Development of an AI-Based Statistical Clinical Decision Support Model for Early Detection of High-Risk Patients in Community Nursing

Mr. Pranjal Kaser[1], Mrs. Nishi Kashyap[2], Mrs. Madhuri Nair[3], Mrs. Shraddha Pradeep Kumar[4]

[1] Statistician, SoS in Statistics, Pt. Ravishankar Shukla University, Raipur, C.G., India.
[2] Associate Professor, Mental health Nursing, Govt. College of Nursing, Rajnandgaon, C.G., India.
[3] Associate Professor, Ob.G. Nursing, Govt. College of Nursing, Rajnandgaon, C.G., India.
[4] Demonstrator Med. Surg. Nursing, Govt. College of Nursing, Rajnandgaon, C.G., India.

Corresponding Author Email: kaserpranjal@gmail.com | ORCID: https://orcid.org/0009-0002-5976-1106

How to Cite this Article:

Kaser, P., Kashyap, N., Nair, M. & Kumar, S. P. (2026). Development of an AI-Based Statistical Clinical Decision Support Model for Early Detection of High-Risk Patients in Community Nursing. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(8), 1-9. https://doi.org/10.55041/ijcope.v2i8.093

https://doi.org/10.55041/ijcope.v2i8.093

Abstract-

Early identification of high-risk patients is one of the most important responsibilities of community nursing because timely intervention reduces disease progression, hospitalization, healthcare costs, and mortality. Traditional risk assessment methods often rely on manual scoring systems and subjective clinical judgment, which may fail to detect subtle relationships among multiple patient characteristics. Recent advances in Artificial Intelligence (AI) and statistical learning provide opportunities to develop intelligent Clinical Decision Support Systems (CDSS) capable of accurately identifying individuals at high risk.

The present study proposes an AI-based Statistical Clinical Decision Support Model integrating classical statistical techniques and machine learning algorithms for early detection of high-risk patients in community nursing settings. The proposed framework combines logistic regression, Random Forest, Gradient Boosting, and Artificial Neural Networks with statistical feature selection techniques. Community health records containing demographic, clinical, behavioral, socioeconomic, and environmental variables are used for model development.

The model performance is evaluated using Accuracy, Sensitivity, Specificity, Precision, Recall, F1-score, ROC-AUC, and calibration statistics. Explainable Artificial Intelligence (XAI) methods such as SHAP values are incorporated to improve transparency and support nursing decision-making. The proposed model is expected to assist community nurses in identifying vulnerable patients earlier, prioritizing home visits, optimizing resource allocation, and improving preventive healthcare outcomes.

Keywords- Community Nursing, Clinical Decision Support System, Machine Learning, Logistic Regression, Random Forest, Early Detection, Healthcare Analytics, Predictive Modeling.

I. INTRODUCTION

Community nursing is a vital component of primary healthcare that focuses on disease prevention, health promotion, early diagnosis, and continuous patient care within the community. Community nurses are often the first healthcare professionals to assess individuals in homes, schools, and primary health centers. Their ability to identify patients who are at high risk of developing serious health conditions is essential for preventing disease progression, reducing hospital admissions, lowering healthcare costs, and improving the overall quality of life.[24]

In recent years, the growing burden of chronic diseases such as diabetes, hypertension, cardiovascular diseases, chronic respiratory disorders, and mental health problems has created significant challenges for healthcare systems worldwide. According to the World Health Organization (WHO), non-communicable diseases account for the majority of global deaths, many of which could be prevented through timely screening and early intervention. As the number of patients continues to increase while healthcare resources remain limited, community nurses require efficient and evidence-based tools to support clinical decision-making.

Traditionally, patient risk assessment in community nursing has relied on clinical guidelines, manual scoring systems, and the professional judgment of healthcare providers. These methods are useful but often depend on a limited number of variables and may not accurately capture the complex interactions among demographic, clinical, behavioral, and socioeconomic factors that influence patient health. Manual assessment is also time-consuming and may produce inconsistent results due to variations in clinical experience.

The rapid advancement of Artificial Intelligence (AI) has introduced new opportunities to improve healthcare decision-making. AI refers to computer systems that can analyze large amounts of data, recognize patterns, and make predictions to assist healthcare professionals. One of the most important applications of AI in healthcare is Machine Learning (ML), which enables computers to learn from historical patient data and predict future health outcomes without being explicitly programmed. Machine learning algorithms have demonstrated excellent performance in disease prediction, patient risk stratification, hospital admission forecasting, and clinical outcome prediction.

One of the major applications of AI in healthcare is the Clinical Decision Support System (CDSS). A CDSS is a computerized system that assists healthcare professionals by providing evidence-based recommendations, identifying high-risk patients, and supporting clinical decision-making. AI-powered CDSS can analyze multiple patient characteristics simultaneously, including age, gender, vital signs, laboratory reports, medical history, lifestyle factors, medication adherence, and socioeconomic conditions, to generate individualized risk predictions. Such systems help healthcare providers make faster, more accurate, and consistent clinical decisions.[26]

Despite their high predictive accuracy, many AI models operate as "black-box" systems, meaning they provide predictions without explaining how those predictions are made. This lack of transparency reduces healthcare professionals' confidence in using AI for clinical decisions. Therefore, integrating Explainable Artificial Intelligence (XAI) techniques is becoming increasingly important. Explainable AI methods, such as SHAP (SHapley Additive exPlanations), identify the contribution of each predictor variable to the final prediction, allowing nurses to understand why a patient has been classified as high risk and increasing trust in AI-assisted decision-making.

Statistical methods continue to play a fundamental role in healthcare research and predictive modeling. Techniques such as descriptive statistics, hypothesis testing, logistic regression, and multivariable analysis help identify significant risk factors and quantify their effects on patient outcomes. Logistic regression, in particular, is widely used because it provides interpretable probability estimates and allows clinicians to understand the relationship between predictor variables and disease risk. However, statistical models may have limited predictive capability when dealing with complex nonlinear relationships and high-dimensional healthcare data.[27]

Machine learning algorithms, including Decision Trees, Random Forests, Gradient Boosting, Support Vector Machines, and Artificial Neural Networks, can overcome many of these limitations by modeling nonlinear relationships and complex interactions among variables. These algorithms often achieve higher prediction accuracy than traditional statistical models. However, they are generally less interpretable. Therefore, combining statistical techniques with machine learning provides a balanced approach that offers both high predictive performance and clinical interpretability.

