.

Machine learning (ML) is an automatic predictive modeling technique capable of searching high-dimensional data for numerous explanatory variables and identifying complex interactions among them. Currently, ML has been widely applied to identify and diagnose psychological problems and disorders such as dyslexia and depression (Bhadra & Kumar, 2022; Castillo-Sánchez et al., 2020; Bu et al., 2023; Dong et al., 2020). ML has also found mature applications in assessing NSSI incidence and key influencing factors (Burke et al., 2019). ML can analyze diverse NSSI variable types, including text, images, and video data (Lu et al., 2023; Xian et al., 2019), substantially improving model predictive power. It can incorporate more NSSI-related variables, model complex associations among them, and generate stronger predictive performance (Fox et al., 2019). ML enables visualization of NSSI predictor importance using random forests and decision trees, constructing decision tree models with higher predictive capacity than traditional models (Walsh et al., 2017). Additionally, ML employs data-driven methods to identify NSSI subtypes based on data patterns, objectively analyzing relationships and structures among symptom characteristics (Lanza & Cooper, 2016) and providing new insights into connections among different NSSI subtypes (Wang et al., 2024). Based on the latest research on ML and NSSI, this paper elaborates on ML applications for improving NSSI predictive power and provides in-depth reflection and future prospects for current research limitations and development directions.

We conducted a literature search on ML and NSSI using databases including CNKI, VIP, Wanfang, Baidu Scholar, Web of Science, Elsevier Science Direct, EBSCO, and PubMed. English search terms included “NSSI/Self-injury/Self-harm” AND “Machine Learning/Deep Learning,” while Chinese terms included “自伤/NSSI” AND “机器学习/深度学习.” Given the relatively short application period of ML in NSSI research and the limited number of studies, we imposed no restrictions on publication language, date, specific ML methods, or study populations. Inclusion criteria were: (1) empirical journal articles and conference papers; (2) studies applying ML to improve NSSI predictive power; and (3) studies using verifiable methods with detailed results and objective conclusions. Based on these criteria, we identified 24 articles for inclusion in this systematic

review (see Table 1). Figure 1 [Figure 1: see original paper] illustrates the literature screening process.

Figure 1 Literature screening flowchart

Table 1 Application of machine learning in the NSSI field

3.1 Data Collection

In terms of data collection methods, ML research primarily considers the requirement for large datasets. Only through training and testing on substantial data can ML generate models with higher predictive power (Yarkoni & Westfall, 2017). Currently, questionnaire surveys, clinical case data from NSSI patients, and social media data can meet this requirement. As shown in Table 1, questionnaire-based data collection dominates ML applications in NSSI, though questionnaire datasets are smaller than other methods. The smallest dataset came from a longitudinal study with 114 participants retested at six-month intervals (Chen et al., 2024). Notably, both this study and other NSSI reviews and meta-analyses demonstrate that China's large-scale questionnaire surveys far exceed other countries in sample size and publication quantity, reflecting China's advantages in big data research. Beyond questionnaire data, retrospectively retrieving patient cases offers higher efficiency and larger sample sizes, typically encompassing tens of thousands of cases. The largest dataset among included studies comprised 6,037,479 NSSI clinical cases (Kumar et al., 2020). Compared to questionnaire data, clinical data can more easily capture numerous NSSI features. Applying Natural Language Processing (NLP) to structured electronic health records enables construction of hundreds of features for analysis (Arora et al., 2023). In addition to questionnaires and clinical cases, social media contains massive amounts of NSSI-related text, images, and videos, also satisfying ML's large-sample requirements. One study analyzing self-injury prevalence on social platforms collected 1.2TB of materials from Instagram, including over 1.7 million self-injury images and 305,000 videos (Xian et al., 2019). However, these materials present greater challenges for quantitative analysis, resulting in extremely limited related research.

Large-sample requirements and methodological choices constrain study population selection. Questionnaire surveys typically use adolescents as research subjects to facilitate large-scale data collection. The largest included study sampled 13,304 adolescents on NSSI in western China (Zhong et al., 2024). Beyond the included journal articles, China has another large-sample study on adolescent NSSI involving over 180,000 middle school, vocational high school, and university students (Xu, 2023). Clinical case data involve not only NSSI patients but also other mental disorder populations, such as individuals with eating disorders (Cliffe et al., 2021) and schizophrenia spectrum disorders (Kappes et al., 2023). Social media data typically selects popular platforms like Instagram, Facebook, and Twitter, filtering data based on NSSI-related hashtags (Xian et al., 2019), though platform anonymity prevents determination of specific participant infor-

mation.

