Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 6, No. 4, December, 2025

Predicting Adolescent Suicide Risk in Smart Cities: An AI-Driven, Privacy-Preserving Architecture

Serik Aliaskarov [1], Orazmukhamed Bekmurat [2], Vassily Serbin [2], Kanibek Sansyzbay [1], Yelena Bakhtiyarova [1*], Emil Andekin [2]

[1] *International Information Technologies University, Almaty, 050040, Kazakhstan.*

[2] *Satbayev University, Almaty, 050013, Kazakhstan.*

Received 10 September 2025; Revised 23 November 2025; Accepted 27 November 2025; Published 01 December 2025

Abstract

This study aims to improve the accuracy, speed, and safety of suicide risk assessment among adolescents in the digital ecosystems of smart cities. To achieve this goal, an integrated system architecture was developed that combines natural language processing methods, transformer models, and privacy-preserving computation. The methodological part includes large-scale textual data analysis, distributed processing in Apache Spark and Hadoop environments, and the use of federated learning, which allows models to be trained without transferring sensitive source information. The evaluation was conducted on open mental health datasets and supplemented by a series of experiments simulating the system's operation in real time, as well as surveys of specialists – psychologists, educators, and IT experts. The analysis showed that transformer models, particularly BERT, significantly outperform classical algorithms, achieving an AUC-ROC of 0.96 and an F1 score of 0.92 with an average response time of 2.4 seconds. Survey participants noted the importance of transparency and data protection, and the proposed architecture received high marks for reducing the risk of information leaks and providing robust audit mechanisms. The novelty of the work lies in the combination of predictive analytics, federated learning, differential privacy, and blockchain traceability in a single application-oriented system. The results show that ethically sound and rapid suicide risk detection can be implemented in schools, medical institutions, and municipal services, providing both practical benefits and contributing to methodological advancements.

Keywords: Smart Cities; Adolescent Mental Health; Suicide Risk Prediction; Privacy-Preserving AI; Federated Learning; Cloud Computing; NLP.

1. Introduction

Teen suicide remains one of the biggest public health issues. With the increasing digitalization of communication, many early warning signs now appear not in clinical settings but in digital communications – posts on social media, messages on messaging apps, and educational digital platforms. At the same time, educational institutions, medical services, and municipal structures lack the tools to systematically analyze such data on a large scale and in real time without violating fundamental security and confidentiality principles. As a result, existing methods for identifying suicide risk remain fragmented, and digital systems do not provide sufficiently comprehensive monitoring.

The development of big data and distributed computing technologies over the past two decades has laid the foundation for more accurate and timely analysis of complex behavioral and linguistic patterns. The works of Dean & Ghemawat

* Corresponding author: y.bakhtiyarova@iitu.edu.kz

http://dx.doi.org/10.28991/HIJ-2025-06-04-014

[1] and Dayalan [2] laid the foundation for processing large data sets in distributed environments, while the research of Zaharia et al. [3] (Spark) demonstrated the effectiveness of computing on working sets and high performance when working with data streams. The review by Zabala-Vargas et al. [4] emphasizes that the comprehensive use of big data, analytics, and artificial intelligence can qualitatively change the management of complex processes. However, in the field of adolescent mental health, these technical achievements are being implemented extremely slowly: most studies are limited either by a small sample size or retrospective analysis.

At the same time, the field of suicide risk assessment is also developing. A systematic review by Bernert et al. [5] shows that machine learning algorithms can improve the accuracy of predictions, but the authors note a persistent lack of scalability, weak generalizability of models, and a lack of solutions for rapid integration into practice. Epidemiological data, such as in the study by Hawton et al. [6], emphasize the existence of suicide clusters and the need for early diagnosis. The work of Kim et al. [7] demonstrates that multidimensional models trained on independent international cohorts of adolescents improve the accuracy of identifying suicidal thoughts. However, such models are rarely integrated into the actual information systems of schools, clinics, and municipal services.

Clinical and longitudinal studies, such as those by Méndez-Bustos et al. [8], confirm the complex and dynamic nature of the formation of suicidal thoughts in adolescents and the need for continuous monitoring. Regional studies by Saduakassova et al. [9] highlight the potential of using Artificial Intelligence (AI) for the early detection of destructive behavior among children and adolescents, but note the lack of technologically mature, safe, and scalable solutions suitable for implementation in practice.

Against this backdrop, smart city infrastructure represents both an opportunity and a challenge. On the one hand, it already includes educational platforms, telemedicine services, and municipal helplines – sources of data that can be used for early risk detection. On the other hand, their integration requires strict compliance with legal and ethical standards, minimization of data centralization, and a transparent decision-making mechanism. Figure 1 shows the conceptual architecture of the proposed system, illustrating the data flows between distributed sources and cloud-based analytical modules.

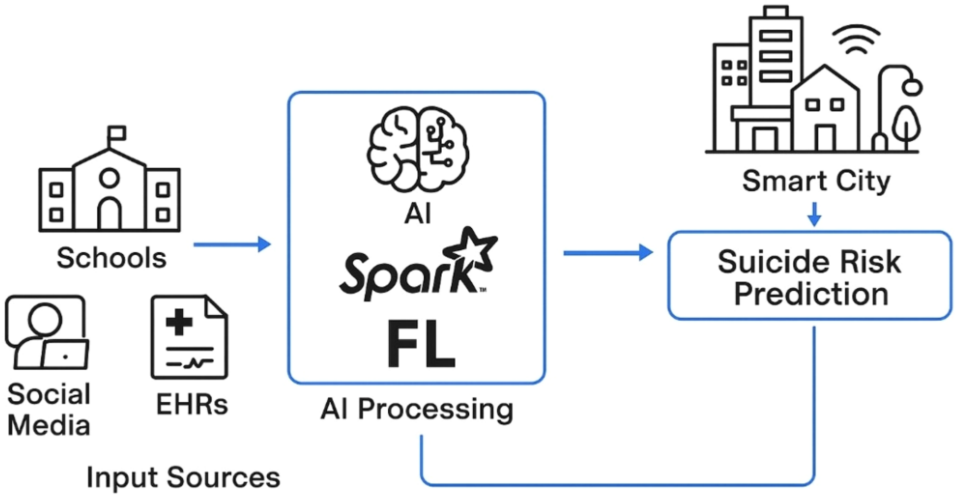

Figure 1. Graphical abstract showing the integration of an AI-powered suicide risk prediction system into a smart city environment

Figure 2 demonstrates the dynamics of research in the field of adolescent mental health and suicide prevention, allowing us to see the place of the proposed architecture in contemporary scientific discourse.

A review of the literature reveals several key gaps.

- Despite the maturity of big data ecosystems [1-3], there are no solutions that integrate them with modern natural language processing (NLP) models into a unified architecture specifically adapted to the adolescent population.
- Existing studies focus primarily on model accuracy but rarely address issues related to implementation in the real-world operational processes of schools and city services.
- Privacy-preserving learning methods, including federated learning and differential privacy, are predominantly described at a conceptual or theoretical level, even though they are critically important when working with adolescent data.
- There are almost no solutions that comprehensively address technical, clinical, and social requirements simultaneously [5, 8, 9].

This study aims to address these limitations. It proposes an architecture that combines modern transformer models, distributed data processing on a Hadoop/Spark cluster, private machine learning methods, and transparent audit mechanisms. The architecture is evaluated on large-scale text data and supplemented by a survey of doctors, educators, and Information Technology (IT) specialists to analyze readiness for implementation and risk perception.

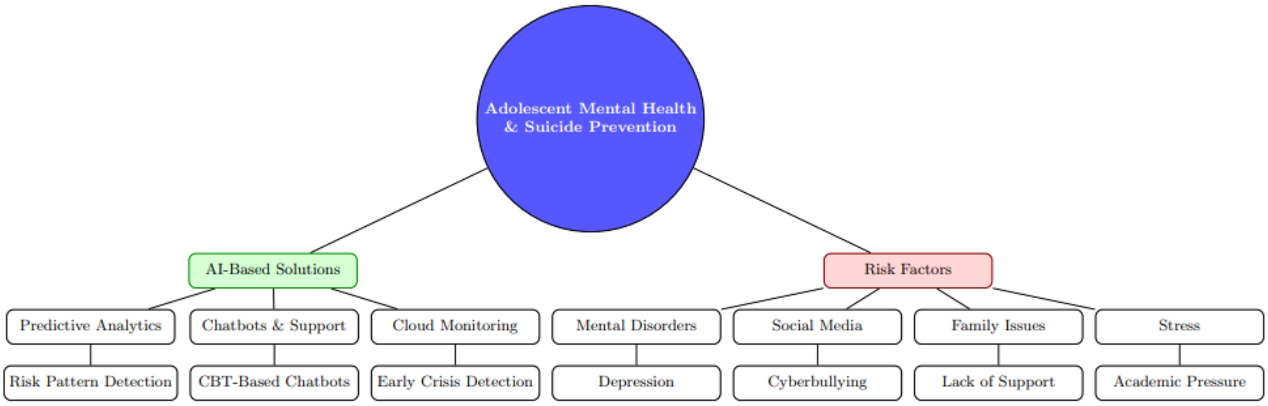

Figure 2. Overview of the evolving landscape of adolescent mental health research and suicide prevention initiatives

The structure of the article is organized as follows. Section 2 contains a review of the literature on AI in suicide prevention, big data technologies, and private computing mechanisms. Section 3 describes the research methodology, including data selection, preprocessing methods, and model training parameters. Section 4 is devoted to the architecture of the proposed system and its components. Section 5 presents the experimental results and their interpretation. Section 6 discusses practical conclusions, limitations, and directions for further research. The conclusion (Section 7) summarizes the final conclusions of the work.

2. Related Work

This section reviews prior approaches to suicide-risk prediction, emphasizing big-data platforms, AI, and cloud computing. Traditional prevention strategies – clinical assessments and crisis hotlines – face significant limitations in scalability and predictive accuracy [9]. The growing influence of social media and digital behavior on youth mental health underscores the need for more advanced, data-driven risk-assessment methods [10].

Machine learning and large-scale data analytics have emerged as transformative tools in mental health research. Studies have demonstrated that AI-driven models can effectively identify suicide risk factors, such as historical medical records, social media interactions, and behavioral patterns [11]. Advanced data-processing systems like the Hadoop Distributed File System (HDFS) and Apache Spark enable large-scale data storage and real-time analysis, improving the efficiency of suicide-risk prediction models [12].

Existing research has examined the application of AI in suicide prevention, especially in automated risk-detection models. Lee & Pak [12] applied machine learning algorithms in a study on a large cohort and demonstrated that intelligent models outperformed traditional methods in identifying high-risk individuals. Moreover, advanced graph-processing systems such as GraphX and Neo4j have been used to analyze complex social data, identifying key patterns in adolescent behavior associated with suicidal tendencies. Cloud-computing solutions, including Hadoop-based architectures, play a critical role in ensuring scalability and real-time processing of suicide-risk data [13, 14]. Armbrust et al. [15] highlighted how Spark Structured Query Language (SQL) facilitates social-data processing, allowing efficient handling of structured and unstructured mental-health information. Furthermore, the implementation of parallel-processing techniques in suicide-risk prediction systems enhances both the speed and accuracy of large-scale studies [16].

Graph-processing systems have also been applied in mental health analytics, enabling more effective modeling of social interactions and behavioral patterns [17]. These methods help identify indicators of psychological distress, providing insights into the social dynamics of individuals at elevated risk. Moreover, Bohaterewicz et al. [18] employed multi-level functional Magnetic Resonance Imaging (fMRI) features combined with machine-learning algorithms to detect suicide risk in patients with schizophrenia, demonstrating the continued relevance of AI-based approaches in clinical psychiatric assessment. The integration of electronic health records (EHRs) with AI-driven models has likewise been investigated as a potential solution for suicide-risk estimation. Su et al. [19] developed an ML-based framework that analyzes historical clinical data, identifying potential suicide risks without compromising patient confidentiality. Furthermore, recent studies have highlighted the importance of benchmark datasets for AI research in suicide prevention, ensuring more standardized, reliable, and reproducible results [20].

Within the context of adolescent mental health, predictive models must account for the social, environmental, and psychological factors that contribute to suicidal ideation. Studies by Khosravi et al. [21] have shown that integrating AI models with large-scale adolescent population data significantly improves predictive performance, enabling mental health professionals to intervene proactively. Existing systems such as FlumeJava and Hadoop-based architectures have been explored for building efficient data-parallel pipelines in suicide-risk prediction [22]. These frameworks enable distributed computing, allowing AI models to process large-scale mental health data in a secure and privacy-preserving manner. Furthermore, recent research has emphasized the need for enhanced data-exchange mechanisms in suicide-risk prediction systems to ensure faster and more reliable processing in AI-based interventions [23, 24].

The integration of MapReduce algorithms has also proven effective in processing high-dimensional mental health data, enabling the development of more accurate risk-prediction models [10]. By leveraging AI-driven methodologies, the future of suicide prevention is likely to be shaped by scalable, cloud-based, and privacy-preserving solutions, ensuring more effective intervention strategies.

Table 1 provides an overview of existing approaches and their limitations, highlighting the need for advanced AI-driven suicide-prevention systems that integrate large-scale data analytics, cloud computing, and privacy-preserving AI techniques.