In community nursing, early identification of high-risk patients enables timely preventive interventions such as health education, home visits, medication management, nutritional counseling, and referral to specialized healthcare services. Early risk prediction also supports better resource allocation, reduces unnecessary hospitalizations, and improves patient outcomes. AI-based predictive models can further assist healthcare administrators in planning community health programs and prioritizing healthcare services for vulnerable populations.[25]

The present study proposes the development of an AI-Based Statistical Clinical Decision Support Model for Early Detection of High-Risk Patients in Community Nursing. The proposed model integrates statistical methods with advanced machine learning algorithms to identify patients who are at increased risk of adverse health outcomes. The model utilizes patient demographic information, clinical measurements, medical history, lifestyle characteristics, and socioeconomic factors to generate individualized risk scores. In addition, explainable AI techniques are incorporated to ensure transparency and improve the usability of the model in nursing practice.[28]

The proposed decision support system is expected to assist community nurses in making timely, accurate, and evidence-based clinical decisions. By facilitating early detection of high-risk patients, the model has the potential to improve preventive healthcare, optimize the use of healthcare resources, reduce disease complications, and enhance the overall quality of community nursing services. Furthermore, the study contributes to the growing field of intelligent healthcare systems by demonstrating how statistical methods and artificial intelligence can be integrated to support evidence-based nursing practice and strengthen primary healthcare delivery.

II. Background

Community nursing serves as the cornerstone of preventive and primary healthcare by facilitating the early identification and management of individuals at risk of adverse health outcomes. During routine patient assessments, community nurses collect a wide range of health-related information that reflects an individual's overall health status. These assessments typically include physiological measurements such as blood pressure, blood glucose level, body mass index (BMI), oxygen saturation (SpO_2), body temperature, and heart rate, along with important clinical and personal information including medical history, medication adherence, nutritional status, family history, lifestyle behaviors (e.g., smoking, alcohol consumption, physical activity), and socioeconomic conditions.

Each of these factors contributes significantly to a patient's health profile, and their combined effects determine the likelihood of developing chronic diseases, disease complications, hospitalization, or other adverse clinical outcomes. However, evaluating multiple variables simultaneously is a complex and time-consuming task, particularly in community healthcare settings where nurses often manage large populations with limited resources. Traditional manual assessment methods primarily rely on clinical experience, standard guidelines, and individual interpretation, making it difficult to recognize complex interactions among numerous risk factors and potentially leading to delayed or inconsistent clinical decisions.

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have created new opportunities to address these challenges. AI-based predictive models can simultaneously process large volumes of multidimensional patient data, identify hidden patterns, and accurately estimate an individual's risk of developing serious health conditions. Unlike traditional methods, machine learning algorithms can capture complex, nonlinear relationships among clinical, demographic, behavioral, and socioeconomic variables, enabling more accurate and personalized risk prediction.[29]

At the same time, statistical models remain essential for healthcare research and clinical decision-making because they provide interpretable estimates of disease risk and identify significant predictors associated with patient outcomes. Statistical techniques such as logistic regression, odds ratios, confidence intervals, and hypothesis testing offer transparency and help clinicians understand the influence of individual risk factors on prediction results.

III. OBJECTIVES

The objectives of the study are:

1. To identify significant predictors associated with high-risk patients.
2. To develop an AI-based statistical prediction model.
3. To compare machine learning algorithms for prediction accuracy.
4. To develop an explainable Clinical Decision Support System.
5. To evaluate model performance using statistical validation methods.

IV. RESEARCH QUESTIONS

1. Which clinical variables significantly influence patient risk?
2. Which machine learning algorithm provides the highest predictive accuracy?
3. Does integrating statistical analysis improve AI performance?
4. Can explainable AI increase nurses' trust in AI recommendations?

5. Hypotheses

H_{01}: The developed AI model does not significantly improve prediction accuracy compared to conventional statistical models.

H_{02}: There is no significant relationship between selected patient characteristics and predicted clinical risk.

V. LITERATURE REVIEW

1. Barredo Arrieta et al. (2019) developed a comprehensive taxonomy of Explainable Artificial Intelligence (XAI) by reviewing existing explainability methods across AI domains. The study emphasized the principles of fairness, accountability, transparency, and interpretability, establishing a foundation for the development of trustworthy AI systems in healthcare.

2. Kim et al. (2022) developed a machine learning-based Clinical Decision Support System (CDSS) for pressure ulcer prevention using hospital patient records. The study demonstrated that the AI system improved nurses' adherence to preventive guidelines, enhanced documentation quality, and reduced the incidence of preventable pressure injuries compared to conventional assessment methods.

3. Bharati et al. (2023) conducted a PRISMA-based systematic review on Explainable AI applications in healthcare. The authors concluded that transparency, fairness, clinician involvement, and accountability are essential for increasing trust and promoting the successful adoption of AI technologies in clinical practice.

4. Prentzas et al. (2023) reviewed 198 studies on Explainable AI in medicine and found that deep learning models dominate healthcare applications. However, the review highlighted limited participation of physicians and nurses during AI development, emphasizing the need for user-centered explainable systems.

5. Islam et al. (2023) systematically reviewed machine learning methods for early sepsis prediction using Electronic Health Records (EHRs). The study reported that Random Forest and XGBoost consistently outperformed traditional clinical scoring systems by providing earlier and more accurate identification of sepsis.

6. Moazemi et al. (2023) reviewed AI applications for monitoring cardiovascular patients in intensive care settings. Their findings showed that machine learning and deep learning algorithms improved prediction of cardiovascular deterioration and supported timely clinical intervention by integrating Electronic Health Records with continuous patient monitoring.

7. Sivaraman et al. (2023) investigated clinicians' acceptance of AI-based treatment recommendations through a human-AI interaction study. The results indicated that clinicians were more willing to trust AI recommendations when clear explanations were provided, highlighting the importance of Explainable AI in healthcare decision-making.

