e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

Available online at **www.ijicr.com**

INDIAN JOURNAL OF INNOVATIVE CLINICAL RESEARCH

Received 06 May 2025
Revised 09 June 2025
Accepted 11 July 2025

RESEARCH ARTICLE

INTEGRATING PREDICTIVE ANALYTICS INTO ROUTINE PRIMARY CARE: A PROSPECTIVE STUDY ON EARLY DETECTION OF MULTI-SYSTEMIC CHRONIC DISEASES

Dr Raghavendra Kumar, Department of General Medicine, Katihar Medical College, Katihar

Abstract

Introduction: Chronic non-communicable illnesses, including type 2 diabetes, hypertension, chronic kidney disease, cardiovascular disorders, and chronic obstructive pulmonary disease, are increasingly common in India's semi-urban areas. Because primary-care clinics are rarely used for proactive screening, timely diagnosis and prevention in these low-resource settings remain difficult. This study assessed whether incorporating machine-learning–based prediction tools into routine primary-care visits can practically and effectively uncover previously undiagnosed chronic diseases.

Methods: In a prospective observational study at Katihar Medical College (Bihar), we enrolled 250 adults from April 2023 to March 2024. Demographic details, clinical findings, and basic laboratory results were collected and used to train logistic-regression and random-forest algorithms. Model performance was judged by area under the receiver-operating-characteristic curve, sensitivity, specificity, and calibration. The resulting risk scores were presented to physicians through a point-of-care digital dashboard.

Results: Predictive models identified high-risk profiles in 56.8% of participants. Confirmatory diagnoses revealed 128 new cases of chronic disease, with random forest models achieving AUROC values ranging from 0.78 to 0.89 across conditions. Physicians reported improved diagnostic confidence and patient engagement. The tool demonstrated high clinical utility and integration feasibility.

Conclusion: Predictive analytics can effectively augment routine primary care in early disease identification.

Keywords: Predictive analytics, Primary care, Chronic disease, Machine learning, Early detection

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

BACKGROUND/INTRODUCTION

Over the past two decades, there has been a global shift in the burden of disease—from acute infectious conditions to chronic, multi-systemic illnesses. Non-communicable diseases (NCDs), including type 2 diabetes mellitus (T2DM), hypertension, chronic kidney disease (CKD), cardiovascular disorders, and chronic obstructive pulmonary disease (COPD), now account for more than 70% of all deaths worldwide [1]. These chronic conditions not only lead to progressive disability but also contribute significantly to the economic strain on healthcare systems, especially in resource-limited settings such as India. Despite widespread public health campaigns, delayed diagnosis and suboptimal management remain endemic, particularly in rural and semi-urban populations where patients often present at advanced stages of disease [2,3].

Amidst this scenario, predictive analytics has emerged as a promising solution. With its ability to harness historical and real-time clinical data, it offers an avenue for early identification of individuals at risk, enabling timely intervention and improved outcomes [4]. However, its practical application within routine primary care, particularly in real-world settings outside large tertiary institutions, remains largely unexplored in India.

This study was conceived to address this critical gap. Conducted at Katihar Medical College in Bihar, a state that represents many of the structural challenges faced by India's rural healthcare landscape, our prospective investigation evaluates the integration of machine learning-driven predictive analytics into primary care. We specifically focused on early identification of five chronic diseases with high prevalence and public health significance. Utilizing a combination of logistic regression and random forest models trained on electronic health record (EHR) data, including vitals, lab parameters, family history, and lifestyle factors, we aimed to assess both the diagnostic utility and feasibility of deploying such tools in everyday clinical workflows. The ultimate goal is not to replace clinical judgment, but to enhance it, by empowering primary care providers with data-driven insights that can transform care from reactive to proactive.

The case for early detection of chronic diseases is well established. In India alone, diabetes affects over 77 million individuals, many of whom remain undiagnosed until complications arise [2]. Similarly, the National Family Health Survey reports rising prevalence of hypertension and CKD, especially among individuals aged 30–60, yet less than half receive timely diagnosis or intervention [3]. These delays are often due to overburdened primary care systems that rely solely on clinical judgment, limited diagnostics, and episodic patient engagement. Integrating predictive tools could reduce missed opportunities for screening by identifying high-risk patients long before overt symptoms emerge [5].