3.2 Model Algorithm Selection

After collecting sufficient NSSI samples, appropriate ML algorithms must be selected for analysis. Figure 2 [Figure 2: see original paper] illustrates current algorithm selection and usage patterns in NSSI research. The field predominantly employs shallow ML, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machines (SVM). This preference stems primarily from NSSI research data being dominated by structured data from questionnaires and demographics, whereas deep learning excels at learning from unstructured data like images and text (Kentopp, 2021). Second, deep learning relies on larger datasets for optimal performance; when data are scarce and inter-item differences are minimal, shallow learning demonstrates superior performance and more robust analysis (Farajzadeh & Sadeghzadeh, 2023). In summary, shallow learning better suits current NSSI data types.

Among shallow learning methods, Random Forest is most frequently used. Because NSSI is complex and multifactorial, Random Forest's high tolerance for data variation enables effective handling of anomalous data and analytical challenges (Jaiswal & Samikannu, 2017), making it widely applied for predicting self-injurious behavior (Burke et al., 2019). Extreme Gradient Boosting ranks second, serving as an optimized distributed gradient boosting library designed to efficiently, flexibly, and portably address NSSI key feature selection and classification problems (Xu et al., 2024). Additionally, both Random Forest and XGBoost include importance estimation for NSSI predictors, similar to variable importance estimation in traditional regression models (standardized coefficients), and are therefore often combined with traditional logistic regression for further model simplification (Chen et al., 2024; Fox et al., 2019)—that is, screening core variables with the greatest influence across multiple algorithms to construct models with higher predictive power (Zhong et al., 2024; Zhou et al., 2024).

Support Vector Machines are also suitable for pattern recognition tasks with small samples, nonlinearity, and high dimensionality. When NSSI sample sizes are small, SVM can avoid the curse of dimensionality (model performance degradation due to excessive feature numbers) and overfitting problems (Yang et al., 2022), effectively selecting key NSSI features and constructing more precise classification and predictive models (Arora et al., 2023; Lang et al., 2024; Lu et al., 2023). Consequently, SVM represents another mainstream application in the NSSI field.

Surprisingly, despite their broad applicability and powerful functionality, unsupervised learning methods have been neglected in NSSI research. Algorithms such as Latent Class Analysis (LCA), Latent Profile Analysis (LPA), and k-means Clustering, as typical unsupervised learning approaches, have found mature application in analyzing NSSI subtypes across features, frequency, and

methods (Case et al., 2020; Christoforou et al., 2021; Guerin-Marion et al., 2021; Murner-Lavanchy et al., 2022). However, in NSSI research, these unsupervised algorithms are often overlooked as important branches of ML. Moreover, unsupervised learning applications remain limited in NSSI; in other psychological issues like dyslexia, depression, and suicide, unsupervised learning has been widely used for key feature selection and extraction (Bhadra & Kumar, 2022; Castillo-Sánchez et al., 2020; Bu et al., 2023), extracting the most relevant and informative feature data from complex raw data (text, fMRI, EEG, eye-tracking) for modeling and analysis. Currently, ML applications in NSSI involve relatively homogeneous data types that rarely require feature extraction, thus limiting unsupervised learning applications.

In summary, current NSSI data are more structured with relatively small sample sizes, making shallow learning more appropriate, particularly Random Forest, XGBoost, and SVM algorithms. Random Forest and XGBoost can also assess and rank predictor importance. Random Forest is more easily understood and implemented, making it most widely applied in NSSI research. XGBoost requires more parameter tuning and optimization, offering higher precision and better application prospects for complex NSSI problems. Additionally, unsupervised ML algorithms remain underutilized for NSSI key feature selection and extraction, representing substantial development potential.

3.3 Model Training

Constrained by NSSI complexity and difficulties in measurement and sample collection, ML may learn not only relationships among NSSI variables but also “noise” between variables, resulting in models more complex than reality—an “overfitting” phenomenon. To select optimal NSSI prediction or classification models and reduce error, ML models require training through methods such as cross-validation (leave-one-out, k-fold, simple hold-out), bootstrapping, and regularization (Dong et al., 2021; Zhou, 2016).