8. Zhang et al. (2023) proposed the "SepsisLab" framework to support collaborative decision-making between clinicians and AI during sepsis diagnosis. The system improved diagnostic confidence by visualizing prediction uncertainty and recommending additional laboratory investigations before final clinical decisions.

9. Kim et al. (2024) reviewed Explainable AI-based Clinical Decision Support Systems and found that SHAP and LIME were the most widely used explanation techniques. The study concluded that Explainable AI significantly improves transparency, clinician trust, and interpretability of machine learning predictions in healthcare.

10. Alenazi et al. (2024) systematically evaluated AI-powered Clinical Decision Support Systems in nursing practice. The review reported improvements in diagnostic accuracy, medication management, treatment planning, and patient safety, while also emphasizing ethical issues such as privacy, transparency, and algorithmic bias.

11. Chua, Rusli, and Aitken (2024) compared Early Warning Score (EWS), qSOFA, and SIRS for sepsis identification through a systematic review and meta-analysis. The authors found that EWS performed comparably or better than conventional scoring systems but suggested that AI-based prediction models could further improve diagnostic performance.

12. Tądel, Dudek, and Bil-Lula (2024) reviewed AI algorithms for sepsis risk prediction and reported that ensemble machine learning methods such as Random Forest, XGBoost, and Neural Networks achieved superior predictive performance compared with traditional statistical models. The review also highlighted AI's potential for personalized treatment recommendations.

13. Khokhar, Gravino, and Palomba (2024) conducted a systematic review of Artificial Intelligence techniques for diabetes prediction. The study concluded that ensemble algorithms, particularly XGBoost and Support Vector Machines, consistently outperformed conventional statistical models and stressed the importance of ethical AI and high-quality healthcare data.

14. Yu and Son (2024) evaluated machine learning models for predicting 30-day readmission among heart failure patients through a systematic review. Random Forest and Gradient Boosting demonstrated high predictive accuracy, while demographic, laboratory, medication, and comorbidity variables were identified as the strongest predictors of readmission.

15. Zhang et al. (2024) critically reviewed prediction models for coronary artery disease readmission and found that existing models showed moderate predictive ability but limited external validation. The authors recommended integrating machine learning with Clinical Decision Support Systems to improve risk prediction and patient management.

16. Khokhar, Gravino, and Palomba (2024) further emphasized that machine learning algorithms provide more accurate diabetes prediction than traditional statistical approaches. Their review also highlighted the importance of interdisciplinary collaboration and ethical implementation for successful deployment of AI-based healthcare systems.

17. Korom et al. (2025) evaluated a Large Language Model (LLM)-based Clinical Decision Support System in primary healthcare using more than 39,000 patient consultations. The AI-assisted system reduced diagnostic errors by approximately 16% and treatment errors by 13%, while clinicians reported greater confidence in their clinical decisions.

18. Buess et al. (2025) reviewed multimodal generative AI applications in medicine and reported that integrating structured clinical records, laboratory findings, medical imaging, and clinical text significantly improved diagnostic accuracy and clinical decision-making compared with single-source AI models.

19. Abbas et al. (2025) conducted a systematic review and meta-analysis on AI-based models for early sepsis detection. The findings demonstrated that machine learning and deep learning algorithms achieved high predictive accuracy for identifying sepsis at an early stage, supporting timely intervention and improved patient outcomes.

20. Ji, Huo, and Dong (2025) performed a meta-analysis involving 16 studies and 159,947 patients to evaluate AI for intensive care unit sepsis detection. The study reported high sensitivity, specificity, and an AUC of approximately 0.87, confirming the effectiveness of AI in early sepsis diagnosis.

21. Yuan et al. (2025) evaluated AI-powered Early Warning Systems through a meta-analysis and found that these systems significantly improved early recognition of clinical deterioration, reduced mortality, and enhanced patient outcomes compared with conventional monitoring approaches.

22. Okere et al. (2025) developed machine learning models to predict hospitalization and 90-day readmission among adults with cardiovascular risk factors in community settings. The Extra Trees algorithm achieved excellent predictive performance, while SHAP analysis identified major clinical and socioeconomic predictors influencing patient outcomes.

23. Havranek et al. (2025) developed machine learning models for predicting unplanned hospital readmissions using multisite Electronic Health Record data. Random Forest and Gradient Boosting improved prediction accuracy compared with traditional approaches, demonstrating the value of personalized risk prediction.

24. Hu et al. (2025) conducted a systematic review and meta-analysis on AI models for ICU readmission prediction. The study concluded that machine learning models achieved higher sensitivity and specificity than conventional scoring systems, supporting better identification of patients at risk of readmission.

25. Brothers et al. (2025) investigated synthetic data-augmented machine learning models for predicting 30-day hospital readmission. The study found that combining synthetic datasets with SHAP-based explanations improved both model robustness and interpretability while identifying important clinical and social determinants of readmission.

26. Wang et al. (2025) compared Logistic Regression with several machine learning algorithms for predicting readmission among elderly patients with diabetes and heart failure. Ensemble learning models outperformed logistic regression, and SHAP analysis identified age, renal function, BNP levels, and comorbidities as the strongest predictors.

27. Alnomasy et al. (2025) systematically reviewed machine learning models for heart failure readmission prediction. The authors reported that Random Forest, XGBoost, and Neural Networks consistently produced better predictive performance than conventional statistical models and recommended their integration into clinical workflows.

28. Mikkonen et al. (2026) systematically reviewed AI technologies supporting nurses' clinical decision-making. The review demonstrated that AI-based Clinical Decision Support Systems improved nursing documentation, workflow efficiency, patient safety, early deterioration detection, and reduced hospital readmissions, highlighting the growing role of AI in evidence-based nursing practice.

VI. METHODOLOGY

6.1 Research Design

This study adopted a **quantitative methodological research design** to develop and validate an AI-based Statistical Clinical Decision Support Model (CDSS) for the early detection of high-risk patients in community nursing. The model integrates statistical techniques with machine learning algorithms to improve risk prediction and support clinical decision-making.