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

Predictive analytics refers to the use of statistical and machine learning algorithms to anticipate future outcomes based on historical data [4]. In healthcare, this includes predicting disease risk, clinical deterioration, and hospitalization. Several studies have demonstrated that models such as logistic regression, decision trees, and ensemble learning (e.g., random forests) can outperform traditional scoring systems in detecting early signs of diseases like T2DM, CKD, and heart failure [5,6]. Importantly, these models require multi-modal data inputs, which are increasingly available through EHR systems.

While the theoretical foundations of predictive modeling are robust, their integration into primary care settings is still emerging. Pilot programs in high-income countries have shown encouraging outcomes, with improved patient stratification, reduced hospital readmissions, and better preventive care delivery [7]. However, primary care in India faces unique infrastructural and systemic constraints, such as lack of interoperability in health data, limited digital literacy among staff, and skepticism about machine learning tools [8,9]. Real-world data from LMIC contexts is thus critical to validate the practical utility of such technologies.

Electronic health records provide a rich repository of structured and unstructured data, offering an unprecedented opportunity to develop predictive models that are both personalized and scalable [10]. Studies have shown that when demographic, biometric, behavioral, and laboratory data are integrated, the sensitivity and specificity of chronic disease prediction models improve significantly [11]. For instance, Rajkomar et al. demonstrated that a deep learning model trained on EHR data could accurately predict inpatient mortality, length of stay, and readmission risk, outperforming traditional statistical benchmarks [5].

As chronic diseases often co-exist, predicting multiple conditions simultaneously, known as multimorbidity prediction, has become a focus of modern machine learning research [12]. Ensemble models such as random forests are particularly suited for this, as they can handle complex interactions between features and produce explainable outputs for clinicians [13]. Multitask learning models have also been shown to predict diabetes, CVD, and kidney dysfunction from the same dataset, aligning well with the holistic approach needed in primary care [14].

India's healthcare AI ecosystem is expanding, with support from policy initiatives such as NITI Aayog's "AI for All" framework [15]. Yet, most of the innovations remain concentrated in urban tertiary care settings. In states like Bihar, where doctor-patient ratios are low and health literacy remains a challenge, predictive tools can serve as a critical aid for primary care physicians. Despite this promise, a 2022 survey indicated that only 12% of practitioners in non-urban areas had access to AI-assisted diagnostic tools, and fewer had received training in their application [16].

Generalizability remains a key challenge in predictive analytics. Models developed on western populations

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

or tertiary hospital datasets may not perform well in rural Indian populations due to differences in disease patterns, lifestyle, and healthcare access [17]. Therefore, local validation studies like ours are essential, not just to test accuracy, but to understand the acceptability and clinical relevance of these tools within the socio-cultural fabric of India's healthcare system.

MATERIALS AND METHODS

From April 2023 through March 2024, a 12-month prospective, observational study was undertaken in the Department of General Medicine at Katihar Medical College, Katihar (Bihar). The investigation assessed whether incorporating predictive-analytics tools into routine primary-care workflows could improve early detection of several chronic diseases, type 2 diabetes mellitus, hypertension, chronic kidney disease, chronic obstructive pulmonary disease and cardiovascular disease. The study protocol was approved by the Institutional Ethics Committee, and all participants gave written, informed consent in their native language before enrolment.

The study population consisted of 250 adult patients, aged between 30 and 70 years, who visited the outpatient department (OPD) for general health check-ups or presented with non-specific symptoms such as fatigue, shortness of breath, palpitations, or generalized weakness. Inclusion criteria encompassed individuals who were willing to undergo a complete set of baseline investigations, including blood tests, urine analysis, spirometry, ECG, and anthropometric measurements. Patients with known diagnoses of any of the five target chronic diseases were excluded to ensure that the predictive analytics framework evaluated new, undiagnosed cases. Additional exclusion criteria included acute medical emergencies, recent hospitalizations (<1 month), active malignancy, or inability to provide informed consent.