As shown in Figure 2, simple hold-out validation is the most commonly used model training method in NSSI research. Researchers typically randomly split NSSI datasets into training and test sets, with training sets comprising approximately 60%–80% of the total and test sets excluded from training, used solely to evaluate model performance (Lang et al., 2024; Swaminathan et al., 2023; Zhong et al., 2024). Compared to simple hold-out validation, k-fold cross-validation better balances statistical power (Yang & Wang, 2015) and is therefore commonly used in NSSI research, typically employing 5-fold or 10-fold cross-validation—partitioning the original NSSI dataset 5 or 10 times, conducting ML training and evaluation for each partition, then averaging all results for the final score (Bao et al., 2024; Kumar et al., 2020; Su et al., 2023; Yin et al., 2021).

Some researchers emphasize the importance of Leave-One-Subject-Out Cross-Validation (LOSO) in NSSI research. Unlike standard k-fold cross-validation,

LOSO uses the entire dataset but learns based on NSSI patients rather than NSSI features, potentially yielding well-fitted models that diverge from the NSSI problem itself (Marti-Puig et al., 2022). Additionally, numerous studies employ flexible algorithms like Random Forest and XGBoost to investigate NSSI. Such algorithms lack additivity or linearity assumptions and allow arbitrary interactions among predictors, making them better suited for internal validation through bootstrapping or repeated cross-validation; otherwise, training sets overestimate model performance (Collins et al., 2015; Steyerberg et al., 2001). In conclusion, NSSI model training methods should be determined jointly by sample size and algorithm selection. Except for studies involving Random Forest and XGBoost that use bootstrapping, other research primarily employs cross-validation, with simple hold-out validation being most common, typically using 70% training/30% testing or 80% training/20% testing splits. Since simple hold-out validation involves only one training and testing round, this study recommends using it as a preliminary evaluation method or when sample sizes are excessively large. If NSSI sample sizes do not require computational cost savings, more advanced cross-validation methods should be adopted for more reliable performance estimation. When NSSI sample sizes are relatively small and sample homogeneity is high, LOSO should be considered for more efficient dataset utilization. When NSSI sample sizes are large with uneven category distributions, k-fold cross-validation and derived methods like stratified k-fold cross-validation should be employed.

3.4 Model Evaluation

As shown in Figure 2, Receiver Operating Characteristic Curve - Area Under Curve (ROC-AUC) is the most commonly used model evaluation metric in NSSI applications. AUC values of 0.70–0.79 indicate moderate performance, 0.80–0.89 good performance, and above 0.90 excellent performance (Simundic, 2009). AUC values for NSSI models in included literature generally range from 0.74 to 0.89, indicating overall moderate to good performance. Only one study showed poor model performance, where Random Forest could not classify NSSI and healthy populations based on neurocognitive features (Murner-Lavanchy et al., 2022). Additionally, larger sample sizes yield higher model performance, even reaching 0.99 (Kumar et al., 2020). With sufficiently large samples, deep learning models demonstrate superior AUC performance compared to shallow learning (Yin et al., 2021). Models built on natural language processing also outperform those based on structured data (Lu et al., 2023; Swaminathan et al., 2023).

AUC represents only a global measure of diagnostic accuracy, reporting no information about individual model parameters such as sensitivity and specificity (Simundic, 2009). Moreover, when NSSI participants are few, models can produce high AUC values with minimal clinical utility (Fox et al., 2019). Therefore, AUC should be combined with other metrics for comprehensive ML model evaluation in NSSI. According to Figure 2, these metrics are used with the following

frequency: Sensitivity, Accuracy, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV), F1 score, and Recall. Accuracy and Recall use the same standards as AUC (Franklin et al., 2017), while PPV and NPV are primarily used in NSSI binary classification and prediction (Chen et al., 2024; Mürner-Lavanchy et al., 2022; Swaminathan et al., 2023; Xu et al., 2024; Zhou et al., 2024).