6.2 Data Source

The study used **secondary data** from the **National Family Health Survey (NFHS-5), 2019–21**, conducted by the **International Institute for Population Sciences (IIPS), Mumbai**, under the Ministry of Health and Family Welfare, Government of India. NFHS-5 is a nationally representative survey containing demographic, clinical, lifestyle, and socioeconomic information suitable for predictive health analytics.

Data Source:

International Institute for Population Sciences (IIPS)
Govandi Station Road, Deonar, Mumbai–400088,
India
Website: http://www.rchiips.org/nfhs

6.3 Study Population

The study population comprised adults aged **18 years and above** included in the NFHS-5 database.

6.4 Sample Size

5,000 eligible records were selected from the NFHS-5 dataset after applying the inclusion and exclusion criteria.

6.5 Inclusion Criteria

- Adults aged ≥18 years.
- Complete demographic and clinical records.
- Complete outcome information.

6.6 Exclusion Criteria

- Duplicate records.
- Incomplete records.
- Missing outcome variable.

6.7. Study Variables

Dependent Variable

High-Risk Patient Status

- **0 = Low Risk**
- **1 = High Risk**

Independent Variables: Socio demographic variables.

6.8 Statistical Analysis

Descriptive statistics including frequency, percentage, mean, median, and standard deviation were used to summarize the study variables.

Inferential analysis included the Chi-square test, Independent t-test, ANOVA, correlation analysis, and binary logistic regression to identify significant predictors of high-risk patient status. Model performance was evaluated using Odds Ratio (OR), 95% Confidence Interval (CI), Hosmer–Lemeshow goodness-of-fit test, Receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC), calibration plot, and 10-fold cross-validation.

VII. MACHINE LEARNING ALGORITHMS

To develop the proposed AI-based Clinical Decision Support Model, **Logistic Regression** and **Random Forest** algorithms were employed. These algorithms were selected because they provide a balance between interpretability and predictive performance, making them suitable for healthcare risk prediction.

7.1 Logistic Regression

Logistic Regression was used as the baseline statistical model for predicting the probability of a patient being classified as **high risk** or **low risk**. It estimates the relationship between the dependent variable and multiple predictor variables and identifies significant risk factors through regression coefficients and odds ratios. Due to its simplicity and interpretability, Logistic Regression is widely used in clinical research and healthcare decision-making.

7.2 Random Forest

Random Forest is an ensemble machine learning algorithm that constructs multiple decision trees and combines their predictions to improve classification accuracy. It effectively captures complex nonlinear relationships among predictor variables, reduces overfitting, and handles high-dimensional healthcare data. In this study, Random Forest was used to improve the predictive performance of the Clinical Decision Support Model and to identify the most important variables associated with high-risk patient status.

VIII. MATHEMATICAL MODEL

The proposed AI-based Clinical Decision Support Model integrates **Logistic Regression** as the baseline statistical model and **Random Forest** as the ensemble machine learning model for predicting the probability of a patient being classified as **high risk** or **low risk**.

8.1 Logistic Regression Model

Let

$$Y = \begin{cases} 1, & \text{High-Risk Patient} \\ 0, & \text{Low-Risk Patient} \end{cases}$$

Suppose $X = (X_1, X_2, \ldots, X_p)$ represents the vector of predictor variables.

For this study,
$X =${Age, Gender, BMI, Diabetes, Hypertension, Smoking, Alcohol, Physical Activity, Hemoglobin, Blood Pressure, Socioeconomic Status, Family History, Nutritional Status, Previous Hospitalization}

Step 1: Probability of High Risk

The probability that an individual belongs to the high-risk group is

$$P(Y = 1|X) = \pi(X)$$

where

$$0 < \pi(X) < 1$$

Step 2: Odds of High Risk

The odds of becoming a high-risk patient are

$$\frac{\pi(X)}{1 - \pi(X)}$$

where

- Numerator = Probability of High Risk
- Denominator = Probability of Low Risk

Step 3: Logit Transformation

Taking natural logarithm,

$$\log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

This is called the **Logit Model**.

Step 4: Logistic Regression Equation

After rearranging,

$$\boxed{P(Y=1|X) = \frac{1}{1+\exp\left[-(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)\right]}}$$

where

- (β_0)= Intercept
- (β_i) = Regression coefficient
- (X_i) = Predictor variable
- $\left(P(Y=1)\right)$ = Probability of High Risk

The coefficients (β_i) are estimated using the **Maximum Likelihood Estimation (MLE)** method.

8.2 Random Forest Model

Random Forest is an ensemble learning algorithm that constructs multiple decision trees using bootstrap samples of the training data.

Assume

$$T_1, T_2, \ldots, T_B$$

are (B) independently grown decision trees.

Each tree predicts $\quad T_b(X) \in 0,1$

where

- 0 = Low Risk
- 1 = High Risk

The final prediction is obtained by majority voting:

$$\boxed{\hat{Y} = \text{Mode}\{T_1(X), T_2(X), \ldots, T_B(X)\}}$$

or equivalently,

$$\hat{Y} = \arg\max_{c} \sum_{b=1}^{B} I\left(T_b(X) = c\right)$$

where

- $(I(\cdot))$= Indicator function
- (c)= Class (0 or 1)
- (B)= Number of trees

The class receiving the maximum number of votes is selected as the final prediction.

8.3 Feature Importance

Random Forest estimates the importance of each predictor by measuring the reduction in node impurity produced by that variable across all trees.

The importance of the (j^{th})predictor is

$$\boxed{FI_j = \sum_{b=1}^{B} \sum_{k=1}^{N_b} \Delta I_{jbk}}$$

where

- (FI_j)= Importance of predictor (j)
- (B)= Number of trees
- (N_b)= Number of nodes in tree (b)
- (ΔI_{jbk}) = Reduction in impurity (Gini Index or Entropy) produced by variable (j) at node (k)

Higher values of (FI_j)indicate greater contribution of that predictor to the classification model.

8.4 Model Prediction Rule

For each patient, both Logistic Regression and Random Forest estimate the probability of belonging to the high-risk category.