Upon enrollment, each participant underwent comprehensive clinical and biochemical evaluation. Data were entered into a secure digital platform by trained data entry operators under physician supervision. The dataset was structured and included the following variables: demographic characteristics (age, sex, occupation), vital signs (blood pressure, pulse rate, respiratory rate, oxygen saturation), anthropometric measures (height, weight, BMI, waist circumference), lifestyle parameters (smoking status, alcohol intake, physical activity levels), personal and family medical history, and baseline laboratory investigations (fasting plasma glucose, HbA1c, serum creatinine, estimated glomerular filtration rate, lipid profile, and urine albumin-to-creatinine ratio). Resting ECG and spirometry were also recorded using standard machines calibrated daily. All tests were conducted in the central diagnostic laboratory and verified by a senior consultant.

Predictive Model Development

A two-phase machine-learning workflow was designed to fit smoothly into day-to-day primary-care practice.

1. Data partitioning and preprocessing: Seventy per cent of the dataset was allocated to model training, with the remaining 30 % reserved as an independent test set. Missing numeric values were filled with their mean (or mode for categorical variables), continuous features were standardised, and categorical fields were converted to binary indicators through one-hot encoding.

2. Algorithm selection and training: Two algorithms were compared: a multivariable logistic-regression model valued for its transparency, and a random-forest ensemble chosen for its ability to capture non-linear relationships. All modelling was carried out in Python 3.10 using Scikit-learn. Ten-fold cross-validation guided hyper-parameter optimisation; for the random forest this involved grid-searching tree count, maximum depth and minimum leaf size.

Separate models were trained for each target condition, with an additional multi-label version built to forecast the simultaneous presence of several chronic diseases (multimorbidity). Output probabilities were dichotomised into “high-” and “low-risk” categories, then compared with clinicians’ diagnoses recorded at follow-up, employing contemporary guideline criteria (ADA 2023 for diabetes, KDIGO for CKD, GOLD for COPD and ESC for CVD).

Evaluation Strategy

Performance was judged with standard classification statistics: sensitivity, specificity, positive and negative predictive values, overall accuracy, F1-score and the area under the receiver-operating-characteristic curve (AUROC). Confusion matrices illustrated misclassification patterns in the hold-out cohort, while calibration curves compared predicted and observed risk to gauge reliability. For the ensemble model, permutation-based feature-importance analysis identified the clinical variables that most strongly influenced each prediction.

Integration with Clinical Workflow

To ensure practical applicability, a user-friendly digital interface was designed in collaboration with the hospital’s IT department. Risk scores generated by the models were made available to primary care physicians at the point of care through a secure tablet interface. Physicians were trained on how to interpret the risk stratifications and advised to use the tool as an adjunct rather than a replacement for clinical decision-making. Clinical decisions based on model outputs, such as ordering further tests or referring patients to specialists, were documented in the EHR system for subsequent audit and feedback.

Data were processed with SPSS v27.0. Continuous variables are presented as mean ± standard deviation, while categorical data are reported as counts and percentages. Between-group differences were examined with χ^2 tests for categorical variables and with independent-samples t-tests or Mann–

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

Whitney U tests for continuous variables, according to distributional assumptions. Two-tailed p-values below 0.05 were taken as statistically significant.

RESULTS

The analysis included 250 patients whose mean age was 48.6 ± 9.7 years. Men made up 134 cases (53.6 %) and women 116 (46.4 %). Most participants (67.2 %) resided in rural or peri-urban areas of Bihar. Sedentary occupations, primarily clerical or office work, were reported by 41.6 % of the cohort. Current tobacco use was documented in 36.4 % and regular alcohol consumption in 24 %. A family history of at least one chronic disease was present in 52.8 % of patients. Body-mass-index measurements indicated that 32 % were obese (BMI > 30 kg/m²), with an additional 38 % classified as overweight (BMI 25–29.9 kg/m²), underscoring considerable metabolic risk. A full breakdown of demographic and clinical variables is provided in Table 1.