4.1 Improving Research Instrument Predictive Power

ML addresses the challenge of low predictive power resulting from measurement difficulties through two pathways. First, by reducing dimensions or item numbers to obtain more concise and efficient questionnaire tools. Second, by enriching other effective measurement instruments. Farajzadeh et al. (2022) argue that self-injury scales should not be overly lengthy, should have clearer themes, and should omit questions that explicitly inquire about self-injury to effectively avoid participant fatigue or dishonesty. Using ML, they reduced 662 items from 17 commonly used NSSI scales plus demographic questionnaires and lie detection items to just 22 items that could accurately distinguish individuals with NSSI behavior (Accuracy = 83.6%). Additionally, when ML sample sizes are small, researchers prevent overfitting by first using Random Forest to analyze all questionnaire items and iteratively deleting the least important item until AUC can no longer improve (Train AUC = 1, 95% CI: 1–1), ultimately reducing 89 items to 34 (Chen et al., 2024) to obtain the most concise item set.

Researchers have also attempted to enrich NSSI research instruments using ML methods to improve predictive power. One study designed an application based on the Experience Sampling Method (ESM) to collect Ecological Momentary Assessment (EMA) data through three main components: emotion, daily activities, and self-injury behavior data. Combined with ML, this approach effectively assessed various emotional states and dysfunctional behaviors associated with NSSI (Accuracy = 84.78%). This method enables richer, more efficient, and convenient data collection while making timely NSSI intervention possible (Martí-Puig et al., 2022). An ongoing study employs a mixed-methods approach, collecting NSSI and emotion assessment questionnaires via EMA three times daily for 14 days while simultaneously using wearable devices to collect participants' heart rate, step count, and sleep patterns. Using dynamic structural equation modeling combined with ML to analyze NSSI internal mechanisms enables more comprehensive analysis and effective prediction of NSSI thoughts and behaviors (Ahn & Lee, 2024). ML also facilitates effective application of social media data for NSSI identification and prediction. NSSI individuals are less likely to seek help from adults or professionals, preferring instead to seek support through text messages, social media, or informal online platforms (Belfort & Miller, 2018), generating massive amounts of NSSI-related images, videos, comments, and posts. Using ML to train classifiers to identify NSSI can achieve 94% accuracy (Xian et al., 2019), far exceeding traditional research. In summary, streamlined questionnaires and enriched research tools open new pathways for

improving NSSI predictive power.

4.2 Increasing Model Complexity and Precision

NSSI involves numerous predictive factors and high complexity. Traditional research incorporates insufficient influencing factors, resulting in low model predictive power. NSSI ML research includes more predictive factors than traditional studies, with most integrating multiple data types such as demographic, behavioral, and physiological data (Lang et al., 2024; Su et al., 2023; Xu et al., 2024), enhancing NSSI predictive power by increasing model complexity (Fox et al., 2019).

Considering clinical applicability, although ML incorporates more NSSI-related variables, it typically retains only the most important predictive factors. Researchers commonly use ML to rank NSSI variable importance and further screen key predictors. One study used SVM as a base learner with a binary dragonfly algorithm to identify feature combinations minimizing the objective function, finding that gender, adolescent education level, childhood physical abuse, non-acceptance of emotional responses, and paranoid, borderline, and histrionic personality features were associated with increased NSSI risk (Yang et al., 2022). Another study selected 15 factors based on ML and further eliminated them using stepwise regression, ultimately retaining 11 factors with the greatest predictive power (Zhong et al., 2024), while others used permutation importance to rank predictor importance (Su et al., 2023), selecting the top 25 as key predictive factors. However, variable importance ranking methods and retention numbers lack uniform standards across studies, resulting in high heterogeneity in selected key predictors. Table 2 summarizes key variables identified across included literature.

Table 2 Key predictive factors for NSSI identified through machine learning

Demographic variables: Gender (female), age, living with father, education level, admission sources from anywhere other than residence, unsettled housing, White ethnicity, young age and unemployment; ethnicity; socioeconomic status

Psychological variables: Psychological distress, depression, emotional reactivity, self-loathing, anxiety, stress, emotional dyscontrol, tolerance, pain avoidance; maladaptive cognitive emotion regulation strategies; self-efficacy for managing negative emotions; emotion regulation self-efficacy; self-efficacy for expressing positive emotions

Cognitive variables: Abstract thinking problems, rumination, general intelligence below normal group; harmful thoughts

Family variables: Family, family dynamics, family functioning, family conflict and parental depression, help-seeking behavior for parental psychological distress, family reunification; childhood physical abuse; parental attachment

Peer variables: Deviant peer relationships; peer attachment; positive communication

Mental disorder variables: Self-cutting in previous month/self-rated likelihood of future NSSI, self-injurious thoughts, suicidal thoughts and behaviors, death wishes, history of self-harm