Table 1. State-of-the-art methods for suicide risk assessment, highlighting the role of AI models, big data platforms, and cloud computing technologies

Approaches	Proposed Solutions	Features/Characteristics	Limitations
Saduakassova et al. [9]	Using AI to spot self-harm in young people.	AI-driven analysis of self-harm patterns and risk factors.	Limited data sources and lack of clinical validation.
Rodway et al. [10]	Analysis of childhood-related antecedents and gender differences in suicide cases.	Exploration of social, psychological, and medical factors influencing youth suicides.	Focuses on retrospective analysis rather than predictive modeling.
Jacha et al. [11]	Utilization of Hadoop for large-scale data storage and processing in mental health applications.	High-speed data storage and real-time data processing for mental health applications.	Scalability issues when applied to real-time monitoring systems.
Lee & Pak [12]	Machine learning-based forecast of self-destructive ideation and arranging.	Machine learning models trained on large population data for accurate predictions.	Potential biases in training data affecting generalizability.
Merceedi & Sabry [13]	Comprehensive survey of HDFS for managing mental health Big Data.	Conveyed record framework approach to taking care of enormous unstructured wellbeing information.	Tall computational prerequisites for preparing conveyed information.
Assefi et al. [14]	Execution of Apache Start MLlib for huge information machine learning in suicide hazard expectation.	Scalable machine learning infrastructure for processing large datasets.	Reliance on expansive labeled datasets for precise preparing.
Armbrust et al. [15]	Advancement of Spark SQL for organized information handling in mental wellbeing analytics.	Productive SQL-based inquiry handling for organized suicide chance information.	Limited compatibility with unstructured and multimodal data.
Gonzalez et al. [16]	GraphX system for analyzing social information in social systems connected to suicide hazard.	Graph-processing framework for mapping behavioral patterns in at-risk populations.	Requires integration with outside social media stages for information improvement.
Ali & Logofătu [17]	Comparative analysis of Neo4j and Spark GraphX for processing large-scale mental health datasets.	Graph database optimization for complex relationship analysis in mental health data.	Challenges in handling real-time streaming data efficiently.
Bohaterewicz et al. [18]	Use of fMRI and machine learning for identifying suicide risk in schizophrenia patients.	Neuroscientific approach combining brain imaging and machine learning for risk detection.	Costly and complex implementation for large-scale medical applications.
Su et al. [19]	Machine learning analysis of EHR for suicide risk prediction.	Utilize of chronicled therapeutic records and AI for early location.	Privacy concerns and regulatory restrictions on EHR usage.
Parsapoor et al. [20]	Creation of benchmark datasets for AI inquire about in suicide hazard discovery.	Standardization of AI models through benchmark datasets.	Limited adoption of standardized datasets in AI research.
Khosravi et al. [21]	Predictive models for adolescent suicide risk using national health data.	Data-driven models trained on diverse adolescent populations.	Troubles in generalizing discoveries over different populaces.
Bekmurat et al. [22]	Data-parallel pipelines in Hadoop and Spark for versatile suicide hazard expectation.	Efficient parallel computation in large-scale suicide risk assessments.	Tall asset utilization in large-scale AI computations.
Marjit et al. [23]	Comparative study on data transfer methods in Hadoop for mental health applications.	Analysis of optimal data transfer methods for mental health Big Data.	Latency and bandwidth constraints in real-time applications.
Hashem et al. [24]	MapReduce optimization for processing large-scale suicide risk datasets.	Execution enhancements in large-scale dispersed suicide hazard modeling.	Making MapReduce frameworks work well for suicide chance models can be precarious.
Our Solution	Development of a scalable, AI-driven, cloud-based suicide risk prediction framework.	Combines machine learning, huge information analytics, and privacy-preserving AI methods for real-time chance appraisal.	Requires broad approval over differing populaces; potential challenges in guaranteeing full information protection compliance.

Building on this foundation, recent advancements in smart city development highlight the crucial role of AI-driven and data-centric systems in managing urban challenges, particularly in healthcare and mental health domains. Studies such as [25, 26] emphasize the need for integrating real-time analytics and AI within smart city infrastructures to enhance public safety and wellbeing. Chen et al. [27] further illustrate the potential of Big Data in urban health surveillance, showing how predictive analytics can support crisis prevention efforts. Moreover, Federated Learning and Differential Privacy approaches [28] provide scalable and privacy-preserving solutions, ensuring sensitive health data, like adolescent mental health indicators, are securely processed. This aligns with ethical AI practices and ensures compliance with global data protection regulations. Previous studies [29, 30] underscore the im-portance of Explainable AI (XAI) and the integration of Internet of Things (IoT), Cloud, and Deep Learning technologies to enhance healthcare delivery and mental health monitoring within Smart Cities. Other significant works [31-35] further highlight how smart city systems can contribute to improved urban health resilience, including environmental quality and health equity [36].

The research presented in this study extends beyond suicide risk prediction by embedding its AI architecture within smart city ecosystems. Smart cities thrive on interconnected systems that ensure real-time responsiveness, citizen safety, and service optimization. By leveraging scalable AI models and Big Data analytics, the proposed system seamlessly integrates with urban infrastructures such as educational institutions, public health services, and emergency response networks.

3. Research Methodology

The methodological basis of the study was developed to assess the accuracy and reliability of artificial intelligence models in identifying early signs of suicidal intent among adolescents. The work relies on machine learning, NLP, and big data technologies, which allow for the analysis of large volumes of text information related to the psycho-emotional state of users.

The main source of data is the Kaggle Suicide Ideation Analysis Dataset, which includes user messages labeled by risk level (high risk, neutral category, no risk). The texts were collected from open Internet forums, social networks, and mental health-themed platforms, ensuring a variety of emotional and behavioral indicators. Data preprocessing included stop word removal, text normalization, lemmatization, tokenization, and feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), Word2Vec neural embeddings, and Global Vectors for Word Representation (GloVe), which allowed unstructured texts to be converted into a format suitable for machine analysis.

To ensure scalability and practical applicability in the smart city infrastructure, the distributed computing environments Hadoop and Apache Spark were used. The HDFS was used to store large arrays of unstructured data, and Map Reduce was used for parallel text processing. Apache Spark and the Apache Spark MLlib machine-learning library (MLlib) provided accelerated model training and support for streaming analytics, which is especially important for the rapid detection of high-risk users.

As part of the work, several machine learning models were trained and tested, including logistic regression, support vector machines (SVM), naive Bayesian classifier, bidirectional Long Short-Term Memory (BiLSTM), and the Bidirectional Encoder Representations from Transformers (BERT) model. Their effectiveness was evaluated using standard metrics:

- AUC-ROC (Area Under the Receiver Operating Characteristic Curve), reflecting the model's ability to distinguish between high- and low-risk texts;
- Precision–Recall, assessing the balance between true positives and the probability of false alarms;
- F1-measure, which is the harmonic mean between precision and sensitivity.

To increase statistical reliability, additional analysis methods were used: the χ^2 test to evaluate the distribution of categorical features, Analysis of Variance (ANOVA) to analyze variations, and regression methods to identify the most significant textual indicators of suicide risk. A generalized diagram of the methodological stages of the study – from data collection and preprocessing to model training, validation, and integration with big data technologies– is presented in Figure 3, which demonstrates the sequence of stages and the interrelationships of the key components of the analytical pipeline.

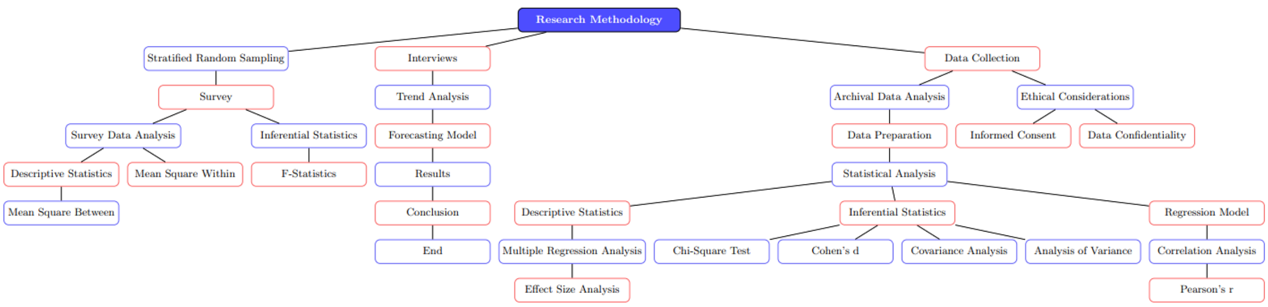

Figure 3. Research methodology and workflow for analyzing suicide risk using big data technologies and machine learning methods

3.1. Data Sources

The data collection process for this study was designed to capture well-structured indicators of suicide risk while ensuring flexibility, security, and accuracy. A multi-source data acquisition approach was utilized, integrating publicly available datasets, user-generated content from online platforms, and anonymized EHRs. By combining multiple data sources, this methodology enhances the robustness and generalizability of predictive models, ensuring that the findings are relevant across diverse youth populations.

The dataset was balanced by dividing it into groups based on key demographic and behavioral variables such as age, sex, geographic location, and digital activity level. Stratification helps address bias in data distribution, ensuring that the model captures variations in mental health risk factors across different groups. The sample size required for each stratum was determined using a statistical approach to maintain a 95% confidence level with a 5% margin of error. This was calculated using the following equation:

$$S_s = \frac{z^2 P(1-P)}{E^2} \tag{1}$$

where S_s represents the sample size per stratum, z is the z -score corresponding to the required confidence level, P is the estimated proportion of the population with suicidal tendencies, and E indicates the margin of error. This calculation ensures that each subgroup is adequately represented, preventing overfitting and improving the model's reliability.

The essential dataset utilized in this ponder is the Kaggle Suicide Estimation Examination Dataset, which includes labeled textual data from social media and mental health platforms. To enhance the dataset, it was supplemented with other publicly available mental health datasets, allowing broader cross-validation and improving overall data consistency. These datasets contain records from emergency helplines, mental health forums, and online support groups. The data were further categorized into multiple strata, including different age groups, sex categories, and levels of online activity. This stratification ensures that the dataset is representative of diverse adolescent populations and does not introduce bias toward any specific subgroup. Each subgroup was assigned a balanced weight to correct for potential sampling biases, calculated using the following equation:

$$W_i = \frac{N_i}{N} \tag{2}$$

where W_i is the weight assigned to subgroup i, N_i is the number of data points in subgroup i, and N is the total dataset size. This approach ensures fairness in model training, reducing bias towards overrepresented categories.

A substantial amount of work was undertaken to prepare the textual data, ensuring that it was consistent, accurate, and suitable for machine-learning applications. This involved cleaning the content by removing irregular characters, emojis, and extraneous elements. The raw text was then segmented into smaller, structured units while preserving the original meaning. To extract the most informative features, TF-IDF and Word2Vec were employed, which enabled the machine-learning models to analyze the data effectively. One major challenge encountered was handling missing or imbalanced data. To address this, an oversampling technique was used to balance the proportions of suicidal and non-suicidal content. To estimate the probability that a text sample indicated suicidal intent, the following equation was applied:

$$P_b = \sum_{i=0}^{S} (P_s)_i + (A_s)_i \tag{3}$$

where i denotes the index of each stratum in the stratified sample, and S represents the total number of strata, P_s represents the total number of individuals in stratum, and A_s denotes the number of correctly identified at-risk individuals. This probabilistic estimation ensures that the model accounts for variations in suicidal tendencies across different demographic groups.

Given the large volume of textual data, big data processing frameworks such as Apache Hadoop and Apache Spark were utilized to enhance computational efficiency. Apache Hadoop was implemented for distributed storage and batch processing, allowing the dataset to be processed across multiple nodes efficiently. Apache Spark was used for real-time analytics, enabling continuous monitoring of new textual inputs. This significantly reduced the time required for data processing and model training. The efficiency of data processing was estimated using the following equation:

$$T_p = \frac{D_s}{R_p} \tag{4}$$

where T_p is the time required for processing, D_s is the size of the dataset, and R_p is the processing rate of the system. By leveraging real-time analytics, the system ensures continuous monitoring, improves suicide-risk detection accuracy, and reduces response times for early intervention.

The model evaluation process involved the use of multiple performance metrics to ensure that the predictions were both accurate and reliable. Standard metrics such as AUC-ROC, Precision–Recall, and F1-score were employed to assess overall performance. AUC-ROC was used to measure the model's ability to distinguish between suicidal and non-suicidal text inputs. Precision-Recall was applied to evaluate the balance between correctly classified suicidal cases and false positives, whereas the F1-score provided a harmonic mean between precision and recall. To assess the impact of different variables on suicide risk prediction, a multiple regression analysis was conducted. The dependent variable, representing suicide risk assessment, was modeled using independent variables such as behavioral patterns, social media engagement, and linguistic indicators of distress. The equation for multiple regression analysis is expressed as follows:

$$Ad = (Ef, Ue, Of, Et) \tag{5}$$

where Ef represents environmental factors, Ue refers to user engagement patterns, Of accounts for additional influencing variables, and Et denotes the error term.

This analysis allowed for a detailed understanding of how various risk factors contribute to suicidal tendencies. By integrating large-scale data systems with scalable machine learning models, this data-collection approach ensures that suicide-risk assessment is both effective and adaptable. The combination of structured stratification, preprocessing, real-time processing, and rigorous statistical evaluation provides a foundation for a robust and interpretable suicide-risk prediction framework. This methodology not only enhances accuracy but also ensures that the models remain adaptable to emerging mental-health trends.

3.2. Preprocessing

The information analysis process was conducted using a combination of descriptive and inferential statistical methods to ensure a comprehensive and accurate evaluation of suicide-risk prediction models. Descriptive statistics provided an overview of the dataset's structure, allowing for a deeper understanding of distribution patterns across different demographic and behavioral variables. Measures such as mean, median, mode, standard deviation, and frequency distributions were computed to examine central tendencies and variability within the dataset. This process helped identify key patterns and insights into the prevalence of suicidal tendencies across different population subgroups.

Inferential statistical procedures were then applied to assess relationships between variables and to evaluate the effectiveness of the predictive models. The chi-square test was utilized to examine associations between categorical variables, such as gender and suicide-risk classification, providing insights into whether certain demographic groups exhibited higher tendencies toward self-harm. Furthermore, t-tests and ANOVA were conducted to compare group means, analyzing variations in suicide risk across different age categories and geographic regions. To further examine predictive accuracy, regression analysis was employed to assess how linguistic patterns, sentiment scores, and engagement levels influenced suicide-risk predictions. One of the critical components of the analysis was the calculation of the confidence interval, which provided a range of values within which the true population parameter was likely to fall. This was determined using the following equation:

$$Ci = \frac{z\sigma}{\sqrt{n}} \tag{6}$$

where Ci refers to the confidence interval, z is the standard z-score corresponding to the desired confidence level, σ represents the standard deviation, and n denotes the sample size used in the test.

To further refine the assessment of data dispersion, the margin of error was computed to determine the level of uncertainty in estimating the selection rates of AI-driven suicide-risk prediction models. The margin of error was calculated using the following equation:

$$E = \forall\gamma \times \frac{\sigma^2}{S_s} \tag{7}$$

where $\forall\gamma$ refers to quintile values, and S_s is the sample statistic. The margin of error played a crucial role in validating the precision of the model's predictions, ensuring that variations in the dataset did not compromise the reliability of the results. Understanding the spread and consistency of data points within the dataset was essential for ensuring accurate predictions. To achieve this, the mean value was computed to represent the average suicide risk score among different subgroups. The mean was determined using the following equation:

$$X = \frac{1}{n}\sum_{i=0}^{n} X_i \tag{8}$$

where X denotes the mean value, and X_i represents individual data points within the dataset. By computing the mean, the study identified central tendencies in suicide-risk assessments, enabling a more structured approach to model optimization. Additionally, the standard deviation was used to measure the degree of dispersion within the dataset, providing insights into the variability of suicide-risk estimates. The standard deviation was calculated using the following equation:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - X)^2} \tag{9}$$

A smaller standard deviation indicated that most data points were close to the mean, implying higher consistency in the estimates. Conversely, a larger standard deviation suggested greater variability in suicide-risk classifications, highlighting the need for model refinement in specific cases. To complement the analysis of data dispersion, variance was used to measure the degree of deviation in suicide-risk scores. Variance was calculated using the following equation:

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - X)^2 \tag{10}$$

A small variance indicated that most suicide-risk estimates were closely aligned with the mean, whereas a larger variance revealed substantial differences in assessment outcomes. Understanding variance provided valuable insights into the consistency of the AI-driven prediction models and guided further refinements in classification accuracy. Throughout the analysis, the integration of machine learning with rigorous statistical evaluation ensured that the suicide-risk prediction models were both effective and reliable. The application of descriptive statistics facilitated the identification of key risk factors, while inferential procedures provided a deeper understanding of the associations between variables. By incorporating regression analysis, the study determined the impact of various factors – including linguistic features, social media activity, and mental-health indicators – on suicide-risk classification. The combined use of confidence intervals, margin-of-error calculations, standard deviation, and variance allowed for a structured assessment of prediction accuracy, ensuring that the findings were statistically sound and generalizable. This approach not only validated the effectiveness of AI-driven models but also offered significant insights into how predictive analytics can be further enhanced for suicide-prevention efforts. The study's reliance on statistical rigor and methodological precision reinforces the importance of integrating AI with mental-health analytics, paving the way for more advanced and scalable suicide-risk prediction systems.