The classification rule is

$$\hat{Y} = \begin{cases} 1, & P(Y=1) \geq 0.5 \\ 0, & P(Y=1) < 0.5 \end{cases}$$

where

- **1 = High-Risk Patient**
- **0 = Low-Risk Patient**

IX. DATA ANALYSIS

Table 1. Descriptive Statistics of Study Participants (N = 4,000)

Variable	Mean ± SD (%)
Age (Years)	46.82 ± 15.36
Male	1,932 (48.3%)
Female	2,068 (51.7%)
BMI (kg/m²)	24.78 ± 4.61
Hypertension	1,148 (28.7%)
Diabetes	624 (15.6%)
Smoking	812 (20.3%)
Alcohol Consumption	694 (17.4%)
Previous Hospitalization	498 (12.5%)
High Risk Patients	1,082 (27.1%)
Low Risk Patients	2,918 (72.9%)

Interpretation: Table 1 presents the descriptive characteristics of the 4,000 study participants included in the analysis. The mean age of the participants was 46.82 ± 15.36 years, indicating a diverse adult population. Females constituted a slightly higher proportion (51.7%) than males (48.3%). The average Body Mass Index (BMI) was 24.78 ± 4.61 kg/m², suggesting that the study population was, on average, within the normal to overweight range. Among the participants, 28.7% had hypertension and 15.6% had diabetes, while 20.3% reported smoking and 17.4% reported alcohol consumption. Additionally, 12.5% of the participants had a history of previous hospitalization. Based on the developed risk prediction model, 27.1% of the participants were classified as high-risk, whereas 72.9% were classified as low-risk. Overall, the findings indicate that a considerable proportion of the study population had chronic disease risk factors, supporting the need for an AI-based Clinical Decision Support Model to facilitate early identification and timely intervention for high-risk individuals in community nursing.

Table 2. Comparison of Variables Between High-Risk and Low-Risk Patients

Variable	High Risk (n=1082)	Low Risk (n=2918)	p-value
Age	58.3 ± 11.2	41.7 ± 13.5	<0.001
BMI	27.1 ± 4.3	23.9 ± 3.8	<0.001
Hemoglobin	10.8 ± 1.7	12.9 ± 1.4	<0.001
Blood Pressure	149 ± 18	122 ± 12	<0.001
Diabetes (%)	43.6	5.3	<0.001
Hypertension (%)	62.5	16.3	<0.001

Interpretation: Table 2 compares the clinical and demographic characteristics of high-risk (n = 1,082) and low-risk (n = 2,918) patients. The results show that high-risk patients were significantly older than low-risk patients (58.3 ± 11.2 vs. 41.7 ± 13.5 years, $p < 0.001$). They also had a higher mean BMI (27.1 ± 4.3 vs. 23.9 ± 3.8 kg/m², $p < 0.001$), indicating a greater prevalence of overweight and obesity. The mean hemoglobin level was significantly lower among high-risk patients (10.8 ± 1.7 g/dL) compared to low-risk patients (12.9 ± 1.4 g/dL, $p < 0.001$), suggesting poorer health status. Similarly, the average blood pressure was markedly higher in the high-risk group (149 ± 18 mmHg) than in the low-risk group (122 ± 12 mmHg, $p < 0.001$). The prevalence of diabetes (43.6% vs. 5.3%) and hypertension (62.5% vs. 16.3%) was also significantly higher among high-risk patients ($p < 0.001$). These findings indicate that age, BMI, hemoglobin level, blood pressure, diabetes, and hypertension are strongly associated with high-risk patient status and are important predictors for the proposed AI-based Clinical Decision Support Model.

Table 3. Chi-square Test for Association Between Risk Status and Predictor Variables

Variable	χ^2	p-value
Gender	8.73	0.003
Smoking	24.68	<0.001
Alcohol Consumption	16.94	<0.001
Diabetes	261.42	<0.001
Hypertension	387.61	<0.001
Physical Activity	59.82	<0.001
Nutritional Status	71.44	<0.001

Interpretation: Table 3 presents the results of the Chi-square test examining the association between risk status and selected predictor variables. The findings indicate that all the selected variables were significantly associated with patient risk status ($p < 0.05$). Gender showed a statistically significant association with risk status ($\chi^2 = 8.73$, $p = 0.003$). Stronger associations were observed for smoking ($\chi^2 = 24.68$, $p < 0.001$), alcohol consumption ($\chi^2 = 16.94$, $p < 0.001$), physical activity ($\chi^2 = 59.82$, $p < 0.001$), and nutritional status ($\chi^2 = 71.44$, $p < 0.001$). Among all the variables, hypertension ($\chi^2 = 387.61$, $p < 0.001$) exhibited the strongest association with high-risk status, followed by diabetes ($\chi^2 = 261.42$, $p < 0.001$), indicating that these chronic conditions are the most influential predictors of clinical risk. Overall, the results demonstrate that demographic, behavioral, and clinical characteristics are significantly related to the prediction of high-risk patients, thereby supporting the inclusion of these variables in the proposed AI-based Clinical Decision Support Model.

Table 4. Logistic Regression Analysis

Variable	β	SE	Odds Ratio (95% CI)	p-value
Age	0.051	0.006	1.05 (1.04–1.07)	<0.001
BMI	0.072	0.013	1.07 (1.04–1.10)	<0.001
Diabetes	1.34	0.16	3.82 (2.81–5.17)	<0.001
Hypertension	1.81	0.14	6.11 (4.66–8.03)	<0.001
Smoking	0.41	0.13	1.51 (1.18–1.93)	0.002
Previous Hospitalization	0.89	0.15	2.43 (1.82–3.26)	<0.001

Model Statistics

- Hosmer–Lemeshow = **χ^2 = 6.81, p = 0.56**
- Nagelkerke R^2 = **0.48**
- AUC = **0.872**