Comorbidity variables: Substance use disorders, drug abuse; psychiatric comorbidities, eating disorder patients, borderline personality disorder; psychiatric history, internet addiction, COVID-19-related PTSD, academic anxiety and sleep, psychopathology and medication, psychotherapy; family history of mental illness

Biological variables: NSSI biological phenotype primarily composed of low oxytocin levels, high white blood cell count, and reduced pain sensitivity; NTRK2 gene; excitatory and inhibitory neuron abnormalities; brain disease

Personality variables: Paranoid, borderline, and histrionic personality

School variables: School psychological awareness; school engagement

Behavioral variables: Sugar-sweetened beverage consumption, pandemic-induced lifestyle changes, screen time; poisoning; asphyxiation; behavioral problems

Protective factors: Life satisfaction, positive youth development; physical health

Among NSSI influencing factors, family factors, emotional factors, and mental disorders appear most frequently as key predictive factors, consistent with existing research. Family factors represent the most important predictors in the NSSI ecosystem (Yin et al., 2023), with family-level risk factors showing cross-temporal stability (Wang et al., 2024). NSSI patients exhibit emotion regulation deficits, and using self-injury for emotion regulation and control constitutes one of NSSI's primary purposes (Hasking et al., 2017, 2023; Lang et al., 2024). Mental disorders show high comorbidity rates with NSSI and represent important risk factors.

Furthermore, ML provides new insights for exploring NSSI predictive factors. By partially departing from traditional psychology's hypothesis-then-verify model and incorporating as many relevant variables as possible, ML can identify critical influencing factors beyond existing NSSI theory: for example, increased screen time and sugar-sweetened beverage intake are associated with elevated NSSI risk, while positive youth development can mitigate the adverse effects of family dysfunction and internet addiction on NSSI (Guo et al., 2024). The identification of impaired abstract thinking as a key NSSI factor was also unexpected, possibly because individuals with abstract thinking difficulties also struggle with social interaction and problem-solving, leading to NSSI behavior (Kappes et al., 2023). These findings further expand NSSI theoretical and empirical research, enhancing model predictive power.

4.3 Distinguishing NSSI Categories and Subtypes

Inconsistent NSSI definitions and measurement methods, combined with highly heterogeneous categories and subtypes, constitute important reasons for low model predictive power. ML addresses this heterogeneity dilemma through two approaches: first, identifying common predictive factors across different NSSI definitions, measurement methods, populations, and subtypes; second, effectively distinguishing different NSSI categories and subtypes to establish higher predictive power models within homogeneous groups. Consequently, ML classification in NSSI falls into three major types. The first involves classifying NSSI presence versus absence to identify key predictive factors. One study used ML to classify self-injury presence based on ten variables including psychiatric history, criminal history, psychopathology, and medication, achieving good predictive capacity with 68% accuracy and 71% AUC. Another study used ML to analyze comprehensive clinical data including saliva and blood samples, heart rate variability, and pain sensitivity, modeling biological marker patterns of NSSI and related psychopathology, and found that combinations of hormonal and inflammatory markers plus pain sensitivity could effectively distinguish NSSI disorder presence (Murner-Lavanchy et al., 2024), facilitating longitudinal prediction of clinical outcomes or treatment responses. Current inability to classify NSSI presence versus absence based on neurocognitive features suggests this cognitive domain does not represent a key NSSI characteristic (Murner-Lavanchy et al., 2022).

The second type involves distinguishing different NSSI categories, primarily separating NSSI from other psychological problems or mental disorders, as high comorbidity rates may obscure true influencing factors. One study used NLP to process and systematically train on 721 electronic medical record chat messages (Swaminathan et al., 2023), successfully classifying NSSI versus suicide/homicide ideation and domestic violence crises (AUC = 0.98). Another study found pain avoidance as a unique pre-existing factor distinguishing adolescent suicide from NSSI, though NSSI model performance was worse than for suicide (Bao et al., 2024). Research distinguishing self-injury from violent assault found that deep neural networks and adaptive boosting algorithms provided more accurate predictions (Yin et al., 2021).