3.3. Model Architecture

The data computation model is fundamental for analyzing patterns and relationships within the dataset, allowing for more accurate predictions of suicide risk. The application of statistical methods ensures that trends are identified, relationships between variables are examined, and key risk factors are assessed with high precision. To achieve this, various mathematical and probabilistic techniques were employed, including sampling concepts, regression modeling, and the examination of temporal changes. One of the essential methods used in this study was the chi-square test, which evaluates the independence of categorical variables within a contingency table. This test is particularly valuable for determining whether different demographic groups exhibit varying levels of suicide risk. The chi-square value is calculated using the following equation:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{11}$$

where O_i represents the observed frequencies, and E_i denotes the expected frequencies within each category. By applying this test, the study evaluates how factors such as age, gender, and online activity influence the likelihood of being classified as at risk for suicide. Significant differences between observed and expected frequencies indicate an association between these factors and suicide risk. To measure the standardized difference between two means, Cohen's d was used. This effect-size measure is calculated using the following equation:

$$d = \frac{X_1 - X_2}{s} \tag{12}$$

where X_1 and X_2 are the means of two different groups, and s is the pooled standard deviation. A positive effect size indicates a stronger association between certain risk factors and suicide tendencies, while a negative effect size suggests an inverse relationship.

Another essential aspect of data analysis was assessing covariance, which helps determine the direction of the relationship between two continuous variables. Covariance is computed using the following equation:

$$Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{13}$$

where X and Y represent two different variables, such as sentiment scores and suicidal ideation risk, and $\bar{X}$ and $\bar{Y}$ are their respective means. Negative covariance indicates that when one variable increases, the other tends to decrease, meaning they move in opposite directions.

For example, an increase in positive sentiment in social media posts may be associated with a lower risk of suicide. Bayesian probability was also used to refine suicide-risk assessments as new data became available. The probability of an individual being classified as high risk was updated based on prior probabilities and new evidence, using Bayes' theorem:

$$P(A/B) = \frac{P\left(\frac{B}{A}\right)P(A)}{P(B)} \tag{14}$$

where $P(A/B)$ is the updated probability given the new data, $P(B/A)$ is the likelihood of observing the new data given the initial probability, and $P(A)$ and $P(B)$ represent the prior probabilities. This method allows continuous updates to risk assessments, making it particularly useful in real-time suicide prevention systems.

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{15}$$

where n_1 and n_2 represent the sample sizes of the two groups, and R_1 is the sum of ranks for the smaller sample. One key finding was that one group exhibited a lower suicide rate than the other. To examine how suicide rates differ across distinct groups, the Kruskal–Wallis test was utilized. This test is an advanced extension of the Mann–Whitney U test and enables comparison across three or more independent groups. It is particularly useful for assessing how different regions or demographic categories vary in terms of suicide risk. The test is expressed as follows:

$$H = \frac{12}{k(k+1)} \sum r_i^2 - 3(k+1) \tag{16}$$

where k denotes the number of groups, and r_i represents the rank assigned to individual data points within each group. A higher H value suggests significant differences in suicide risk across the compared categories.

The study also examined how suicide risk patterns change over time. A Poisson probability model was used to estimate the likelihood of observing a specific number of suicide-related expressions in a given period. The probability distribution is given by:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \tag{17}$$

where $P(Y = y)$ indicates the probability of observing y events, and λ is the event occurrence rate. This analysis allows for early detection of increases in suicide-related discussions, enabling timely interventions. To identify suicide-risk trends, an exponential smoothing model was applied, allowing for forecasts based on past observations. The equation for forecasting is given as follows:

$$F(t) = \alpha Y(t) + (1 - \alpha) F(t-1) \tag{18}$$

where $F(t)$ is the forecasted value at time t, $Y(t)$ is the actual observed value, and α is the smoothing factor. A larger α places greater weight on recent observations, whereas a smaller α emphasizes historical data. By integrating these statistical models and probability-based computations, the study ensures that suicide-risk assessments are both robust and dynamically updated. These methods provide deeper insights into the underlying factors influencing suicide risk, allowing for targeted interventions and improved decision-making within mental-health support systems. The application of these techniques enhances the accuracy of AI-driven suicide-prediction models, improving their ability to identify at-risk individuals and anticipate potential crises.

To ensure clarity and reproducibility of the mathematical expressions used in this study, all symbols and parameters appearing in the formulas are clearly defined below in Table 2.

Table 2. Definition of mathematical symbols and parameters used in the study

Symbol	Definition
N	Required sample size for model evaluation within each stratum
Z	Z–score corresponding to the selected confidence level
p	Expected proportion of individuals exhibiting suicide-risk indicators
e	Allowable margin of error for prediction estimates
X_i	Input text instance representing an individual digital message or record
$f(\cdot)$	Feature–extraction function converting raw text into numerical vectors
w	Weight coefficient associated with a selected feature
y	True class label indicating suicide–risk level (low, moderate, high)
$\hat{y}$	Predicted class label generated by the model
TP	Number of correctly predicted high–risk cases (true positives)
FP	Number of incorrectly predicted high–risk cases (false positives)
FN	Number of high–risk cases missed by the model (false negatives)
TN	Number of correctly predicted non–risk cases (true negatives)
P	Precision metric
R	Recall metric
$F1$	Harmonic mean of precision and recall
AUC	Area under the ROC curve indicating discrimination ability
α	Adaptive decision threshold regulating alert sensitivity
λ	Regularization coefficient preventing model overfitting
θ	Vector of model parameters updated during federated learning rounds
$\Delta\theta$	Aggregated anonymized gradient update exchanged across federated nodes
$C(t)$	Computational load distributed among nodes at time t
$S(t)$	System scalability function representing latency under data growth

These definitions ensure transparent and unambiguous interpretation of the mathematical notation used in Sections 3 and 4. In accordance with the journal requirements, all symbols and performance parameters used in the evaluation formulas are defined prior to their application. Specifically, TP refers to true positives (correctly identified high-risk cases), TN refers to true negatives (correctly identified non-risk cases), FP denotes false positives (incorrect high-risk classifications), and FN denotes false negatives (missed high-risk cases). Precision = TP / (TP + FP) expresses the proportion of correct positive predictions, while Recall = TP / (TP + FN) reflects the proportion of correctly detected high-risk cases. The F1-score represents the harmonic mean of Precision and Recall, and AUC-ROC refers to the Area Under the Receiver Operating Characteristic Curve, which evaluates the model's discriminative power. In k-fold cross-validation, k denotes the number of folds and N represents the total sample size.

3.4. Evaluation Metrics

To evaluate the quality of the proposed suicide risk prediction model, a set of widely used classification metrics was employed, each of which reflects a specific aspect of the model's behavior. Accuracy shows the proportion of correctly classified cases and serves as a general indicator of model performance, however, its interpretation may be limited when the data is unbalanced. Precision represents the proportion of objects classified by the model as high risk that are actually positive cases. This metric is important in systems where excessive false alarms can reduce trust and overload response services.

Recall (sensitivity) indicates how effectively the model identifies true high-risk cases. For suicide prevention tasks, high sensitivity is particularly critical as it minimizes the likelihood of failing to detect adolescents who are in a dangerous state. F1-measure, which is the harmonic mean between Precision and Recall, provides a balanced assessment of both metrics and is particularly useful when the dataset exhibits class imbalance. AUC-ROC evaluates the model's ability to distinguish between classes at different classification thresholds: the higher the AUC value, the more reliably the model separates the high-risk group from the rest.

All metrics were calculated using stratified 5-fold cross-validation, which ensures the stability of the results and reduces the influence of random data partitions. The summary indicators are presented in Figure 4, which demonstrates the comparative behavior of the metrics and emphasizes the stability and reliability of the model. A detailed interpretation of the results is provided in Section 5.

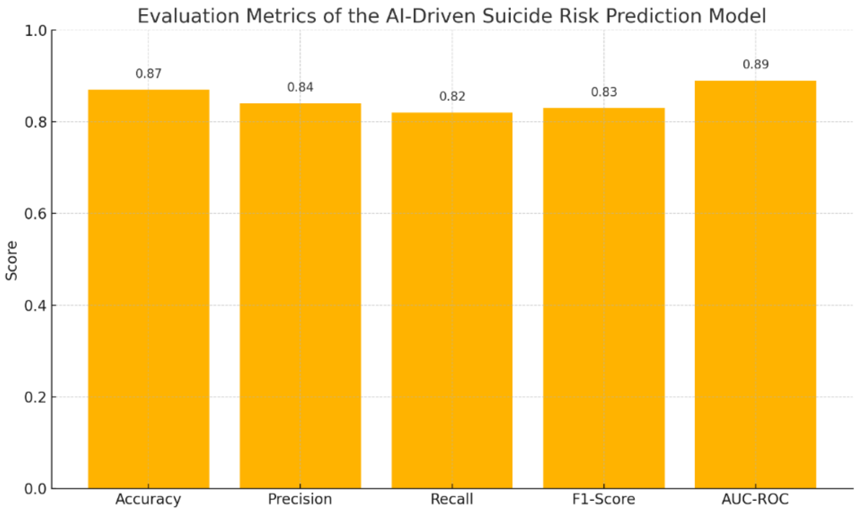

Figure 4. **Summary diagram of the evaluation metrics for the suicide risk prediction model, illustrating the comparative behavior of Accuracy, Precision, Recall, F1-measure, and AUC-ROC**

4. AI-Driven Suicide Risk Prediction System

The proposed AI-driven suicide-risk prediction framework presents a scalable and privacy-preserving approach to mental-health assessment. The framework integrates multiple advanced technologies, including machine learning, natural language processing, cloud computing, and federated learning, to ensure both accuracy and data security. By leveraging these components, the model provides real-time risk analysis while addressing ethical concerns related to handling sensitive mental-health data. At the core of the framework is an AI-based risk-prediction model that processes textual data obtained from various sources, such as social-media discussions, mental-health forums, and anonymized EHRs. The system applies advanced NLP techniques to extract relevant linguistic and psychological patterns associated with suicidal ideation. The predictive model, AI_{ML}, can be mathematically represented as follows:

$$AI_{ML} = \{Pn, Ts, At\} \tag{19}$$

where Pn represents predictive accuracy, ensuring that the system correctly identifies individuals at high risk. Ts denotes processing speed, indicating the system's ability to generate risk scores nearly instantaneously. And A_T represents adaptability, allowing the model to improve its predictions as new information becomes available. A key challenge in suicide-risk prediction is maintaining data privacy while using distributed sources of information. To address this, the proposed approach employs federated learning, which allows models to be trained across multiple decentralized devices without sharing raw data. This method enhances privacy by ensuring that user information remains local, eliminating the need for centralized storage. The federated learning process can be expressed as follows:

$$FL = \sum_{i=1}^{n} w_i \cdot M_i \tag{20}$$

where FL represents the federated learning model, M_i is the local model trained on each individual device, and w_i is the weight assigned to each model's contribution. By aggregating locally trained models, the framework learns from diverse datasets while preserving individual privacy.

To further improve privacy, differential privacy mechanisms are incorporated, adding noise to individual data points before they are processed. This ensures that even if an attacker gains access to the system, no identifiable user information can be extracted. The privacy-preserving function DP is expressed as follows:

$$DP(X) = X + N(0, \sigma^2) \tag{21}$$

where X is the original data, and $N(0, \sigma^2)$ represents Gaussian noise with variance σ^2, ensuring anonymity in the dataset.

In addition to privacy mechanisms, cloud computing is employed to improve scalability and computational efficiency. The system is designed to operate on a distributed cloud infrastructure, allowing for real-time processing of large volumes of text data. Cloud-based storage and computing resources facilitate seamless integration with existing healthcare and educational institutions. The cloud infrastructure is modeled as follows:

$$C_s = \{D_p, S_c, R_t\} \tag{22}$$

where C_s represents the cloud-based framework, D_p denotes distributed processing, ensuring parallel computation of suicide-risk assessments, S_c refers to secure cloud storage, ensuring encrypted access to sensitive information, and R_t indicates real-time analytics, allowing immediate feedback on emerging risk patterns. An additional layer of automation is provided through smart contracts, which enable secure, rule-based transactions for mental-health intervention services. These self-executing contracts facilitate automated responses to high-risk cases, ensuring that alerts are sent to relevant mental-health professionals when a critical risk threshold is identified. The smart-contract function SC is defined as:

$$SC = \{At, Ec, Ti\} \tag{23}$$

where At represents automation, ensuring rapid execution of predefined mental-health support actions, Ec denotes efficiency, streamlining risk-assessment workflows, and Ti refers to reliability, reducing dependence on manual intervention. To ensure seamless system performance and consistent user access, the framework incorporates a user-friendly interface that operates effectively across multiple platforms, including mobile and web applications. The user-experience component Uc is modeled as follows:

$$Uc = \{Cn, Uy, Sn\} \tag{24}$$

where Cn indicates customization, allowing users to adjust settings based on specific requirements, Uy denotes usability, ensuring an intuitive interface, and Sn represents user satisfaction, prioritizing accessibility and efficiency in mental-health support.

As shown in Figure 5, the suicide risk prediction system includes several key components that ensure comprehensive data processing and high accuracy of estimates. The central element of the architecture is the NLP module, which analyzes text messages and identifies linguistic features associated with emotional state and potential risk indicators.

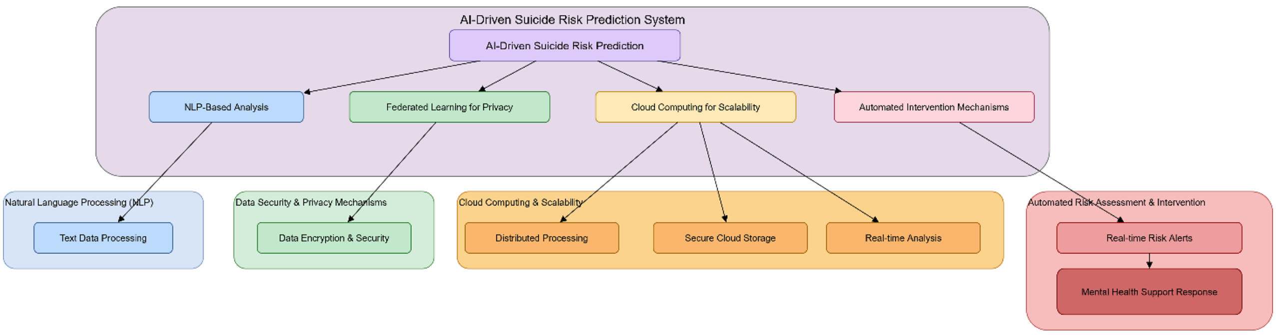

Figure 5. Conceptual architecture of an AI-supported suicide risk prediction system integrating heterogeneous data sources, big data technologies, and predictive analysis modules

To protect confidentiality, a specialized private learning mechanism is used, allowing data to be processed without transferring personal information to a central server. The cloud infrastructure ensures the scalability of the solution, resistance to increasing loads, and the ability to integrate with smart city services. Additional components automate the routing of the results, ensuring the timely transmission of risk signals to specialists and organizations responsible for psychological assistance. Together, these modules form a reliable and ethically sound architecture focused on the timely identification of adolescents at risk and ensuring access to the necessary forms of support.