Interpretation: Table 4 presents the results of the binary Logistic Regression analysis for identifying factors associated with high-risk patient status. The findings indicate that all the selected predictor variables were statistically significant ($p < 0.05$). Age ($\beta = 0.051$, OR = 1.05, 95% CI: 1.04–1.07, $p < 0.001$) and BMI ($\beta = 0.072$, OR = 1.07, 95% CI: 1.04–1.10, $p < 0.001$) were positively associated with the likelihood of being classified as a high-risk patient, suggesting that increasing age and higher BMI increase the probability of clinical risk. Patients with diabetes were 3.82 times more likely to be classified as high risk (OR = 3.82, 95% CI: 2.81–5.17, $p < 0.001$), while those with hypertension had the strongest association, being 6.11 times more likely to belong to the high-risk group (OR = 6.11, 95% CI: 4.66–8.03, $p < 0.001$). Similarly, smoking (OR = 1.51, $p = 0.002$) and previous hospitalization (OR = 2.43, $p < 0.001$) were also identified as significant predictors of high-risk status. The model demonstrated a good fit to the data, as indicated by the Hosmer–Lemeshow goodness-of-fit test ($\chi^2 = 6.81$, $p = 0.56$), suggesting no significant difference between observed and predicted outcomes. Furthermore, the Nagelkerke R^2 value of 0.48 indicates that approximately 48% of the variation in high-risk patient status was explained by the predictor variables included in the model. The model also showed good discriminatory ability with an Area Under the ROC Curve (AUC) of 0.872, indicating excellent performance in distinguishing between high-risk and low-risk patients. Overall, these findings confirm that the selected demographic and clinical variables are significant predictors of patient risk and demonstrate the effectiveness of Logistic Regression as a baseline predictive model for the proposed AI-based Clinical Decision Support System.

Table 5. Random Forest Variable Importance

Rank	**Variable**	**Importance Score**
1	Hypertension	0.214
2	Age	0.181
3	Diabetes	0.173
4	Blood Pressure	0.146
5	BMI	0.112
6	Hemoglobin	0.081
7	Physical Activity	0.049
8	Smoking	0.044

Interpretation: Table 5 presents the variable importance scores obtained from the Random Forest model, indicating the relative contribution of each predictor in identifying high-risk patients. Among all the variables, hypertension emerged as the most influential predictor with the highest importance score (0.214), followed by age (0.181) and diabetes (0.173), highlighting their critical role in risk prediction. Blood pressure (0.146) and BMI (0.112) also made substantial contributions to the model, suggesting that cardiovascular and metabolic health indicators are key determinants of patient risk. Moderate importance was observed for hemoglobin (0.081), while physical activity (0.049) and smoking (0.044) had comparatively lower, yet meaningful, contributions to the prediction model. Overall, the Random Forest analysis demonstrates that both clinical and lifestyle-related factors contribute to the accurate classification of high-risk patients, with hypertension, age, and diabetes being the most influential predictors. These findings support the effectiveness of the proposed AI-based Clinical Decision Support Model in prioritizing the most significant risk factors for early identification and timely intervention in community nursing.

Table 6. Performance Comparison of Machine Learning Models

Performance Metric	Logistic Regression	Random Forest
Accuracy	84.7%	**91.5%**
Precision	82.1%	**90.2%**
Recall (Sensitivity)	80.5%	**92.6%**
Specificity	86.2%	**91.0%**
F1-score	81.3%	**91.4%**
ROC-AUC	0.872	**0.951**

Interpretation: Table 6 compares the predictive performance of the Logistic Regression and Random Forest models using various evaluation metrics. The results indicate that the Random Forest model consistently outperformed Logistic Regression across all performance measures. Random Forest achieved a higher accuracy (91.5%) compared to Logistic Regression (84.7%), demonstrating its superior ability to correctly classify high-risk and low-risk patients. Similarly, the precision of the Random Forest model (90.2%) was higher than that of Logistic Regression (82.1%), indicating greater reliability in identifying true high-risk patients. The recall (sensitivity) was also substantially higher for Random Forest (92.6%) compared to Logistic Regression (80.5%), reflecting its enhanced capability to detect patients who are genuinely at high risk. In terms of specificity, Random Forest achieved 91.0%, exceeding the 86.2% obtained by Logistic Regression, thereby reducing the likelihood of false-positive classifications. Furthermore, the F1-score, which balances precision and recall, was higher for Random Forest (91.4%) than for Logistic Regression (81.3%), indicating better overall classification performance. The ROC-AUC value also improved from 0.872 for Logistic Regression to 0.951 for Random Forest, demonstrating excellent discriminative ability.

Hypothesis 1

H_{01}: The developed AI model does not significantly improve prediction accuracy compared to conventional statistical models.

Table 7. Comparison of Predictive Performance between Logistic Regression and Random Forest

Performance Metric	Logistic Regression	Random Forest	Improvement
Accuracy (%)	84.7	**91.5**	+6.8
Precision (%)	82.1	**90.2**	+8.1
Recall (Sensitivity) (%)	80.5	**92.6**	+12.1
Specificity (%)	86.2	**91.0**	+4.8
F1-score	0.813	**0.914**	+0.101
ROC-AUC	0.872	**0.951**	+0.079
10-Fold CV Accuracy (%)	84.69 ± 0.30	**91.54 ± 0.24**	+6.85

Result

Table 7 compares the predictive performance of the conventional Logistic Regression model with the proposed Random Forest model for identifying high-risk patients. The results indicate that the Random Forest model consistently outperformed Logistic Regression across all evaluation metrics. Specifically, the Random Forest achieved a higher accuracy (91.5% vs. 84.7%), precision (90.2% vs. 82.1%), recall/sensitivity (92.6% vs. 80.5%), and specificity (91.0% vs. 86.2%). Similarly, the F1-score increased from 0.813 to 0.914, indicating a better balance between precision and recall. The ROC-AUC value also improved from 0.872 for Logistic Regression to 0.951 for Random Forest, demonstrating superior discriminatory ability in distinguishing high-risk from low-risk patients. Furthermore, the 10-fold cross-validation accuracy increased from 84.69 ± 0.30% to 91.54 ± 0.24%, confirming the robustness and stability of the proposed AI model. Overall, the Random Forest model improved prediction accuracy by 6.8%, with notable improvements in precision (8.1%) and sensitivity (12.1%) compared to the conventional statistical model.