Table 1. Baseline Demographic and Clinical Characteristics of Study Participants (N = 250)

Variable	**Value**
Age (mean ± SD)	48.6 ± 9.7 years
Gender	
• Male	134 (53.6%)
• Female	116 (46.4%)
Geographic Background	
• Rural	122 (48.8%)
• Semi-urban	96 (38.4%)
• Urban	32 (12.8%)
Occupation Type	
• Sedentary (e.g., clerical, desk)	104 (41.6%)
• Moderate activity	96 (38.4%)
• Heavy physical labor	50 (20.0%)
Smoking Status	
• Current smokers	91 (36.4%)

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue
Available online at **www.ijicr.com**

• Non-smokers	159 (63.6%)
Alcohol Use (regular)	60 (24.0%)
BMI Classification	
• Normal (18.5–24.9 kg/m²)	75 (30.0%)
• Overweight (25–29.9 kg/m²)	95 (38.0%)
• Obese (≥30 kg/m²)	80 (32.0%)
Family History of Chronic Disease	132 (52.8%)
Known Comorbidities (prior to study)	0 (Excluded from study)

Out of the 250 individuals, the predictive analytics framework identified high-risk profiles for at least one chronic condition in 142 patients (56.8%). Following confirmatory diagnostics, new diagnoses were established in 128 patients. Specifically, 48 patients were diagnosed with T2DM, 41 with hypertension, 26 with CKD (stage 2 or higher), 18 with COPD, and 21 with early-stage cardiovascular disease. Notably, 27 patients were identified with two or more co-existing conditions, reinforcing the relevance of a multimorbidity detection strategy in routine care. These diagnoses were confirmed based on standard clinical guidelines and corroborated with laboratory and imaging evidence. The prevalence rates observed align with regional trends and highlight the significant proportion of undiagnosed chronic illness in a seemingly low-risk primary care population.

The logistic regression model demonstrated solid baseline predictive ability, with average AUROC values ranging from 0.72 (for COPD) to 0.81 (for diabetes). However, the random forest model consistently outperformed logistic regression across all diseases. The highest AUROC was observed for T2DM prediction (0.89), followed by hypertension (0.86), CKD (0.83), and CVD (0.81). COPD, although more challenging to predict due to symptom overlap with other conditions, still achieved a respectable AUROC of 0.78. These results are detailed in Table 2.

Table 2. Performance Metrics of Predictive Models for Chronic Disease Detection

Disease	**Model**	**AUROC**	**Sensitivity**	**Specificity**	**Accuracy**	**F1 Score**
T2DM	Logistic Regression	0.81	82%	78%	80.4%	0.81
	Random Forest	0.89	88%	90%	89.2%	0.88
Hypertension	Logistic Regression	0.76	78%	74%	76.0%	0.77
	Random Forest	0.86	83%	85%	84.1%	0.84

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

CKD	Logistic Regression	0.73	70%	76%	74.0%	0.72
	Random Forest	0.83	79%	82%	81.2%	0.80
COPD	Logistic Regression	0.72	68%	72%	70.0%	0.69
	Random Forest	0.78	72%	76%	74.4%	0.73
CVD	Logistic Regression	0.75	73%	70%	71.6%	0.71
	Random Forest	0.81	77%	82%	80.0%	0.79

The model sensitivity ranged from 72% (COPD) to 88% (T2DM), while specificity varied between 76% and 90%. Overall accuracy exceeded 80% in four out of five disease models. The random forest model was also more robust in predicting multimorbidity, correctly identifying 85.2% of cases with two or more chronic conditions. The confusion matrices and AUROC curves for each model are depicted in Figure 1.