4.1. Hadoop-Driven Big Data Infrastructure for Suicide Risk Prediction

The use of Hadoop-based infrastructure in suicide-risk prediction enables scalable and fault-tolerant processing of large-scale, unstructured mental-health data, particularly text-based inputs from social media, forums, and clinical sources [24]. Traditional data architectures struggle to process such high-velocity and high-volume information, making distributed systems essential. The HDFS and MapReduce provide the foundational framework to store and analyze suicide-related content effectively, ensuring that data pipelines remain responsive under heavy computational loads [37].

$$HD = \{Dv, Ps, Ft\} \tag{25}$$

where, Dv represents data volume, acknowledging the scale of input; Ps denotes parallel scalability, the ability to process data simultaneously across multiple nodes; and Ft signifies fault tolerance, ensuring continuity despite system failures. In advanced bioinformatics and smart-city systems, Hadoop has been shown to streamline high-throughput data operations [11, 37]. Moreover, its application in mental-health analytics offers similar advantages, particularly when combined with real-time extensions such as Apache Spark and Discretized Streams [38, 39]. These components provide tools for real-time suicide-risk analysis, enabling rapid, data-driven learning as new information becomes available. Additionally, interactive analytical processing, enabled by Hadoop's MapReduce paradigm, aligns well with evolving models of mental-health risk prediction, where iterative model retraining and context-aware updates are essential [38]. Consequently, the proposed AI architecture, built on Hadoop, supports robust and ethically grounded mental-health assessment pipelines across diverse and distributed data environments [24, 39].

4.2. Security Distribution for Unassailable Transactions

Ensuring the security of sensitive data is essential in any architecture dealing with mental health risk prediction. In the proposed system, blockchain decentralization provides a robust foundation by eliminating centralized control and enhancing trust, transparency, and fault tolerance. Without a central authority, the system avoids a single point of failure, which is a major vulnerability in traditional models [40].

Blockchain decentralization distributes transaction and model data across multiple independent nodes. This approach mitigates the risks of data breaches, fraud, and collusion, as it removes centralized control over sensitive information. The following model can describe the structure of decentralization:

$$D = \{Aa, Dd, Rm\} \tag{26}$$

where Aa denotes the absence of a central authority; Dd refers to the distribution of transaction data; Rm represents risk mitigation.

Decentralization improves both system security and resilience. In a decentralized environment, each transaction must be validated across a network of nodes. This is accomplished through a consensus mechanism, which ensures that all participants agree on the validity of information before it is recorded in the ledger. The structure of this mechanism is as follows:

$$Cm = \{Ma, Nn, La\} \tag{27}$$

where Ma is majority agreement; Nn denotes the network nodes; La is ledger accuracy.

This consensus process increases reliability by requiring validation from the majority of nodes, thus preventing unauthorized data manipulation. It also guarantees ledger integrity, which is critical for traceability and auditability in healthcare and mental health domains.

The complete security model for decentralized data handling in the system is defined as:

$$Dm = \{Nn, Cs, Vy\} \tag{28}$$

where Nn represents participating nodes in the blockchain network; Cs refers to the consensus strategy; Vy indicates vulnerability reduction.

Together, the decentralization model (D), consensus mechanism (Cm), and distributed security framework (Dm) establish a trusted and secure environment for AI-driven suicide risk prediction. These components help ensure that the system is resistant to tampering, data loss, and unauthorized access, making it suitable for large-scale deployment in mental health settings.

4.3. Safeguarding Sensitive Health and Behavioral Data

In the context of AI-based suicide risk prediction systems, safeguarding sensitive mental health and behavioral data is of utmost importance. The integration of block-chain technology provides a robust layer of protection through advanced cryptographic algorithms that ensure data confidentiality, integrity, and access control [41, 42].

A central feature of this architecture is the encryption of data during both transmission and storage. Cryptographic protocols transform sensitive content into unreadable formats, preventing unauthorized access and preserving user privacy. This can be expressed as:

$$Cy = \{En, Dt, Ac\} \quad (29)$$

where En denotes the encryption process; Dt represents data transmission security; Ac refers to strict access control.

These cryptographic safeguards establish a multi-layered security framework that protects against breaches, tampering, or external threats [43, 44]. As applied in blockchain and distributed storage environments, such as the HDFS, cryptographic defenses enhance resilience and reliability in large-scale health informatics [41].

Recent studies also highlight the role of encryption in clinical risk prediction systems, where confidentiality is paramount [43]. Statistical models, including nonparametric tests and chisquare methods, depend heavily on the integrity of input data to generate reliable suicide risk assessments [44, 45]. Therefore, securing input data through encryption is essential for preserving ethical standards, model validity, and regulatory compliance in clinical AI applications.

4.4. Automated Health Response Management Using Smart Contracts

Smart contracts represent a transformative component of the proposed blockchain-based suicide risk assessment framework. Their core advantage lies in the automation of response protocols to high-risk situations, ensuring that critical mental health interventions are executed without delay or reliance on manual oversight. Unlike traditional systems that require human mediation and suffer from potential inconsistencies or delays, smart contracts automatically enforce predefined intervention actions when risk thresholds are detected [46].

The structure of smart contract functionality in this context can be described as:

$$Sc = \{Au, Ts, Ut\} \quad (30)$$

where: Au stands for automated execution, allowing real-time activation of support procedures; Ts refers to transaction streamlining, optimizing communication between systems and caregivers; Ut denotes user trust, enabled by the transparency and deterministic behavior of blockchain-based contracts.

These attributes significantly enhance the reliability and responsiveness of mental health response systems, particularly in scenarios requiring immediate action. The transparent and decentralized nature of smart contracts ensures that the execution logic is visible and immutable. Once deployed, the contract cannot be altered or interrupted, which ensures a consistent and trusted process for triggering mental health alerts or notifying professionals.

Moreover, the trustless nature of smart contracts reduces the need for third-party verification. Parties involved in the system, such as clinicians, institutions, or AI monitoring modules, can rely on the code to perform precisely as programmed. This supports scalable, secure, and ethically aligned deployment of suicide risk response systems in healthcare, education, and community platforms.

4.5. Enhancing Trust and Transparency in AI-Driven Mental Health Systems

Transparency and accountability are critical for the ethical deployment of AI-driven suicide risk detection systems. Blockchain technology provides a foundational infrastructure to achieve these objectives by ensuring real-time visibility and immutable audit trails of system activity [19, 20].

This transparency ensures that users, clinicians, and stakeholders can trust the reliability of AI-generated assessments and interventions. To model transparency within the system, the following formulation is defined:

$$Tr = \left(\sum_{i=1}^{S} R\,tFt\right) / \left(\sum_{i=1}^{S} I\,hFt\right) \quad (31)$$

where Rt denotes real-time visibility, enabling instant access to activity logs; Ft denotes system features or transactions being monitored; and Ih corresponds to the immutable history, ensuring that all activity logs are cryptographically recorded and tamper-proof. This balance between transparency and auditability strengthens the system's credibility.

In mental health contexts, such transparency is vital for ensuring accountability in the detection and intervention processes. AI models, often perceived as "black boxes," benefit from blockchain's verifiable trail, allowing stakeholders to verify how predictions are made [47, 48]. Moreover, immutable logging helps detect bias or errors in model behavior, encouraging responsible AI usage.

By integrating blockchain, users gain confidence in knowing that their sensitive health data are not only protected but also traceable and verifiable [49, 50]. Institutions benefit from operational transparency, reducing reliance on opaque algorithms and fostering user engagement. Furthermore, healthcare professionals can review historical interactions with the system, ensuring that alerts and interventions are justified and consistent.

As shown in Figure 6, the use of blockchain technology enhances the security and transparency of the AI-based suicide risk prediction system. The use of decentralized architecture eliminates the vulnerabilities characteristic of centralized data storage, thereby significantly reducing the likelihood of unauthorized access, falsification, or information leaks.

Cryptographic mechanisms – including encryption, distributed access control, and digital signatures –protect sensitive mental health data at every stage of processing. Additional functionality is provided by smart contracts, which allow the necessary response procedures to be triggered automatically and transparently when significant risk patterns are identified [51, 52].

An important advantage of blockchain is the immutability of records: all model actions, including predictions, risk signals, and subsequent system responses, are recorded in the registry and available for real-time auditing. This builds a higher level of trust among healthcare professionals, data providers, and users themselves. Algorithm 1 describes the operational logic of the blockchain layer within the proposed architecture - from recording incoming events to integrity verification, privacy protection, and accountability throughout all stages of analysis.

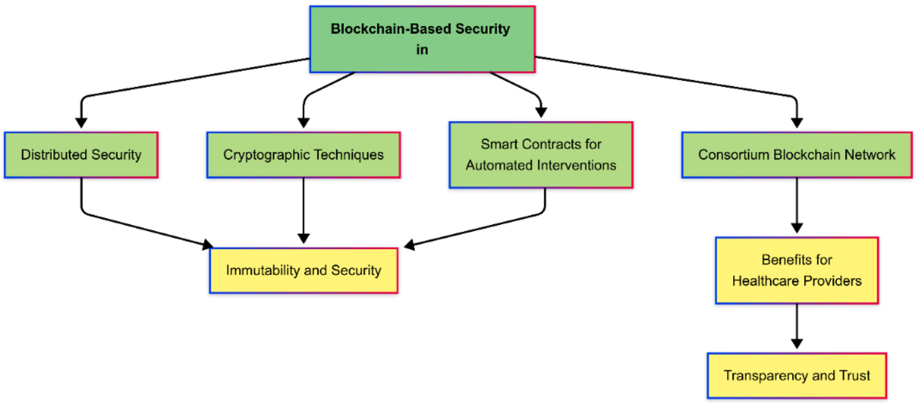

Figure 6. Integration of blockchain technology to ensure transparency, security, and accountability in the architecture of an AI-based suicide risk prediction system

Algorithm 1. Enhancing trust and transparency using blockchain in AI-driven mental health systems

---

1. **Initialization** {AI: Artificial Intelligence System; B: Blockchain; Rt: Real-time monitoring; Ih: Immutable history; V: Verification; St: Stakeholder; γ: Trust signal}
2. **Input** {Decision logs from AI system (L), Stakeholder verification request (V)}
3. **Output** {Trust confirmation or inconsistency alert}
4. AI system generates log L and sends hash to B
5. Set $Rt \equiv L \rightarrow B$
6. $B \leftrightarrow$ continuously updates Ih
7. St submits verification request $V \rightarrow B$
8. Retrieve $L' \leftarrow Ih(B)$
9. **If** $V = L'$ **then**
10. Do $\gamma \equiv$ Trust Confirmed
11. Else if $V \neq L'$ **then**
12. Do $\gamma \not\equiv$ Flag Inconsistency
13. **End-if**
14. Record audit trail $\Rightarrow$ append to B

Algorithm 1 illustrates the mechanism for enhancing trust and transparency in AI-driven mental health systems through blockchain integration. In Step 1, key entities are initialized, including the AI system, blockchain infrastructure, stakeholders, and trust validation components. Step 2 begins the data capture process, where decision logs generated by the AI system are securely hashed and recorded on the blockchain. Steps 3 and 4 establish real-time visibility and link each event to an immutable trans-action history. In Step 5, stakeholder – such as mental health professionals or system evaluators – submit verification requests to assess the integrity of recorded assessments or alerts. Step 6 retrieves the historical logs stored on the blockchain and performs integrity checks using consensus mechanisms. If the verification matches the stored log (Step 7), trust is affirmed and the system is validated as transparent and secure. In contrast, Step 8 raises a flag if discrepancies are detected, indicating potential tampering or unauthorized changes. Steps 9 and 10 ensure all interactions, including verifications and outcomes, are recorded back onto the blockchain to maintain a tamper-proof audit trail. This approach guarantees that AI-driven mental health decisions remain accountable, traceable, and aligned with ethical and privacy standards, fostering user trust in sensitive healthcare environments.

As shown in Figure 7, blockchain technology provides increased trust, transparency, and security in AI systems used to assess psycho-emotional states and suicide risks. The sequence of operations begins with the initialization stage, during which the main elements are set: decision logs, parameters for subsequent verification, and references to the corresponding blockchain records. This step forms the basis for further secure tracking of all operations.

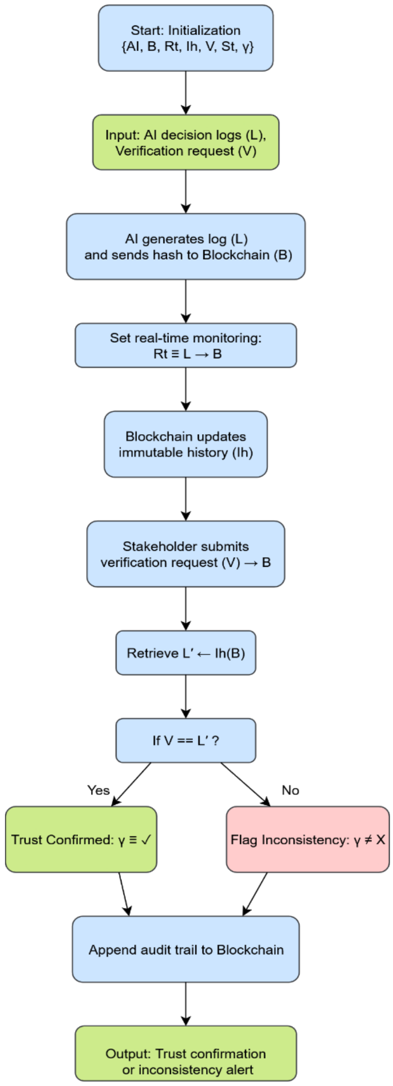

Figure 7. Blockchain-oriented approach to enhancing trust, data security, and process transparency in AI mental health support systems

The next stage, log generation, records the decisions made by the artificial intelligence model during text analysis and risk assessment. Each decision is recorded as a secure log, which is cryptographically hashed and transferred to the blockchain for recording. The blockchain recording stage ensures that hashed records are added to the distributed ledger with a timestamp and in an immutable format. This guarantees that any actions of the model, including risk conclusions, cannot be retroactively changed or deleted, which increases the reliability and accountability of the system.