Decision: Since the proposed AI model demonstrated consistently superior predictive performance across all evaluation metrics, the **null hypothesis (H_{01}) is rejected**, and the **alternative hypothesis (H_{11}) is accepted**. Therefore, it can be concluded that the developed AI model significantly improves prediction accuracy compared to the conventional Logistic Regression model and is more effective for the early detection of high-risk patients in community nursing.

Hypothesis 2

H_{02}: There is no significant relationship between selected patient characteristics and predicted clinical risk.

Table 8. Logistic Regression Analysis of Factors Associated with High-Risk Patient Status

Variable	β	Odds Ratio	p-value	Remarks
Age	0.051	1.05	<0.001	Sig.
BMI	0.072	1.07	<0.001	Sig.
Diabetes	1.34	3.82	<0.001	Sig.
Hypertension	1.81	6.11	<0.001	Sig.
Smoking	0.41	1.51	0.002	Sig.
Previous Hospitalization	0.89	2.43	<0.001	Sig.

Table 8 presents the results of the binary Logistic Regression analysis examining the association between selected patient characteristics and predicted clinical risk. The findings indicate that all the selected predictor variables were statistically significant ($p < 0.05$), suggesting that they have a significant influence on the likelihood of being classified as a high-risk patient. Age was positively associated with clinical risk ($\beta = 0.051$, OR = 1.05, 95% CI: 1.04–1.07, $p < 0.001$), indicating that the probability of high-risk status increases with advancing age. Similarly, Body Mass Index (BMI) showed a significant positive association ($\beta = 0.072$, OR = 1.07, 95% CI: 1.04–1.10, $p < 0.001$), demonstrating that individuals with higher BMI were more likely to be categorized as high risk.

Among the clinical variables, hypertension emerged as the strongest predictor of high-risk patient status ($\beta = 1.81$, OR = 6.11, 95% CI: 4.66–8.03, $p < 0.001$), indicating that hypertensive individuals were more than six times as likely to be classified as high risk compared with those without hypertension. Diabetes was also a significant predictor ($\beta = 1.34$, OR = 3.82, 95% CI: 2.81–5.17, $p < 0.001$), increasing the odds of high-risk classification by nearly four times. In addition, previous hospitalization ($\beta = 0.89$, OR = 2.43, 95% CI: 1.82–3.26, $p < 0.001$) and smoking ($\beta = 0.41$, OR = 1.51, 95% CI: 1.18–1.93, $p = 0.002$) were found to be significant predictors of clinical risk, indicating that both prior healthcare utilization and unhealthy lifestyle behaviors contribute substantially to the prediction of high-risk patients.

The Logistic Regression model identified hypertension, diabetes, previous hospitalization, age, BMI, and smoking as significant determinants of predicted clinical risk. These findings demonstrate that demographic, clinical, and behavioral characteristics play an important role in the early identification of high-risk individuals and support their inclusion in the proposed AI-based Clinical Decision Support Model for community nursing.

Decision: Since all the selected patient characteristics showed statistically significant associations with predicted clinical risk ($p < 0.05$), the null hypothesis (H_{02}) is rejected, and the alternative hypothesis (H_{12}) is accepted. Therefore, it is concluded that there is a significant relationship between selected patient characteristics and predicted clinical risk, confirming the usefulness of these variables in developing an effective AI-based Clinical Decision Support Model for the early detection of high-risk patients.

X. CONCLUSION

This study developed an AI-Based Statistical Clinical Decision Support Model (CDSS) for the early detection of high-risk patients in community nursing by integrating Logistic Regression and the Random Forest algorithm. The proposed framework combined conventional statistical analysis with machine learning techniques to enhance the accuracy of patient risk prediction while maintaining clinical interpretability. The findings demonstrated that several patient characteristics, including age, BMI, diabetes, hypertension, smoking, and previous hospitalization, were significant predictors of high-risk patient status, emphasizing their importance in community-based risk assessment.

The comparative evaluation revealed that the Random Forest model outperformed the conventional Logistic Regression model across all performance measures. It achieved higher accuracy (91.5%), precision (90.2%), recall (92.6%), specificity (91.0%), F1-score (0.914), and ROC-AUC (0.951), demonstrating superior predictive capability for identifying high-risk patients. Furthermore, the model exhibited consistent performance during 10-fold cross-validation, confirming its robustness and reliability. The Logistic Regression analysis further identified hypertension and diabetes as the strongest predictors of clinical risk, while the Random Forest variable importance analysis ranked hypertension, age, and diabetes as the most influential variables contributing to prediction accuracy.

The proposed AI-based Clinical Decision Support Model provides an effective and reliable approach for supporting community nurses in the early identification of vulnerable individuals. By enabling timely risk stratification, the model can facilitate preventive interventions, optimize healthcare resource allocation, reduce avoidable hospitalizations, and improve patient outcomes. The integration of statistical modeling with machine learning offers a balanced framework that combines predictive accuracy with clinical interpretability, making it suitable for real-world community healthcare settings. Future research should validate the proposed model using multicenter datasets, incorporate additional clinical and laboratory variables, and integrate explainable AI techniques into real-time clinical decision support systems to further enhance transparency, usability, and adoption in nursing practice.

XI. Nursing Implication

A) Nursing Education:

- Integrate AI-based clinical decision support system into the nursing curriculum.

- Conduct continuing education programmes on AI education in community nursing.

B) Nursing Practice:

- Clinical decision-making through AI-assisted risk prediction while maintaining nurses' clinical judgment.

C) Nursing Research:

- AI models should validate across different community settings & populations.

- Explore ethical, legal & patient acceptance issues related to AI in community nursing.

D) Nursing Administration:

- Organize training programmes for nurses on AI-based CDS.

- Ensure data privacy, quality assurance and adequate infrastructure for AI integration.

ACKNOWLEDGMENT

I sincerely express my heartfelt gratitude to my fellow economist, **Dr. Gaurav Jain (gjain793@gmail.com)**, for his dedicated support and valuable contribution to the **modelling and analytical framework** of this study. His expertise and thoughtful inputs were particularly helpful in developing and refining the modelling approach involving **logistic regression and machine learning techniques** for the proposed AI-based Clinical Decision Support Model. His valuable guidance greatly contributed to the analysis and interpretation of the study findings. I am truly thankful for his time, cooperation, and expertise, which played an important role in the successful completion of this work.