The third type uses unsupervised algorithms to classify NSSI internal functions, frequencies, and purposes—that is, investigating NSSI subtypes to establish high predictive power models within homogeneous groups. One study used Latent Class Analysis to identify two subtypes among adolescent depression clinical samples based on NSSI features and severity: “high suicide ideation NSSI group” and “low suicide ideation NSSI group,” finding that being female, single-parent family, left-behind experience, depression, and peer bullying increased probability of belonging to the high suicide ideation group (He et al., 2023). Another study using Latent Profile Analysis identified five NSSI motivation subtypes, finding that individuals whose motivations involved strong desire to communicate with others or punish themselves exhibited more frequent NSSI

behaviors over time (Dixon-Gordon et al., 2022). After analyzing 21 NSSI subtyping studies, researchers found current NSSI subtyping results show obvious heterogeneity with insufficient evidence, but emphasized the important role of NSSI function and psychological pain in subtyping (Wang et al., 2024).

NSSI presence/absence classification uncovers more key predictive factors (Lang et al., 2024; Xu et al., 2024) and establishes models with higher predictive power within complex NSSI data (Lang et al., 2024). Classifying NSSI versus other psychological or mental disorders holds both important research value and high clinical utility, facilitating clinical crisis information classification and triage, saving diagnostic time, and reducing diagnostic errors (Swaminathan et al., 2023; Yin et al., 2021).

High heterogeneity in NSSI subtypes further demonstrates the complexity of NSSI itself. Existing evidence indicates that psychopathology is heterogeneous in many aspects, with different etiological mechanisms potentially related to the same disorder and multiple outcomes possibly existing within the same individual (Feczko et al., 2019). Therefore, further investigating NSSI classification issues not only clarifies NSSI itself but also provides references for classification and identification of other psychopathologies.

In summary, ML holds promise for addressing NSSI measurement difficulties, high complexity, and high heterogeneity, improving NSSI model predictive power, as illustrated in Figure 3 [Figure 3: see original paper].

Figure 3 Applications of machine learning to improve NSSI predictive power

5.1 Further Enriching Data Types to Improve Model Performance

ML-based NSSI models have achieved moderate to good performance, representing substantial improvement over traditional research. However, compared to ML applications in other mental health/psychiatric problems, NSSI-ML research clearly has room for improvement. Difficulties in NSSI measurement lead to relatively small sample sizes and overreliance on single data types like questionnaires or clinical cases, which are key factors affecting ML performance. Future research should focus on expanding multi-type, multi-indicator NSSI models and employing multiple measurement methods for improvement.

Regarding measurement, Ecological Momentary Assessment (EMA) represents the most promising method for NSSI ML applications. Although China faces challenges in collecting case data due to factors like lack of unified medical information management systems (Dong et al., 2020), it holds advantages in coordinated questionnaire surveys. Further integration of questionnaires with EMA technology can assess exposures and outcomes in real-time and authentically, obtaining large datasets. EMA does not reactivate NSSI patients' negative thoughts, and combined with ML analysis can accurately analyze daily self-injury situations and their internal and external triggers (Gee et al., 2020; Martinez-Ales & Keyes, 2019).

Combining multimodal, multi-indicator data also represents an important pathway for improving NSSI model performance. A study on NSSI behavior among prisoners found that models using unstructured datasets far outperformed those using single questionnaire datasets, even surpassing combined comprehensive datasets (Lu et al., 2023). Text data contains more information than structured data, achieving AUC-ROC of 0.862 alone, with sensitivity and specificity reaching 0.816 and 0.738 respectively. However, using only unstructured text also introduces identification risks, as classification of clinical case data depends on physician documentation containing grammatical errors, jargon, and special abbreviations that increase NLP misclassification opportunities (Cliffe et al., 2021). Therefore, combining large-scale multi-indicator, multimodal data is essential for truly leveraging ML advantages and represents the key breakthrough for NSSI model performance.

5.2 Integrating Existing Theories and Methods to Expand In-Depth Research

While ML has made important discoveries regarding high-prediction factors and key features in NSSI, completely ignoring NSSI theory and existing research would divert ML from psychological research, and attempting to exhaust all NSSI influencing factors is impossible (Siddaway et al., 2020). Therefore, as an emerging research method, ML should be further integrated with traditional NSSI theories and research methods to build upon existing findings.