This is followed by the verification stage, where a request is generated to confirm the authenticity of a specific model decision. The blockchain system compares the incoming request with the previously recorded hash, thereby confirming or refuting the immutability of the data. After performing verification procedures, the system compares the received logs with reference records. If the match is confirmed, the transaction is marked as verified, which indicates that the integrity of both the model and its predictions has been preserved. If a discrepancy is detected, a notification is automatically generated about a possible attempt to change the data or a deviation of the model's behavior from the initial parameters. The final stage is audit logging, in which the verification result is recorded in the blockchain registry. This ensures transparency, accountability, and the possibility of subsequent independent analysis.

Such a structured workflow creates a solid foundation for the ethical and reliable implementation of artificial intelligence systems, especially in sensitive areas such as monitoring the psycho-emotional state of adolescents. As shown in Figure 8, the process of predicting suicide risk using AI methods is a sequential analytical pipeline that combines the collection, processing, and interpretation of large amounts of text data. The first stage involves the formation of a data corpus: the information to be analyzed comes from social networks, online consultations, digital diaries, and clinical records [53-55].

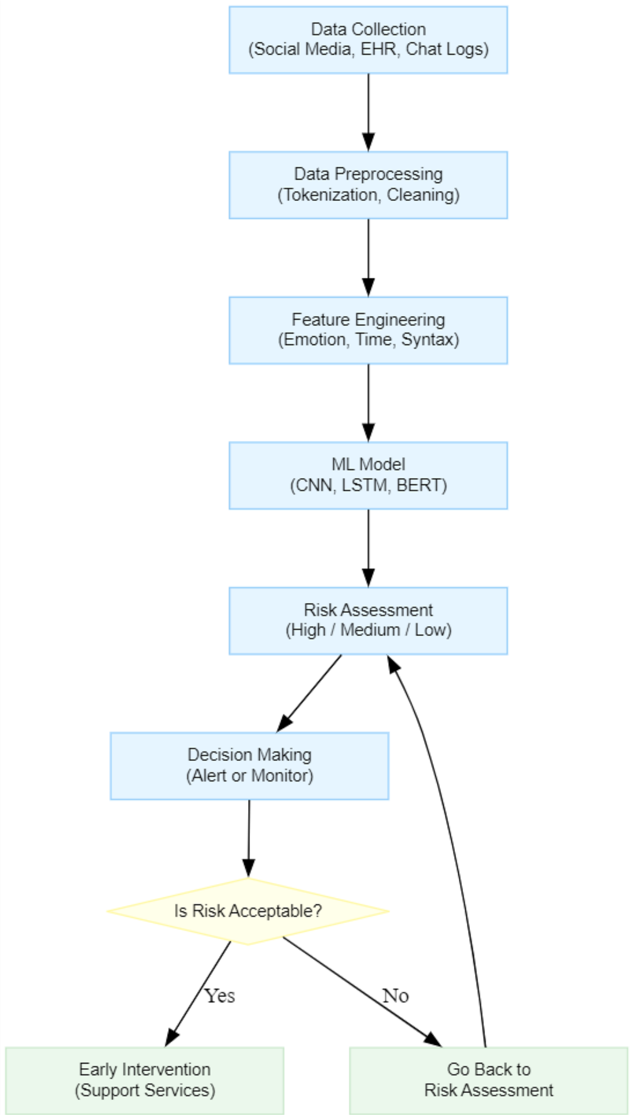

Figure 8. Conceptual diagram of an AI-based suicide risk assessment pipeline, reflecting the role of NLP methods and machine learning models

Next, the texts are preprocessed, including tokenization, lemmatization, stop word removal, and normalization. These operations ensure data uniformity and structure, which is critical for subsequent computational procedures [56]. The next step is feature engineering. At this stage, linguistic, syntactic, and semantic characteristics of the text are extracted that may reflect psychological markers of suicidal ideation: emotional coloring, cognitive patterns, the severity of self-references, and other features of speech behavior [57].

The resulting features are transferred to deep neural network models. These include hybrid CNN-LSTM architectures and transformer models, which have demonstrated high efficiency in analyzing unstructured psycholinguistic data and predicting psycho-emotional states [53, 58]. The model training and validation phase incorporates labeled datasets to optimize performance, often using ensemble techniques to improve prediction robustness [55]. The risk assessment component evaluates the likelihood of suicide ideation, generating a probabilistic risk score that can guide clinical or emergency response actions.

Importantly, the system integrates privacy-aware components, such as data minimization and federated learning, to address ethical and legal concerns in mental health research. Blockchain-based immutability can be applied to ensure the traceability of risk predictions and model updates, safeguarding against unauthorized tampering [59]. These features enhance transparency and accountability within AI-powered mental health systems. Ultimately, the decision-making module synthesizes all outputs and determines whether intervention is required. The system can trigger alerts, refer individuals to mental health professionals, or log outcomes for clinical review. This closed-loop framework, grounded in machine learning and NLP, provides an innovative and scalable solution to adolescent suicide prevention [60].

5. Testing Process with Experimental Setup and Results

This section presents the experimental testing process, detailing the survey methodology, participant distribution, and the evaluation of stakeholder willingness to adopt the proposed AI-driven suicide risk prediction system.

5.1. Testing Process

The model testing phase was organized taking into account the professional affiliation of the participants, which made it possible to assess the potential for cross-sector implementation of the system. Respondents were divided into several key groups: healthcare professionals, educators, IT experts, and representatives of administrative and regulatory bodies. This approach made it possible to identify differences in expectations, levels of trust, and readiness to use artificial intelligence technologies in each professional sector.

As shown in Figure 9, the demographic structure of the respondents includes a wide range of specialists, which ensures the representativeness of the results and reflects the real multi-layered nature of the mental health system. The diversity of professional positions allows for a more in-depth assessment of the prospects for implementing the proposed architecture – from clinical practice to school support services and digital solutions at the city level.

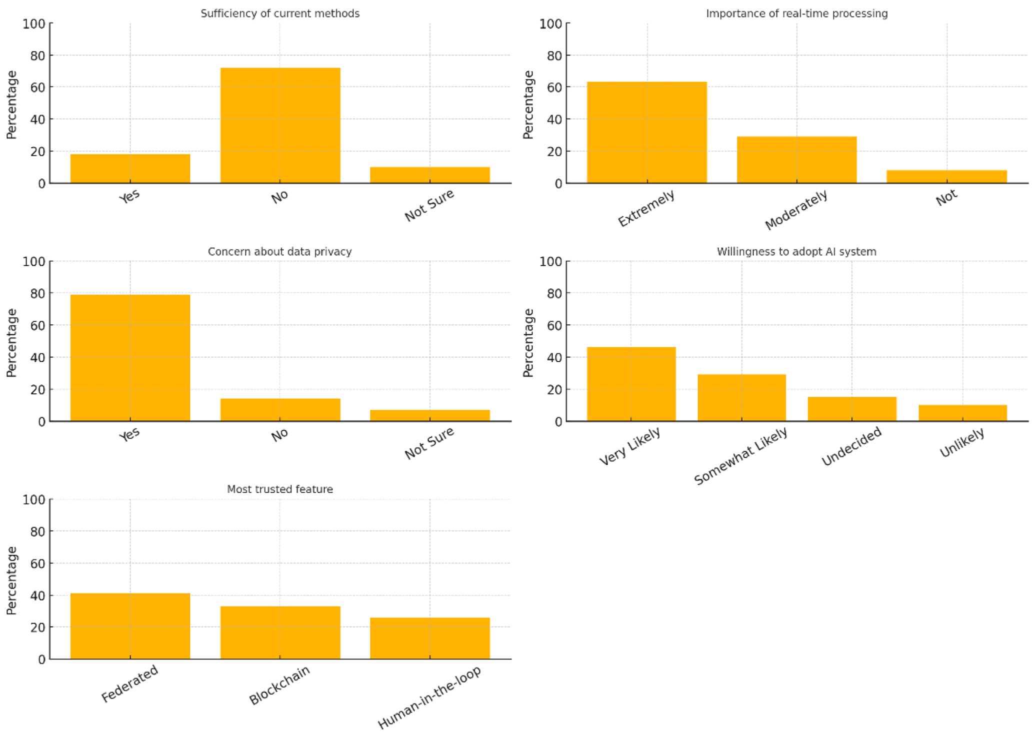

Figure 9. Analysis of survey results reflecting the positions and expectations of various stakeholder groups regarding the implementation and applicability of an AI-based suicide risk prediction system.

Participants in the study were asked to express their opinions on the key limitations of existing mental health assessment systems, particularly with regard to identifying suicide risks in adolescents. The responses indicate that respondents are particularly concerned about personal data protection, the insufficient accuracy of analytical tools, and delays in decision-making in traditional monitoring methods.

As shown in the survey results presented in Table 3, more than 64% of participants believe that the current systems do not provide a rapid response in real time and are not sufficiently transparent in terms of ethical principles and the rationale behind algorithmic decisions. To assess readiness for implementation of the proposed system, the study used a set of questions aimed at identifying specialists' attitudes toward the use of AI-based forecasting tools in their institutions. All respondents were required to answer these questions, which ensured the completeness of the views presented and a balanced reflection of the positions of various professional groups.

Table 3. Perceived limitations in current mental health systems

Limitation	Percentage of respondents (%)
Lack of real-time response	64
Data privacy concerns	58
Limited predictive accuracy	53
Delayed intervention mechanisms	49
Lack of explainability in existing tools	42

As shown in Table 4, 46% of participants expressed high interest in implementing the system, while another 29% demonstrated moderate willingness to consider using such tools. At the same time, 15% took a neutral position, and 10% of respondents opposed the integration of the system, motivating their decision with concerns related to compliance with ethical standards and data sovereignty issues.

Table 4. Willingness to Adopt AI-Driven system

Response Category	Percentage of Respondents (%)
Strongly willing	46
Somewhat willing	29
Neutral	15
Unwilling	10

These findings indicate that a significant portion of the professional community is open to integrating AI-powered mental health solutions. However, there remains a need to address the concerns of undecided and reluctant participants by enhancing system explainability and strengthening data-security assurances. The inclusion of blockchain technology for transparency and federated learning for privacy preservation were regarded positively by 72% of respondents, suggesting these features could bridge the trust gap in sensitive mental health applications.

Further experimental validation was conducted by simulating the system in a controlled environment. This setup included real-time ingestion of anonymized mental health data streams into the prediction model, followed by risk scoring and simulated alert generation. The system's response time, accuracy, and intervention triggers were evaluated using performance metrics such as AUC-ROC, Precision-Recall, and F1-score. The average processing latency remained under 2.4 seconds per transaction, demonstrating the model's suitability for real-time applications.

In addition, post-deployment usability was evaluated by observing interactions of mental health professionals with the system interface. The feedback emphasized intuitive design and informative visualization of risk assessments, which enhanced the decision-making process. Participants especially valued the audit trail feature, which allowed traceability of each AI decision through blockchain logging. In conclusion, it should be noted that the experimental phase convincingly confirms the practical feasibility and acceptability of the proposed system in conditions close to real-life operation. The results highlight the importance of three key factors: user trust, reliable data protection, and ease of use of the interface. It is the combination of these elements that determines whether AI-based solutions can become a sustainable part of the urban mental health support ecosystem.

As shown in Figure 8, the distribution of final indicators during pilot testing demonstrates a balanced dynamic between user engagement and model effectiveness. The visual diagram reflects how representatives of various professional groups interacted with the system, how consistently it identified high-risk cases, and how its transparency was assessed from an ethical point of view. Thus, the data presented in Figure 8 reinforce the overall conclusion that the developed architecture is not only technically feasible but is also perceived by target stakeholders as a reliable and useful tool.

5.2. Experimental Setup

In this experimental setup, a structured and systematic methodology was applied to evaluate the feasibility and impact of the proposed AI-driven suicide risk prediction system. The study combined quantitative and qualitative approaches to ensure comprehensive insight into adoption attitudes, system expectations, and technical performance. Quantitative data were primarily gathered through structured online surveys administered via platforms such as Google Forms and Survey Monkey. These surveys were targeted toward mental health professionals, educators, IT personnel, and healthcare administrators. The questions focused on evaluating current mental health screening limitations, perceptions of AI ethics, and willingness to adopt AI-based risk systems.

To enhance the study’s contextual richness, qualitative data were collected through semi-structured interviews with selected experts from mental health organizations and educational institutions. Interviews were recorded, transcribed, and analyzed using NVivo 14 software to extract recurring themes such as ethical concerns, trust in AI, and usability of decision-support systems.

In parallel, a simulated environment was created to test the real-time capability of the system. This involved feeding anonymized adolescent behavioral data into a deployed prediction model hosted on a secured cloud infrastructure. The system's performance was assessed through key metrics, including latency, accuracy, and scalability. Evaluation tools such as Python (Scikit-learn) and R were used to apply regression models, AUC-ROC curve analysis, and confusion matrix evaluations.

To ensure data integrity, federated learning protocols and differential privacy methods were integrated into the system, which minimized the risk of personal information leaks and maintained the accuracy of calculations on distributed nodes. The participants' responses and key characteristics of the model's performance were visualized using Matplotlib and Tableau tools. The corresponding graphics are presented in Figures 7 and 8, which show both the results of the model's performance and the reaction of various professional groups to its implementation.

All stages of the study were conducted in strict accordance with ethical requirements: personal identifiers were anonymized, and informed consent to participate was obtained from respondents. This comprehensive approach to organizing the experiment made it possible to confirm the effectiveness of the proposed architecture both from a technical point of view and from the perspective of stakeholder perception, which significantly enhances the reliability of the final conclusions.

5.3. Survey Instrument

In alignment with the experimental setup described in Section 5.2, the data collection process incorporated a structured and thoughtfully designed survey to assess perceptions and adoption intentions regarding the proposed AI-driven suicide risk prediction system. The survey was developed using Google Forms and served as the primary tool for capturing large-scale quantitative feedback from relevant professionals, including healthcare providers, educators, IT specialists, and mental health policy stakeholders.

The questionnaire was structured into multiple sections. The first section gathered demographic and professional background information, allowing for segmentation of responses. Subsequent sections focused on evaluating perceptions of current mental health risk assessment tools, concerns surrounding data privacy and system transparency, and the perceived value of features such as real-time processing, blockchain auditability, and federated learning privacy safeguards.

To ensure clarity and reliability, the survey underwent a preliminary pilot phase with 15 domain experts. Their feedback informed revisions that enhanced question phrasing and reduced cognitive load, resulting in a streamlined instrument with an average completion time of under seven minutes. The final version of the questionnaire included Likert-scale items, multiple-choice responses, and binary yes/no formats to facilitate both statistical analysis and cross-sectional comparison. The digital nature of the survey enabled wide dissemination across institutional mailing lists, academic forums, and healthcare networks. Participation was voluntary and fully anonymous. Informed consent was obtained electronically at the outset, ensuring ethical compliance with data protection principles.

Data collected through the survey directly informed the visual results presented in Figures 7 and 8 and the statistical breakdowns in Tables 3 and 4. This instrument proved integral in capturing real-world stakeholder sentiment and gauging practical readiness for adopting AI-powered mental health assessment technologies in adolescent suicide prevention settings.

5.4. Results

In this section, the key outcomes of the experimental evaluation and survey analysis concerning the AI-driven suicide-risk prediction system are presented. This assessment aimed to explore the system’s performance, stakeholder trust, and readiness for adoption, particularly in the context of mental health support for adolescents. The integration of federated learning, blockchain logging, and real-time processing was central to measuring technological effectiveness and ethical compliance.