REFERENCES

1. Alenazi, A. S., Alqahtani, F. M., & Alshammari, M. A. (2024). Artificial intelligence-powered clinical decision support systems in nursing: A systematic review. *Saudi Journal of Medicine and Public Health, 1*(2), 55–67.

2. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

3. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA, 319*(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

4. Bharati, S., Podder, P., & Mondal, M. R. H. (2023). Explainable AI for healthcare: A systematic review. *Artificial Intelligence Review, 56*, 1–38.

5. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. *Springer*.

6. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.

7. Buess, M., et al. (2025). From large language models to multimodal AI: A scoping review on the potential of generative AI in medicine. *NPJ Digital Medicine*.

8. Chua, W. L., Rusli, K. D. B., & Aitken, L. M. (2024). Comparing early warning score, qSOFA and SIRS for sepsis prediction: A systematic review and meta-analysis. *Journal of Clinical Nursing*, 33(6), 2005–2018.

9. Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

10. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.

11. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). *Wiley*.

12. Hu, X., et al. (2025). Artificial intelligence models for intensive care unit readmission prediction: A systematic review and meta-analysis. *Journal of Critical Care*.

13. Islam, K. R., et al. (2023). Machine learning approaches for early sepsis prediction using electronic health records: A systematic review. *Journal of Clinical Medicine*, 12(17), 5658.

14. Ji, X., Huo, H., & Dong, L. (2025). Artificial intelligence for early diagnosis of sepsis in intensive care units: A systematic review and meta-analysis. *Journal of Intensive Care Medicine.*

15. Kim, M. S., et al. (2022). Machine learning-based clinical decision support system for pressure injury prevention in nursing practice. *Computers, Informatics, Nursing*, 40(11), 693–701.

16. Kim, S. Y., et al. (2024). Explainable artificial intelligence for clinical decision support systems: A systematic review. *Applied Sciences,* 14(15), 6638.

17. Khokhar, P. B., Gravino, C., & Palomba, F. (2024). Advances in artificial intelligence for diabetes prediction: A systematic literature review. *Artificial Intelligence in Medicine*.

18. Korom, R., et al. (2025). AI-based clinical decision support for primary care: A real-world study. NPJ Digital Medicine.

19. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.

20. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

21. Mikkonen, K., et al. (2026). Artificial intelligence technologies supporting nurses' clinical decision-making: A systematic review. Journal of Clinical Nursing, 35(4), 1525–1540.

22. Moazemi, S., et al. (2023). Artificial intelligence in cardiovascular intensive care: A systematic review. Frontiers in Medicine, 10, 1109411.

23. Okere, A. N., et al. (2025). Machine learning prediction of hospitalization and readmission using community health survey data. BMJ Health & Care Informatics, 32(1).

24. Kaser, P. (2025). Chhattisgarh ke paanch vishesh roop se kamzor janjatiya samuhon (PVTGs) mein shiksha ki sthiti. *Janajatiya Bhugol*, 111–119. https://doi.org/10.52711/book.anv.tribalgeography-10

25. Kaser, P. (2025). Effect of vaccination on predictive model of Covid-19 in Chhattisgarh. *NewBioWorld: A Journal of Alumni Association of Biotechnology,* 7(2), 1–3. https://doi.org/10.52228/NBW-JAAB.2025-7-2-1

26. Kaser, P. (2026). Double sampling plan formulation for quality assessment. *International Journal of Science and Research (IJSR), 15*(3), 1622–1625. https://doi.org/10.21275/SR26326185226

27. Kaser, P. (2026). Generalized estimator on operations research guided sampling (ORGS). *International Scientific Journal of Engineering and Management (ISJEM),* 5(4), 1–6. https://doi.org/10.55041/ISJEM05915

28. Kaser, P. (2026). Sampling estimators and properties of a good estimator: A theoretical and applied perspective. *International Journal of Science, Strategic Management and Technology, 2*(3). https://doi.org/10.55041/ijsmt.v2i3.343

29. Kaser, P. (2026). Systematic review of literature: Sampling estimators. *International Journal of Creative and Open Research in Engineering and Management, 2*(3). https://doi.org/10.55041/ijcope.v2i3.181

30. Prentzas, N., et al. (2023). Explainable artificial intelligence applications in the medical domain: A systematic review. Artificial Intelligence Review.

31. Rajkomar, A., Dean, J., & Kohane, I. S. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

32. Sivaraman, V., et al. (2023). Understanding clinician acceptance of AI-based treatment recommendations. Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL).

33. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

34. Wang, L., et al. (2025). Machine learning models for predicting readmission among elderly patients with diabetes and heart failure. Frontiers in Cardiovascular Medicine, 12, 1673159.

35. World Health Organization. (2023). Noncommunicable diseases. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

36. Yuan, S., et al. (2025). Artificial intelligence-powered early warning systems for detecting clinical deterioration: A systematic review and meta-analysis. BMC Medical Informatics and Decision Making, 25, 203.

37. Yu, M. Y., & Son, Y. J. (2024). Machine learning models for predicting hospital readmission in heart failure: A systematic review. European Journal of Cardiovascular Nursing, 23(7), 711–719.

38. Zhang, S., et al. (2023). Rethinking human-AI collaboration in complex medical decision making: A case study on sepsis diagnosis. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).

39. Zhang, Y., et al. (2024). Prediction models for coronary artery disease readmission: A systematic review. Risk Management and Healthcare Policy, 17, 1023–1038.

40. Tądel, K., Dudek, A., & Bil-Lula, I. (2024). Artificial intelligence in sepsis prediction and management: A systematic review. Journal of Clinical Medicine, 13(19), 5959.

41. International Institute for Population Sciences (IIPS) & ICF. (2021). National Family Health Survey (NFHS-5), 2019–21: India Report. Mumbai, India: IIPS. https://www.rchiips.org/nfhs/.