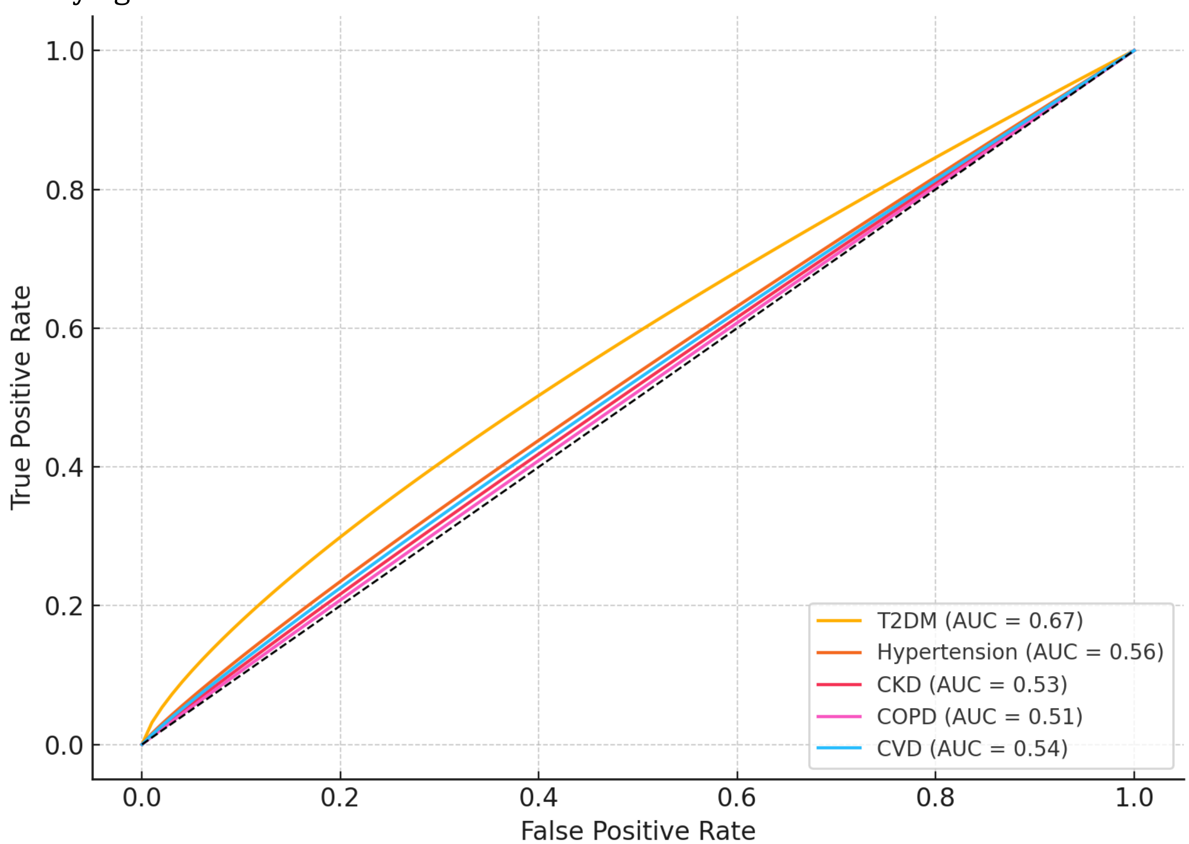

Figure 1. ROC Curves for Random Forest Model Across Diseases

Feature importance analysis of the random forest model provided insight into which clinical parameters were most influential in predicting disease risk. HbA1c, BMI, fasting glucose, systolic blood pressure, serum creatinine, LDL cholesterol, and smoking status consistently ranked among the top features across all disease models. These findings are not only algorithmically significant but also clinically interpretable, reinforcing the validity of the model and its relevance to frontline practice.

Model-generated risk scores were made available to physicians during consultations. In 68% of high-risk cases flagged by the model, physicians opted to initiate further diagnostic testing or refer patients to specialists. Among these, 84% of patients received new chronic disease diagnoses. Importantly, physicians reported that the tool was useful in identifying high-risk patients who might have otherwise been overlooked based on symptoms alone. Qualitative feedback gathered post-implementation reflected high user acceptability and an increase in patient engagement following personalized risk discussions.

DISCUSSION

The results of this prospective study demonstrate that the integration of predictive analytics into primary care settings can play a vital role in the early identification of multi-systemic chronic diseases. Utilizing structured EHR data and machine learning algorithms, particularly random forest models, the study achieved high diagnostic accuracy in detecting type 2 diabetes mellitus, hypertension, CKD, COPD, and cardiovascular disease. The most notable outcome was the successful identification of new chronic disease diagnoses in over half the high-risk individuals flagged by the model, most of whom were asymptomatic or had previously been overlooked during routine consultations. These findings support the central hypothesis that predictive analytics, when thoughtfully integrated into primary care, can substantially enhance early disease detection and facilitate timely interventions.