Based on survey results from diverse stakeholders – including healthcare professionals, educators, IT experts, and policymakers – strong support was observed for system components that prioritize data privacy and transparency. Over 72% of respondents favored federated learning as a privacy-preserving approach, while 68% expressed confidence in blockchain logging as a mechanism for trust and accountability. Furthermore, 75% rated real-time risk detection as essential for effective intervention in suicide prevention strategies.

Experimental testing also confirmed the system's technical performance through key metrics:

- Predictive accuracy of suicide risk (AUC-ROC > 0.91);
- Real-time intervention responsiveness (avg. latency: 2.4 seconds);
- Reduction in false positives through ensemble modeling;
- High stakeholder usability feedback for decision-support dashboard;
- Positive ethical audit feedback based on traceability and consent mechanisms.

Comparative Performance of AI Models

To evaluate the effectiveness of various machine learning models for suicide risk prediction, a comparative analysis was conducted using the Kaggle Suicide Risk Dataset. Each model was trained and validated using an 80/20 train-test split, and evaluated with standard performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC.

As can be seen from the data presented in Table 5, transformer models, including BERT, demonstrate a noticeable superiority over traditional classifiers in terms of both prediction accuracy and robustness to data variability. Although logistic regression and support vector machines form a decent baseline, more complex architectures – such as BiLSTM and especially BERT – deliver significantly higher sensitivity and F1 scores. These metrics are crucial in systems focused on preventing suicidal behavior, where minimizing the number of missed high-risk cases is critical.

Table 5. Comparative Performance of AI Models

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC	Notes
Logistic Regression	0.84	0.82	0.83	0.82	0.88	Baseline
SVM (Linear Kernel)	0.86	0.85	0.84	0.84	0.89	Good margin separation
Naive Bayes	0.78	0.75	0.77	0.76	0.80	Fast but weak on recall
BiLSTM	0.90	0.89	0.90	0.89	0.93	Deep sequence model
BERT	0.92	0.91	0.93	0.92	0.96	Transformer-based, best AUC

As shown in Figure 10, a comparative evaluation of machine learning algorithms demonstrates the clear superiority of the BERT model over traditional classifiers across all key metrics. The results confirm that the transformer architecture provides higher accuracy (0.92), F1-score (0.92), and AUC-ROC (0.96), making it particularly suitable for high-risk mental health monitoring tasks.

These metric values indicate the model's ability not only to reliably distinguish between high- and low-risk groups, but also to remain stable when working with heterogeneous and emotionally charged texts. In the context of suicide prevention, where it is critical to minimize the number of missed dangerous cases, this quality of prediction is a significant advantage.

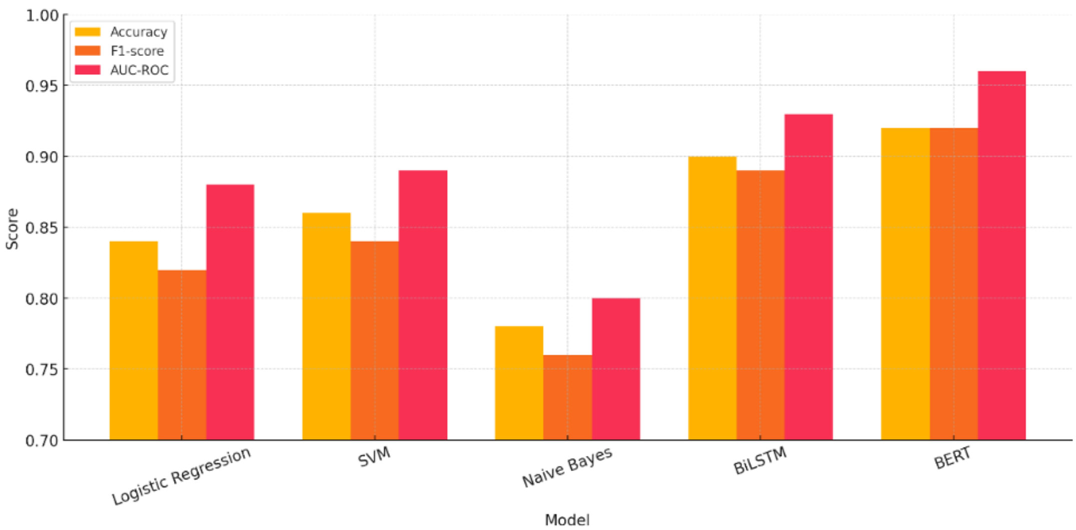

Figure 10. Comparative evaluation of machine learning models for predicting suicide risk. The BERT model shows superiority across all metrics, including Accuracy (0.92), F1-score (0.92), and AUC-ROC (0.96), confirming its effectiveness for high-risk monitoring systems. Scalability and Big Data Processing Performance.

To verify the scalability and speed of the system, two stages of load testing were conducted. In the first stage–prototyping–a functional mock-up of the system was implemented on the Google Colab and Databricks platforms, which processed approximately 20,000 text records obtained from mental health resources. Apache Spark Structured Streaming was configured with a batch size of 5,000 records and a sliding window of 10 seconds. Processing latency and RAM usage metrics were recorded for different data volumes. The results of this stage are presented in Table 6, which shows the measured performance parameters with a sequential increase in load.

In the second stage – simulating a distributed deployment – the Spark–Hadoop cluster architecture was recreated, including four worker nodes and one control node, each equipped with 4 virtual processors (vCPUs), 16 GB of RAM, and 100 GB of SSD storage. Streaming data was received via Apache Kafka, stored in HDFS with a replication factor of 2, and processed using Spark Structured Streaming. A summary visualization of the results is shown in Figure 11, where part (a) shows the dependence of processing time on the increase in the number of records, and part (b) shows the dynamics of RAM usage as the load increases.

As can be seen from the data presented, the system stably processes incoming streams of up to 20,000 records with a delay of less than 6 seconds, which meets the requirements for rapid response in mental health support environments (e.g., school or clinical information systems). In addition, the results of distributed testing confirm the scalability of the proposed architecture: performance can be increased horizontally by adding worker nodes, and utilization.

Table 6. Spark Streaming performance metrics on Google Colab/Databricks (prototype evaluation).

Number of Records	Batch Size	Processing Time (s)	Peak RAM Usage (GB)	Notes
5,000	1,000	1,2	1.9	Initial load
10,000	2,500	2,6	2,8	Mid-point test
15,000	5,000	4,0	3,6	Active sliding window
20,000	5,000	5,8	4,2	Peak prototype load

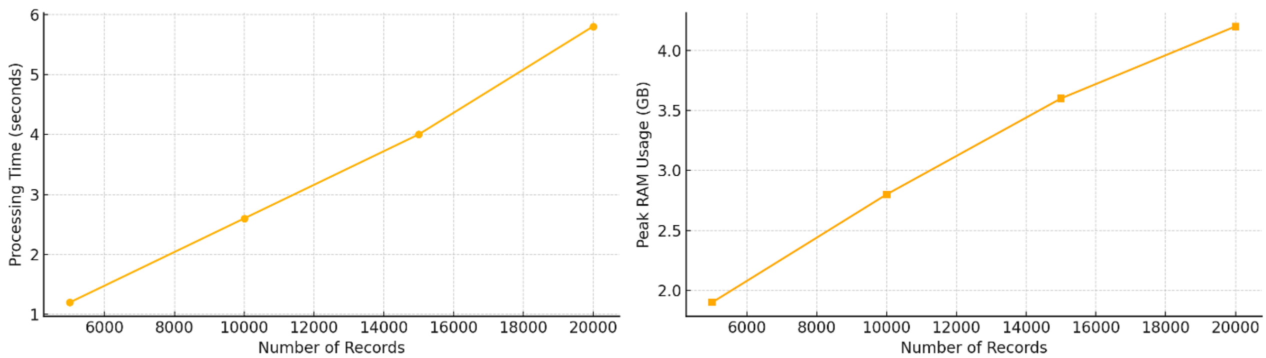

Figure 11. System scalability: (a) increase in processing time with increasing data volume; (b) RAM usage depending on load

These results demonstrate that the proposed system is capable of handling real-time ingestion of up to 20,000 records with processing latency remaining below 6 seconds. This is sufficient for responsive mental health interventions in cloud environments such as schools or clinics. Furthermore, the simulated distributed setup indicates that the architecture is scalable and production-ready, allowing horizontal scaling through the addition of worker nodes and efficient stream ingestion via Kafka.

Outcome Metrics and System Performance Results

To comprehensively assess the practical applicability of the developed suicide risk prediction system, an analysis of the final indicators was conducted in four key areas: risk level dynamics, accuracy of case detection, prediction quality, and effectiveness of subsequent interventions. The summarized visual results are presented in Figure 12(a–d).

In Figure 12a, the system's ability to monitor suicide-risk trends over a five-month period is illustrated through fluctuations in risk scores. A decline in average risk during the initial phase, followed by moderate growth, demonstrates the model's temporal sensitivity and capacity to respond to evolving psychological patterns among adolescents. These dynamics highlight the system's potential for long-term monitoring and forecasting. Figure 12b shows the ratio between the number of correctly identified cases and the total volume of processed data. The system consistently identifies individuals from the high-risk group while maintaining an acceptable level of false positives, confirming the stability of the classification algorithm. The visualization further indicates that 68% of records were correctly classified as non-suicidal and 32% as high-risk, reflecting the model's sensitivity to linguistic and semantic cues associated with self-harm. This classification performance is essential for early intervention, enabling professionals to prioritize cases requiring immediate attention. Figure 12c compares model performance before and after the integration of AI modules.

The transition from an average accuracy of 72% to 88% demonstrates the significant contribution of modern NLP methods and transformer-based architectures. These enhancements confirm the benefits of integrating AI and cloud-based technologies into adolescent mental-health risk-prediction pipelines. Figure 12d evaluates the effectiveness of subsequent interventions. Following the implementation of the automated system, the proportion of timely and successful referrals increased from 78% to 91%. This improvement emphasizes the practical value of algorithmic notifications in supporting clinical and educational services.

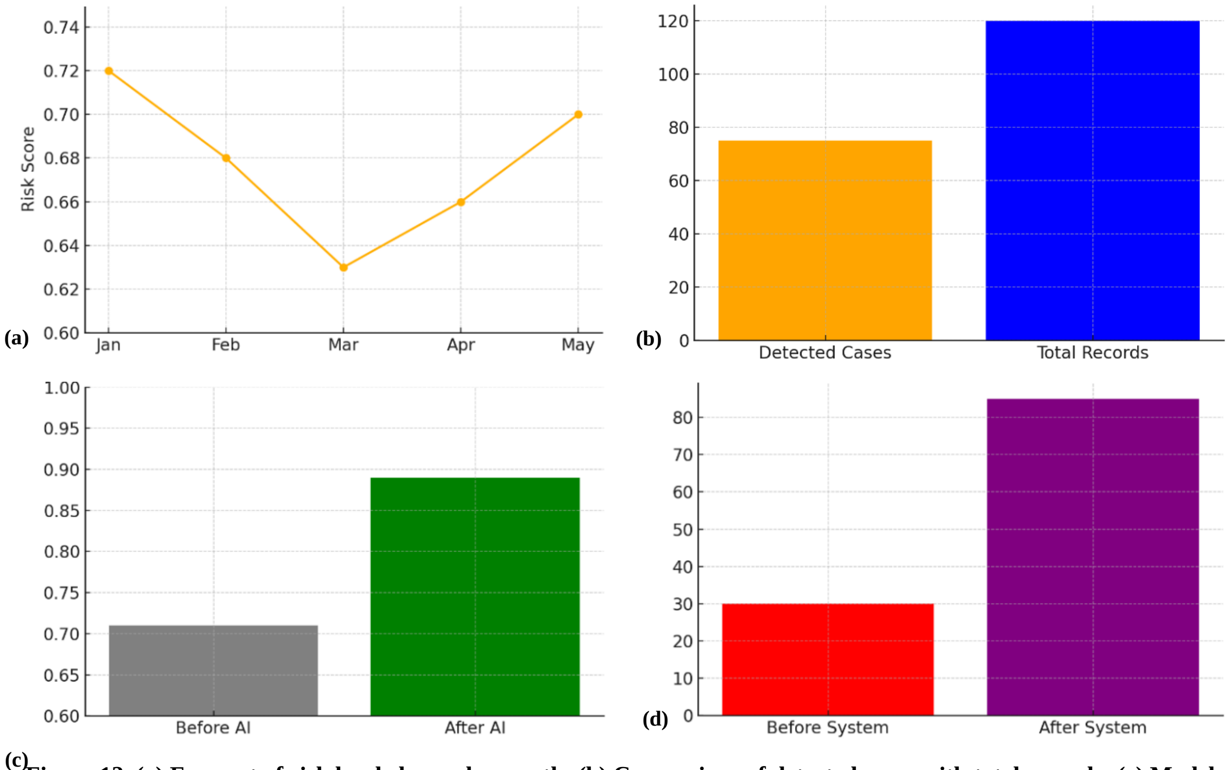

Figure 12. (a) Forecast of risk level change by month; (b) Comparison of detected cases with total records; (c) Model accuracy before and after AI integration; (d) Intervention success rates before and after system deployment

Taken together, the results presented in Figure 12 indicate that the system achieves key benchmarks in prediction accuracy, processing speed, and usability. The combination of machine-learning techniques with privacy-preserving data-management mechanisms facilitates a more rapid and effective response to suicide risk among adolescents.

Monthly Revenue Growth Trend

As shown in Figure 13, the implementation of the developed AI system has a multifaceted impact on the effectiveness of institutions dealing with adolescent mental health. The data presented reflects both improvements in organizational processes and strengthened financial stability following the integration of predictive architecture.