Only two studies employed scientific inclusion and screening methods when using ML to build NSSI prediction models. One study selected 20 relevant predictive factors based on NSSI family systems theory (Zhou et al., 2024). Another adopted a big data perspective, selecting 497 NSSI-related factors from 4,496 total factors for analysis (Su et al., 2023). Future research should consider combining NSSI theory with ML algorithms for variable screening in early stages, selecting strong predictors from large databases, then combining with stepwise regression or structural equation modeling to further improve prediction precision (Burke et al., 2019). For example, Nock's (2009) integrated NSSI model posits that NSSI results from combined biological, psychological, and social influences involving genetics, childhood trauma, negative emotions, negative cognition, stress response, and social competence. ML could first screen important predictors from different domains before model construction. Alternatively, Hooley and Franklin's (2018) NSSI benefits and barriers model could use ML to further clarify beneficial and barrier factors for NSSI behavior and their complex interactions, improving NSSI behavior prediction.

ML should also be deeply integrated with traditional NSSI measurement methods, such as complex data and image data generated by fMRI, fNIRS, EEG, and eye-tracking methods, combined with unsupervised and deep learning for data dimensionality reduction, key feature extraction, and analysis, rather than being limited to structured data and shallow learning.

Furthermore, ML research should learn from limitations of traditional NSSI research, which predominantly uses cross-sectional designs (He et al., 2020) and overemphasizes risk factors while neglecting protective and positive factors (Yang et al., 2023). ML should incorporate more variable types and further improve longitudinal predictive power for NSSI.

5.3 Enhancing Replicability and Comparability to Increase Application Value

Current research shows high heterogeneity in key NSSI predictive factors, stemming from both NSSI's inherent high complexity and ML's use of different participant populations, variable inclusion/exclusion criteria, analytical algorithms, training methods, and evaluation metrics, making comparisons across similar studies difficult (Burke et al., 2019) and limiting clinical diagnostic and intervention assistance applications.

This study proposes following the principle of “similarity before difference, abundance before scarcity” to improve ML replicability and comparability in NSSI research. “Similarity before difference” means first improving model predictive power for homogeneous NSSI groups or subtypes, involving similar NSSI populations such as same-gender, same-age, or same-mental-disorder groups, and similar NSSI types such as same-function, same-frequency, or same-purpose subtypes. Subtypes can first be identified using latent profile analysis, latent class analysis, or other clustering algorithms before investigating groups with larger sample sizes. Because NSSI patients' daily experiences show variability (Kyrön et al., 2021), it is best to control sample variability and shorten follow-up periods to improve NSSI model predictive capacity (Franklin et al., 2017). “Abundance before scarcity” means, after building high-predictive-power models for homogeneous groups with abundant samples, applying transfer learning to different types with smaller samples. Transfer learning transforms learning achieved in data-rich source domains to related target domains with scarce data. Researchers previously pre-trained a deep neural network model for NSSI in a convenient college sample, then fine-tuned it in hospitalized patients, comparing it with models built solely from hospitalized patients. Results showed that deep transfer learning did not improve classification accuracy but could buffer overfitting (Kentopp, 2021). Therefore, if model performance does not change substantially after transfer learning, it suggests identical influencing factors exist across different populations/types. If performance improves or declines significantly, it indicates heterogeneous groups requiring new model construction. This approach merges different NSSI populations within influencing factors or merges identical predictive factors across different populations, ultimately establishing stable classification and high-predictive-power models for homogeneous groups to inform clinical diagnosis and intervention decisions.

6 Research Conclusions and Contributions

This study included 24 studies and used Sankey diagrams to visualize the ML process from sample collection through algorithm selection, model training, and evaluation. We summarized ML applications and advantages for improving NSSI predictive power, finding that ML can streamline questionnaires or enrich other effective research instruments, potentially solving the problem of low predictive power caused by measurement difficulties. Additionally, ML improves predictive power by discovering and identifying key predictive factors, building complex models, and enabling NSSI classification and subtyping. However, model performance and result replicability and comparability require improvement. Therefore, this study proposes the recommendation of “subtyping first, then transfer learning”—first building high-predictive-power ML models using abundant samples from homogeneous types, then applying transfer learning to other populations. This approach both verifies model replicability and accounts for heterogeneity across different groups, ultimately achieving stable and efficient NSSI prediction. Furthermore, ML should be deeply integrated with existing NSSI theories, measurement methods, and analytical approaches, with more in-depth research conducted on this foundation to further improve longitudinal predictive capacity.

References marked with asterisks were included in the systematic analysis

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.