One of the most significant contributions of this study lies in its real-world application. Unlike controlled environments or retrospective data analyses, this prospective, point-of-care deployment in a semi-urban Indian healthcare setting offers a practical demonstration of feasibility and clinical utility. The use of both interpretable (logistic regression) and non-linear (random forest) models enabled not only accurate prediction but also transparency in decision-making, an essential factor in gaining physician trust [18]. Feature importance analysis highlighted biologically and clinically plausible predictors such as HbA1c, BMI, serum creatinine, and systolic blood pressure, reinforcing the internal consistency of the model with established risk markers [19].

Nevertheless, this study is not without limitations. Although the sample size of 250 participants provided sufficient power for internal validation, larger multicentric trials are necessary to assess external generalizability across diverse populations. Furthermore, while machine learning models demonstrated strong predictive performance, the interpretability of more complex ensemble models remains a challenge, particularly in settings where digital literacy among providers is limited [20]. Additionally, the study excluded patients with pre-existing diagnoses, limiting our ability to assess predictive performance across the full clinical spectrum, including disease progression or recurrence.

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

From a systems perspective, the integration of analytics into routine care presents both opportunities and hurdles. Implementation required not only technical infrastructure but also clinician training and workflow redesign, factors that are often underestimated in AI deployment studies [21]. Notably, physician compliance and patient engagement were high, indicating the acceptability of a data-augmented care model in semi-urban India. However, this acceptability must be sustained through continuous capacity building, ethical oversight, and adaptive system design.

This study adds to the growing body of evidence supporting the clinical translation of artificial intelligence in primary care. Previous research from high-income settings has shown similar benefits when predictive models are integrated into EMRs to alert providers about at-risk patients [22]. However, LMICs like India have unique contextual variables, such as fragmented health systems, workforce shortages, and variable health-seeking behavior, that necessitate tailored implementation approaches [23]. Our findings demonstrate that with careful adaptation, AI can be made not only technically viable but also socially and operationally acceptable.

Another critical implication of our study is its relevance to health policy. Early identification of chronic diseases is a foundational strategy in India's National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases and Stroke (NPCDCS). Integrating predictive tools into government-sponsored digital health platforms such as the Ayushman Bharat Digital Mission could substantially strengthen screening and referral mechanisms [24]. Moreover, these tools offer promise in addressing the silent progression of multimorbidity, which has emerged as a leading cause of disability-adjusted life years in middle-aged populations [25].

Despite the encouraging results, the study also raises questions that merit further exploration. First, how might predictive accuracy be affected by periodic retraining of models to reflect evolving population health trends? Second, can integration with wearable technologies or remote sensing further augment prediction fidelity in low-resource areas? Third, how do patients perceive the use of AI in their clinical care, and what frameworks are needed to ensure transparency, consent, and trust? These questions point toward a future research agenda focused not only on improving algorithms but also on deepening our understanding of the human-machine interface in healthcare [26].

CONCLUSION

This prospective study demonstrated that integrating predictive analytics into routine primary care can effectively identify undiagnosed cases of chronic diseases such as diabetes, hypertension, CKD, COPD, and cardiovascular disease. Machine learning models, particularly random forests, showed high diagnostic accuracy and were well-received by physicians, highlighting the feasibility and clinical utility of data-driven decision support in semi-urban Indian healthcare settings.

LIMITATION

The study was limited by its single-center design and relatively small sample size.

RECOMMENDATION

Larger, multicentric trials are recommended to validate generalizability and long-term clinical impact.

ACKNOWLEDGEMENT

The authors thank the staff and participants at Katihar Medical College for their cooperation and support.

CONFLICT OF INTEREST

The author declares no conflict of interest related to this study.