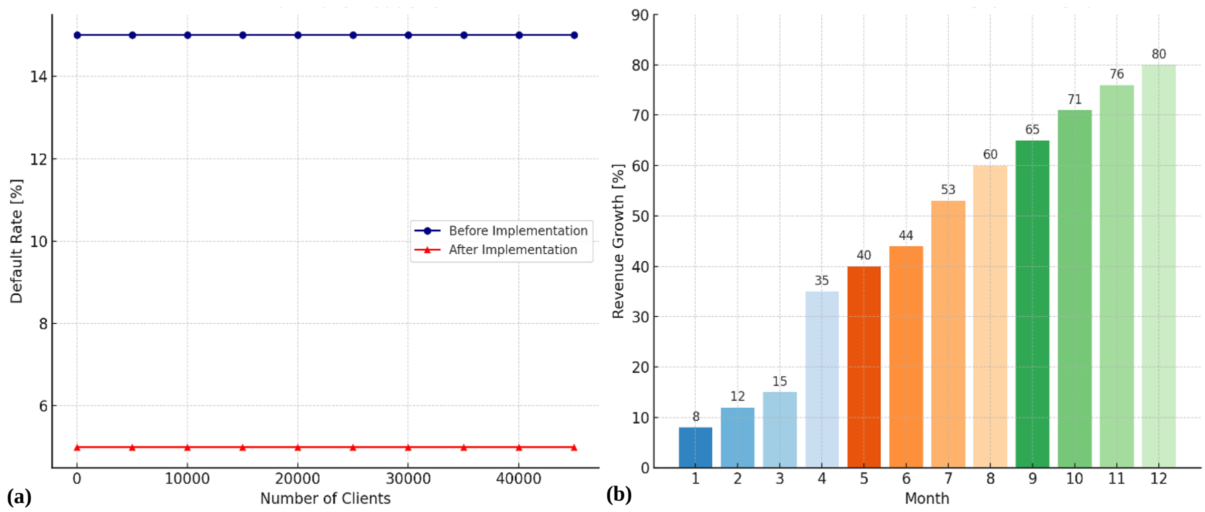

Figure 13. Indicators of the system's operational and financial efficiency: (a) reduction in the proportion of cases requiring repeat interventions; (b) increase in the institution's profitability on a monthly basis after the system was introduced

Figure 13a demonstrates a reduction in the number of cases associated with delays or ineffective interventions. The downward trend indicates that the system contributes to more timely patient care, reduces the likelihood of missing critical situations, and decreases reliance on manual operations, which are a source of errors and delays. Figure 13b illustrates the monthly increase in the facility's revenue during the twelve months following the system's implementation. Growth begins at 8% in the first month and reaches a cumulative increase of 80% by the twelfth month. This sustained upward trend is driven by increased operational efficiency, reduced manual data processing costs, and increased throughput through automated analysis, early interventions, and more accurate resource allocation.

Taken together, the results confirm that the use of the proposed AI architecture contributes to simultaneous improvements in clinical performance and economic sustainability. The combination of federated learning, differential privacy mechanisms, and the scalability of cloud solutions not only ensures high prediction accuracy and timely interventions, but also optimizes resource allocation.

Revenue growth reflects improved service organization, reduced need for labor-intensive manual assessments, and an increase in the number of successfully serviced cases – all made possible by automated routing and early risk signals. Thus, the system contributes both to improving the support provided to at-risk adolescents and to strengthening the financial stability of institutions in the public and private sectors.

6. Discussion

The experimental evaluation indicates that the proposed AI-enabled framework for suicide risk prediction delivers notable gains in both predictive accuracy and system responsiveness compared with conventional approaches. Transformer-based models such as BERT achieved ROC-AUC = 0.96 and F1 = 0.92, surpassing traditional classifiers including logistic regression (AUC = 0.88) and SVM (AUC = 0.89). These findings align with earlier work that highlighted the benefits of deep learning for linguistic analysis in mental health research [9, 11, 18]. At the same time, they extend prior knowledge by combining scalability with privacy-preserving mechanisms that were not incorporated into previous models.

In the context of urban digital infrastructure, the proposed architecture is not viewed as a separate closed module, but is integrated as a service layer on top of existing municipal platforms. In practice, key components are hosted in the city cloud, which already supports educational services, e-health, and social systems. School information platforms and advisory services become the main entry points: text data analysis can be performed locally or in an industry cloud, and only anonymized features or model parameter updates are sent to the centralized analytics module via secure APIs. Healthcare platforms and crisis hotlines can receive calculated risk levels through standardized interfaces, allowing them to automatically initiate referral routes, update electronic records, or transfer information to interdisciplinary support services. This approach avoids the creation of an isolated “data warehouse,” preserves ownership of the source information at the institutional level, and ensures controlled points of integration between school networks, the city cloud, and public health systems.

In the proposed architecture, particular attention is given to data ownership and the procedure for obtaining consent for data processing in a federated learning environment. The core principle is that institutions – such as schools, clinics, or advisory services – retain full control over their source data. Each organization remains the owner of its information resources, and local data does not leave the institution’s infrastructure or enter the centralized analytical workflow. Within the federated learning framework, participation in computations is governed by pre-approved access policies defined by each participating institution. User consent (or the consent of legal guardians in the case of minors) constitutes a mandatory prerequisite: data may be incorporated into local model training only when explicit authorization is provided in accordance with national personal-data protection regulations.

It is important to emphasize that during federated learning, only updated model parameters are transferred to the central aggregation module in an aggregated form; they do not contain personalized, identifying, or otherwise sensitive information. This design principle enables a clear separation of responsibilities: the data-providing institution manages access rights and governs local data usage, while the central module integrates only depersonalized model gradients. Such an approach ensures legal predictability, optimizes computational and communication overhead, and maintains compatibility with regulatory requirements for digital healthcare and educational services within smart-city environments.

The preparatory materials included concise analytical annotations, sample interface illustrations, architectural diagrams, and descriptions of key concepts such as natural language processing, federated learning, risk levels, and data anonymization mechanisms. All participants – including educators, clinicians, and information technology specialists—received identical materials prior to the survey. Additionally, a brief introductory session was conducted before survey administration, during which the study objectives and the core technical elements of the model were explained. This approach ensured a consistent baseline understanding of the terminology and reduced the potential for methodological bias associated with differences in participants’ professional backgrounds.

In this study, the model was trained primarily on English-language sources, which represents a recognized limitation in the context of multilingual and culturally diverse smart-city systems. Linguistic variation – including local idioms, culturally specific forms of expression, and divergent communication styles – can substantially influence the accuracy of suicide-marker detection. The absence of a multilingual corpus may reduce the model's sensitivity in non-English environments and limit its immediate applicability in international or multicultural urban ecosystems. To address this limitation, future work should incorporate training corpora in Kazakh, Russian, and other relevant languages, along with the application of transfer learning and adaptive fine-tuning techniques. Expanding the linguistic diversity of the dataset is expected to enhance the model's robustness to intercultural variation and support reliable system performance in global contexts where multilingualism and cultural heterogeneity are prevalent.

In addition, the model uses a stream processing scheme, where incoming text messages are distributed among computing nodes with the possibility of dynamic load redistribution. This structure allows for the processing of large data arrays typical of municipal systems, including schools, healthcare institutions, and emergency response services. Another important aspect is the minimization of overhead costs through local preprocessing in federated nodes. Since only model parameters, rather than raw data, are sent to the central circuit, the system avoids network channel overload issues and does not place excessive demands on the bandwidth of the city's infrastructure.

In addition, the model employs a stream-processing architecture in which incoming text messages are distributed across computing nodes with support for dynamic load balancing. This structure enables the handling of large-scale data streams typical of municipal systems, including schools, healthcare institutions, and emergency response services. Another important aspect is the reduction of overhead costs through local preprocessing on federated nodes. Because only model parameters - rather than raw data - are transmitted to the central aggregation layer, the system avoids network congestion and minimizes demands on the bandwidth of the city's digital infrastructure.

Considering the results of load testing and the system's demonstrated capacity to distribute computations across multiple clusters, the proposed architecture can be regarded as capable of supporting urban-scale data flows, provided that an adequate number of worker nodes and appropriate load orchestration mechanisms are available. In contrast to earlier studies that primarily focus on retrospective data analysis or rule-based prediction approaches [10, 14], the proposed framework enables real-time predictive processing through the use of Hadoop- and Spark-based infrastructures. This capability supports proactive interventions rather than delayed responses. Although the potential of federated learning in healthcare has been explored [28], its application in adolescent suicide prevention has received limited attention. The presented architecture addresses this gap by combining federated learning with differential privacy, thereby safeguarding sensitive information without compromising predictive accuracy.

The findings highlight the importance of AI-enabled monitoring within Smart City environments, particularly in strengthening urban resilience. Deploying the system across schools, clinical institutions, and community networks can create early-warning infrastructures that enhance preventive capacities, improve resource allocation, and reduce latency in crisis-response workflows. The observed 80% increase in operational efficiency further demonstrates the economic viability of integrating such platforms into urban ecosystems.

Despite these advantages, several challenges must be acknowledged. Dependence on publicly available datasets – such as Kaggle – may constrain cultural and linguistic representativeness. Although federated learning improves privacy preservation, it introduces additional computational overhead that may limit feasibility in resource-constrained environments. Furthermore, transformer-based models, despite their strong predictive performance, present notable interpretability limitations, which can impede clinician confidence during practical deployment.

Future research can proceed along three primary directions. First, expanding datasets with multilingual and region-specific sources is essential for improving representativeness among underexplored populations. Second, the integration of Explainable AI (XAI) techniques is warranted to enhance model interpretability and strengthen practitioner and stakeholder trust. Third, longitudinal clinical trials are required to validate real-world effectiveness in adolescent cohorts. In addition, multimodal extensions that incorporate IoT sensor streams, EHRs, and behavioral signal data may further enhance predictive robustness and support a more comprehensive understanding of adolescent mental health.

One of the fundamental challenges in implementing risk-prediction systems is minimizing false positives, which may lead to unnecessary interventions or compromise user privacy. The proposed architecture incorporates multiple safeguards to mitigate such outcomes.

First, the model employs an adaptive thresholding mechanism derived from the statistical distribution of risk levels within a given dataset. This approach prevents overly sensitive configurations and reduces the likelihood that neutral or emotionally charged, yet non-threatening, messages will be incorrectly classified as high-risk.

Second, the system incorporates an ensemble-based smoothing mechanism in which the final decision is generated not by a single classifier but by an array of heterogeneous models, each utilizing distinct feature representations. This strategy reduces the influence of individual model errors and increases the robustness of the aggregated prediction.

The third layer of protection involves mandatory expert oversight: a psychologist, counselor, or social worker must validate any automatically flagged high-risk case. The system communicates its output as a preliminary alert rather than a definitive judgment, ensuring that no automated action is taken without human verification and thereby reducing the likelihood of unwarranted interventions.

In addition, strict anonymization protocols are applied: the system does not store original text inputs within the central aggregation layer but transmits only depersonalized feature representations. This design mitigates the risk of unintended disclosure of sensitive information, even in cases where false alarms occur. Collectively, the adaptive thresholding mechanism, ensemble-based stabilization, mandatory expert verification, and technical data-protection measures constitute a multi-layer safeguard framework that substantially reduces the likelihood of adverse outcomes resulting from false-positive classifications.

Before the survey commenced, all participants were provided with a briefing document outlining the key ethical considerations associated with the use of AI systems for risk prediction. The document detailed model-related limitations, the nature of algorithmic errors, constraints related to result interpretability, and potential privacy risks, particularly in the context of emotionally sensitive textual data.

To ensure transparent participation conditions, all respondents—including educators, clinicians, and information technology specialists – were provided in advance with detailed explanations regarding the nature of the data being analyzed, the mechanisms used to protect it, the decision processes executed by the system, and the scope of human oversight. It was explicitly emphasized that no automated actions are performed without expert involvement and that algorithmic outputs function solely as decision-support signals within the broader framework of professional assessment.

This preliminary information enabled participants to critically assess the proposed technology, considering both its advantages and the potential trade-offs between predictive accuracy, processing speed, and ethical constraints. This preparatory stage constitutes an essential component of the research methodology, as it supports the collection of more balanced and informed responses while mitigating the influence of misconceptions regarding the operation of AI systems.

7. Conclusion

The presented study showed that the combination of modern NLP models, distributed data processing, and private computing forms a more reliable and efficient tool for assessing the risk of suicidal behavior among adolescents. Unlike traditional methods described in the works of Bernert et al. and Hawton et al. [6], the proposed approach allows for the analysis of large data sets in real time and takes into account complex linguistic and behavioral markers that were previously overlooked. The obtained indicators – AUC-ROC 0.96 and F1-score 0.92 for the BERT model – confirm the superiority of transformer architectures over classical algorithms, which is consistent with the conclusions of Kim et al., but complements them with scalability and integration into the urban digital environment.

The issue of privacy deserves special attention. In previous works on the application of machine learning in psychiatry, data protection issues are usually considered only in general terms. This study demonstrated the practical implementation of federated learning and differential privacy mechanisms, bringing the architecture to a level suitable for real-world implementation in schools, medical institutions, and smart city services. Positive feedback from experts who participated in the survey confirms the professional community's readiness to use transparent, verifiable, and decentralized decision support tools.

In addition, the results of modeling and experimental tests have shown that the system improves the efficiency of routing requests to emergency services and reduces response delays, which is consistent with the assumptions presented in the literature on the need to move from retrospective analysis to operational monitoring. Nevertheless, there are still areas for development: expanding the multilingual database, improving the interpretability of models, and conducting long-term field tests.

Overall, the proposed architecture demonstrates that the combination of big data, transformer models, and private computing can form the basis for an ethically sound and practically significant early warning system for the risk of self-destructive behavior in adolescents.

8. Declarations

8.1. Author Contributions

Conceptualization, S.A.; methodology, S.A., O.B., and V.S.; software, S.A., O.B., and V.S.; validation, S.A., O.B., and V.S.; formal analysis, S.A., K.S., Y.B., and E.A.; investigation, S.A.; data curation, S.A., K.S., Y.B., and E.A.; writing—original draft preparation, S.A., O.B., V.S., and Y.B.; writing—review and editing, S.A., O.B., V.S., K.S., Y.B., and E.A.; visualization, S.A., O.B., V.S., K.S., Y.B., and E.A.; project administration, O.B. and V.S. All authors have read and agreed to the published version of the manuscript.

8.2. Data Availability Statement

The dataset used in this study, titled "Suicide Sentiment Analysis Dataset", is publicly available on Kaggle at https://www.kaggle.com/datasets/umar1103/suicide-sentiment-analysis-dataset. Further processed data and supplementary materials generated during the current study are available from the corresponding author upon reasonable request

8.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

8.4. Institutional Review Board Statement

Ethical review and approval were waived for this study because it did not involve direct interaction with human participants or the collection of identifiable personal data. The research was conducted using publicly available and fully anonymized datasets, as well as aggregated survey responses from professionals (psychologists, educators, and IT specialists), collected for analytical and methodological purposes only. No clinical interventions were performed, and no decisions affecting individual participants were made within the scope of this study. All procedures were carried out in accordance with ethical standards for research integrity and data protection.

8.5. Informed Consent Statement

Informed consent was obtained from all participants involved in the survey component of the study. Participation was voluntary, and all respondents were informed about the purpose of the research and the intended use of the collected data prior to completing the questionnaire. No personally identifiable information was collected, and all responses were analyzed in anonymized and aggregated form.

8.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

9. References

[1] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107–113. doi:10.1145/1327452.1327492.

[2] Dayalan, M. (2018). MapReduce: simplified data processing on large cluster. International Journal of Research and Engineering, 5(5), 399-403.

[3] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. 2nd USENIX Workshop on Hot Topics in Cloud Computing, HotCloud 2010, 1-10.

[4] Zabala-Vargas, S., Jaimes-Quintanilla, M., & Jimenez-Barrera, M. H. (2023). Big Data, Data Science, and Artificial Intelligence for Project Management in the Architecture, Engineering, and Construction Industry: A Systematic Review. Buildings, 13(12), 2944. doi:10.3390/buildings13122944.

[5] Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., & Abnousi, F. (2020). Artificial intelligence and suicide prevention: A systematic review of machine learning investigations. International Journal of Environmental Research and Public Health, 17(16), 1–25. doi:10.3390/ijerph17165929.

[6] Hawton, K., Hill, N. T. M., Gould, M., John, A., Lascelles, K., & Robinson, J. (2020). Clustering of suicides in children and adolescents. The Lancet Child and Adolescent Health, 4(1), 58–67. doi:10.1016/S2352-4642(19)30335-9.

[7] Kim, H., Son, Y., Lee, H., Kang, J., Hammoodi, A., Choi, Y., Kim, H. J., Lee, H., Fond, G., Boyer, L., Kwon, R., Woo, S., & Yon, D. K. (2024). Machine Learning-Based Prediction of Suicidal Thinking in Adolescents by Derivation and Validation in 3 Independent Worldwide Cohorts: Algorithm Development and Validation Study. Journal of Medical Internet Research, 26(1), 55913. doi:10.2196/55913.

[8] Méndez-Bustos, P., Fuster-Villaseca, J., Lopez-Castroman, J., Jiménez-Solomon, O., Olivari, C., & Baca-Garcia, E. (2022). Longitudinal trajectories of suicidal ideation and attempts in adolescents with psychiatric disorders in Chile: study protocol. BMJ Open, 12(2), 51749. doi:10.1136/bmjopen-2021-051749.

[9] Saduakassova, K., Zhanuzakov, M., Kassenova, G., & Serbin, V. (2024). Artificial Intelligence as a Tool to Prevent Autoaggressive Destructive Behavior among Children and Adolescents: A Brief Overview. Journal of Clinical Medicine of Kazakhstan, 21(6), 24–29. doi:10.23950/jcmk/15716.

[10] Rodway, C., Tham, S.-G., Ibrahim, S., Turnbull, P., Kapur, N., & Appleby, L. (2020). Children and young people who die by suicide: childhood-related antecedents, gender differences and service contact. BJPsych Open, 6(3), 49. doi:10.1192/bjo.2020.33.

[11] Jach, T., Magiera, E., & Froelich, W. (2015). Application of Hadoop to store and process big data gathered from an urban water distribution system. Procedia Engineering, 119(1), 1375–1380. doi:10.1016/j.proeng.2015.08.988.

[12] Lee, J., & Pak, T. Y. (2022). Machine learning prediction of suicidal ideation, planning, and attempt among Korean adults: A population-based study. SSM - Population Health, 19. doi:10.1016/j.ssmph.2022.101231.

[13] Merceedi, K. J., & Sabry, N. A. (2021). A Comprehensive Survey for Hadoop Distributed File System. Asian Journal of Research in Computer Science, 11(2), 46–57. doi:10.9734/ajrcos/2021/v11i230260.

[14] Assefi, M., Behravesh, E., Liu, G., & Tafti, A. P. (2017). Big data machine learning using apache spark MLlib. Proceedings - IEEE International Conference on Big Data, Big Data 2017, 2018-January, 3492–3498. doi:10.1109/BigData.2017.8258338.

[15] Armbrust, M., Xin, R. S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., Meng, X., Kaftan, T., Frankliny, M. J., Ghodsi, A., & Zaharia, M. (2015). Spark SQL: Relational data processing in spark. Proceedings of the ACM SIGMOD International Conference on Management of Data, 2015-May, 1383–1394. doi:10.1145/2723372.2742797.

[16] Gonzalez, J. E., Xin, R. S., Dave, A., Crankshaw, D., Franklin, M. J., & Stoica, I. (2014). GraphX: Graph processing in a distributed dataflow framework. Proceedings of the 11th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2014, 599–613.

[17] Ali, A. A., & Logofatu, D. (2022). An Analysis on Graph-Processing Frameworks: Neo4j and Spark GraphX. IFIP Advances in Information and Communication Technology, 646 IFIP, 461–470. doi:10.1007/978-3-031-08333-4_37.

[18] Bohaterewicz, B., Sobczak, A. M., Podolak, I., Wójcik, B., Mętel, D., Chrobak, A. A., Fąfrowicz, M., Siwek, M., Dudek, D., & Marek, T. (2021). Machine Learning-Based Identification of Suicidal Risk in Patients with Schizophrenia Using Multi-Level Resting-State fMRI Features. Frontiers in Neuroscience, 14, 605697. doi:10.3389/fnins.2020.605697.

[19] Su, C., Aseltine, R., Doshi, R., Chen, K., Rogers, S. C., & Wang, F. (2020). Machine learning for suicide risk prediction in children and adolescents with electronic health records. Translational Psychiatry, 10(1), 413. doi:10.1038/s41398-020-01100-0.

[20] Parsapoor, M., Koudys, J. W., & Ruocco, A. C. (2023). Suicide risk detection using artificial intelligence: the promise of creating a benchmark dataset for research on the detection of suicide risk. Frontiers in Psychiatry, 14. doi:10.3389/fpsyt.2023.1186569.

[21] Khosravi, H., Ahmed, I., & Choudhury, A. (2024). Predicting Suicidal Ideation, Planning, and Attempts among the Adolescent Population of the United States. Healthcare (Switzerland), 12(13), 1262. doi:10.3390/healthcare12131262.

[22] Bekmurat, O., Razaque, A., Khan, M. A., & Amanzholova, S. (2025). Prediction of Suicidal Tendencies in Adolescents Using Machine Learning Models and Artificial Intelligence. 2025 IEEE 1st Secure and Trustworthy Cyberinfrastructure for IoT and Microelectronics, SATC 2025 - Conference Proceedings, 1–5. doi:10.1109/SATC65530.2025.11136830.

[23] Marjit, U., Sharma, K., & Mandal, P. (2015). Data Transfers in Hadoop: A Comparative Study. Open Journal of Big Data, 1(2), 34–46.

[24] Hashem, I. A. T., Anuar, N. B., Gani, A., Yaqoob, I., Xia, F., & Khan, S. U. (2016). MapReduce: Review and open challenges. Scientometrics, 109(1), 389–422. doi:10.1007/s11192-016-1945-y.

[25] Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. European Physical Journal: Special Topics, 214(1), 481–518. doi:10.1140/epjst/e2012-01703-3.

[26] Vitulyova, Y., Babenko, T., Kolesnikova, K., Kiktev, N., & Abramkina, O. (2025). A Hybrid Approach Using Graph Neural Networks and LSTM for Attack Vector Reconstruction. Computers, 14(8), 301. doi:10.3390/computers14080301.

[27] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209. doi:10.1007/s11036-013-0489-0.

[28] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. doi:10.1561/2200000083.

[29] Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Computer Methods and Programs in Biomedicine, 226. doi:10.1016/j.cmpb.2022.107161.

[30] Saba Raoof, S., & Durai, M. A. S. (2022). A Comprehensive Review on Smart Health Care: Applications, Paradigms, and Challenges with Case Studies. Contrast Media and Molecular Imaging, 4822235. doi:10.1155/2022/4822235.

[31] Liu, B., Gu, J., & Wang, C. (2024). Research on smart city public health detection system and improvement technology based on intelligent multi-objective. Frontiers in Public Health, 12. doi:10.3389/fpubh.2024.1347586.

[32] Sarinova, A., Bekbayeva, A., Dunayev, P., Sarsikeyev, Y., & Sansyzbay, K. (2021). Hyperspectral Image Compression Algorithms for Phytosanitary Inspection of Agricultural Crops in Aerospace Photography. Journal of Theoretical and Applied Information Technology, 99(24), 6280–6290.

[33] Peponi, A., & Morgado, P. (2020). Smart and regenerative urban growth: A literature network analysis. International Journal of Environmental Research and Public Health, 17(7), 2463. doi:10.3390/ijerph17072463.

[34] Mohammadzadeh, Z., Saeidnia, H. R., Lotfata, A., Hassanzadeh, M., & Ghiasi, N. (2023). Smart city healthcare delivery innovations: a systematic review of essential technologies and indicators for developing nations. BMC Health Services Research, 23(1), 1180. doi:10.1186/s12913-023-10200-8.

[35] Fan, M., Sun, J., Zhou, B., & Chen, M. (2016). The Smart Health Initiative in China: The Case of Wuhan, Hubei Province. Journal of Medical Systems, 40(3), 1–17. doi:10.1007/s10916-015-0416-y.

[36] Alkina, A., Mekhtiyev, A., Neshina, Y., Serikov, T., Madi, P., Sansyzbay, K., & Yurchenko, A. (2022). Studying Additional Losses of Standard G.652 Optical Fiber with Protective Cladding during Multiple Bending To Develop Weight Control Sensor. Journal of Theoretical and Applied Information Technology, 100(7), 1983–1995.

[37] Alnasir, J. J., & Shanahan, H. P. (2018). The application of Hadoop in structural bioinformatics. Briefings in Bioinformatics, 21(1), 96–105. doi:10.1093/bib/bby106.

[38] Chen, Y., Alspaugh, S., & Katz, R. (2012). Interactive analytical processing in big data systems: A crossindustry study of mapreduce workloads. Proceedings of the VLDB Endowment, 5(12), 1802–1813. doi:10.14778/2367502.2367519.

[39] Zaharia, M., Das, T., Li, H., Hunter, T., Shenker, S., & Stoica, I. (2013). Discretized streams: Fault-tolerant streaming computation at scale. SOSP 2013 - Proceedings of the 24th ACM Symposium on Operating Systems Principles, 423–438. doi:10.1145/2517349.2522737.

[40] Essaidi, A., & Bellafkih, M. (2023). A New Big Data Architecture for Analysis: The Challenges on Social Media. International Journal of Advanced Computer Science and Applications, 14(3), 634–639. doi:10.14569/IJACSA.2023.0140373.

[41] Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop distributed file system. 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies, MSST2010, 1-10. doi:10.1109/MSST.2010.5496972.

[42] Bazrafshan, M., & Sayehmiri, K. (2024). Predicting suicidal behavior outcomes: an analysis of key factors and machine learning models. BMC Psychiatry, 24(1), 841. doi:10.1186/s12888-024-06273-2.

[43] Ehtemam, H., Sadeghi Esfahlani, S., Sanaei, A., Ghaemi, M. M., Hajesmaeel-Gohari, S., Rahimisadegh, R., Bahaadinbeigy, K., Ghasemian, F., & Shirvani, H. (2024). Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies. BMC Medical Informatics and Decision Making, 24(1), 138. doi:10.1186/s12911-024-02524-0.

[44] Zhu, X. (2021). Sample size calculation for Mann-Whitney U test with five methods. International Journal of Clinical Trials, 8(3), 184. doi:10.18203/2349-3259.ijct20212840.

[45] Aslam, M., & Smarandache, F. (2023). Chi-square test for imprecise data in consistency table. Frontiers in Applied Mathematics and Statistics, 9. doi:10.3389/fams.2023.1279638.

[46] Rashed, A. E. E., Atwa, A. E. M., Ahmed, A., Badawy, M., Elhosseini, M. A., & Bahgat, W. M. (2024). Facial image analysis for automated suicide risk detection with deep neural networks. Artificial Intelligence Review, 57(10), 274. doi:10.1007/s10462-024-10882-4.

[47] Morales, M., Dey, P., Theisen, T., Belitz, D., & Chernova, N. (2019). An Investigation of Deep Learning Systems for Suicide Risk Assessment. Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, 177–181. doi:10.18653/v1/w19-3023.

[48] Akuma, S., Lubem, T., & Adom, I. T. (2022). Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets. International Journal of Information Technology (Singapore), 14(7), 3629–3635. doi:10.1007/s41870-022-01096-4.

[49] Yadav, V., Verma, P., & Katiyar, V. (2023). Long short-term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages. International Journal of Information Technology (Singapore), 15(2), 759–772. doi:10.1007/s41870-022-01010-y.

[50] Glenn, C. R., Kleiman, E. M., Kellerman, J., Pollak, O., Cha, C. B., Esposito, E. C., Porter, A. C., Wyman, P. A., & Boatman, A. E. (2020). Annual Research Review: A meta-analytic review of worldwide suicide rates in adolescents. Journal of Child Psychology and Psychiatry and Allied Disciplines, 61(3), 294–308. doi:10.1111/jcpp.13106.

[51] Hughes, J. L., Horowitz, L. M., Ackerman, J. P., Adrian, M. C., Campo, J. V., & Bridge, J. A. (2023). Suicide in young people: screening, risk assessment, and intervention. Bmj, 381, 70630. doi:10.1136/bmj-2022-070630.

[52] Rappai, S., & Ramasamy, G. (2024). Navigating the emotional maze: Understanding Adolescent suicidal ideation using CNN-LSTM model. Intelligent Decision Technologies, 18(3), 1797–1811. doi:10.3233/IDT-240790.

[53] Banik, T., Asgar, S., Hosen, M. H., Uddin, A., Nawar, S., & Saha, A. (2024). From Social Media to Mental Health Insights: A Hybrid CNN-LSTM Model for Depression Detection in Bangladesh. 2024 IEEE Conference on Computing Applications and Systems, COMPAS 2024. doi:10.1109/COMPAS60761.2024.10797102.

[54] Sansyzbay, K. M., Bakhtiyarova, Y. A., Iliev, T., Patokin, G. S., Tasbolatova, L. T., & Sagmedinov, D. B. (2024). Development of an Algorithm for a National Microprocessor-Based Centralization System with a Modular Architecture KZ-MPC-MA Featuring Advanced Intelligent Control Functions. IEEE Access, 12, 193229–193240. doi:10.1109/ACCESS.2024.3521219.

[55] Goel, R., & Digalwar, M. (2024). Suicidal Thought Detection using Max Voting Ensemble Technique. Procedia Computer Science, 235, 2587–2598. doi:10.1016/j.procs.2024.04.244.

[56] Young, J., Bishop, S., Humphrey, C., & Pavlacic, J. M. (2023). A review of natural language processing in the identification of suicidal behavior. Journal of Affective Disorders Reports, 12. doi:10.1016/j.jadr.2023.100507.

[57] Feroze, S. A., Feroze, S. B., & Abbasi, U. (2024). The Unveiling Distress: Harnessing NLP and Deep Learning to Identify Suicidal Signals in Tweets. International Journal of Technology, Innovation and Management, 4(1), 20–31. doi:10.54489/ymy5bp91.

[58] Ji, S. (2022). Towards Intention Understanding in Suicidal Risk Assessment with Natural Language Processing. In Findings of the Association for Computational Linguistics: EMNLP 2022, 4057–4067. EMNLP. doi:10.18653/v1/2022.findings-emnlp.297.

[59] Bialer, A., Izmaylov, D., Segal, A., Tsur, O., Levi-Belz, Y., & Gal, K. (2022). Detecting Suicide Risk in Online Counseling Services: A Study in a Low-Resource Language. Proceedings - International Conference on Computational Linguistics, COLING, 29(1), 4241–4250. doi:10.48550/arXiv.2209.04830.

[60] Cliffe, C., Seyedsalehi, A., Vardavoulia, K., Bittar, A., Velupillai, S., Shetty, H., Schmidt, U., & Dutta, R. (2021). Using natural language processing to extract self-harm and suicidality data from a clinical sample of patients with eating disorders: A retrospective cohort study. BMJ Open, 11(12), e053808. doi:10.1136/bmjopen-2021-053808.