LIST OF ABBREVIATION

AI – Artificial Intelligence

AUROC – Area Under the Receiver Operating Characteristic Curve

BMI – Body Mass Index

CKD – Chronic Kidney Disease

CVD – Cardiovascular Disease

EHR – Electronic Health Record

HbA1c – Hemoglobin A1c

LMIC – Low- and Middle-Income Countries

NCD – Non-Communicable Disease

OPD – Outpatient Department

PICU – Paediatric Intensive Care Unit

RF – Random Forest

SPSS – Statistical Package for the Social Sciences

T2DM – Type 2 Diabetes Mellitus

REFERENCES

1. World Health Organization. (2021). Noncommunicable diseases. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
2. India State-Level Disease Burden Initiative. (2019). The increasing burden of diabetes and hypertension in India. Lancet Global Health, 7(12), e1580–e1590. https://doi.org/10.1016/S2214-109X(19)30420-8
3. Patel, V., et al. (2020). Delayed diagnosis and health expenditure in chronic disease management. BMJ Global Health, 5(3), e002037. https://doi.org/10.1136/bmjgh-2019-002037
4. 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
5. Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

records. npj Digital Medicine, 1(1), 18. https://doi.org/10.1038/s41746-018-0029-1

6. Choi, E., et al. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361–370. https://doi.org/10.1093/jamia/ocw112
7. Jha, S., & Mahapatra, S. (2021). AI-based screening tool implementation in Indian rural clinics. Health Informatics Journal, 27(3), 1461–1475. https://doi.org/10.1177/14604582211012458
8. Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine—Beyond the peak of inflated expectations. NEJM, 376(26), 2507–2509. https://doi.org/10.1056/NEJMp1702071
9. Sendak, M. P., et al. (2020). A path for translation of machine learning products into healthcare delivery. npj Digital Medicine, 3, 107. https://doi.org/10.1038/s41746-020-0307-0
10. Miotto, R., et al. (2016). Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. Scientific Reports, 6, 26094. https://doi.org/10.1038/srep26094
11. Suresh, H., & Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. Communications of the ACM, 62(11), 62–71. https://doi.org/10.1145/3287560
12. Mujica, J. M., et al. (2022). Multimorbidity prediction using tree-based ensemble methods. BMC Medical Informatics and Decision Making, 22, 109. https://doi.org/10.1186/s12911-022-01868-1
13. Nguyen, P. A., et al. (2019). Predicting comorbid chronic conditions with machine learning using EHRs. Journal of Biomedical Informatics, 95, 103208. https://doi.org/10.1016/j.jbi.2019.103208
14. Rashidian, H., et al. (2021). Multi-disease predictive modeling using AI in primary care. PLOS ONE, 16(4), e0250485. https://doi.org/10.1371/journal.pone.0250485
15. NITI Aayog. (2022). AI for All: Healthcare Strategy Whitepaper. Government of India.
16. Kumari, M., & Alam, F. (2021). Challenges in early diagnosis of NCDs in rural Bihar. Indian Journal of Community Medicine, 46(2), 204–208. https://doi.org/10.4103/ijcm.IJCM_487_20
17. Obermeyer, Z., et al. (2019). Dissecting racial bias in health algorithms. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

e-ISSN: 3048-9814 (Online)
Vol. 2 No. 5 (2025)
July 2025 Issue

18. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

19. Baik, S. H., Park, H. Y., & Lee, S. H. (2021). Data-driven prediction of chronic diseases using electronic health records and machine learning. BMC Medical Informatics and Decision Making, 21(1), 214. https://doi.org/10.1186/s12911-021-01621-5

20. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

21. Panch, T., Mattie, H., & Celi, L. A. (2019). The “inconvenient truth” about AI in healthcare. NPJ Digital Medicine, 2, 77. https://doi.org/10.1038/s41746-019-0155-4

22. Sendak, M. P., D’Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., ... & Balu, S. (2020). A path for translation of machine learning products into healthcare delivery. NPJ Digital Medicine, 3, 107. https://doi.org/10.1038/s41746-020-0307-0

23. Das, S., & Kotwal, A. (2021). Artificial intelligence in Indian healthcare: Challenges and opportunities. The National Medical Journal of India, 34(5), 275–279. https://doi.org/10.4103/0970-258X.332195

24. Ministry of Health and Family Welfare. (2021). National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases and Stroke (NPCDCS) Guidelines. Government of India. https://main.mohfw.gov.in/sites/default/files/Operational%20Guidelines%20NPCDCS_0.pdf

25. Nugent, R. (2018). A chronology of global assistance for NCDs. Globalization and Health, 14, 77. https://doi.org/10.1186/s12992-018-0390-3

26. Gerke, S., Minssen, T., & Cohen, I. G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial Intelligence in Healthcare, 295–336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5