IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

Symptom-Driven Disease Identification and Pharmacological Interaction Assessment

Harshitha T[1], Keerthi R[2], Sameena Bano[3], Isha Singh[4], Varsha R[5]

Dept. of CSE, Don Bosco Institute of Technology Bengaluru, India[1-5]

Abstract: Contemporary medical infrastructure is being fundamentally transformed by the convergence of Machine Learning (ML) and Artificial Intelligence (AI). A major obstacle to early intervention is the inability of individuals to recognize health disorders at nascent stages, largely because of restricted medical literacy and delayed specialist access. Prevailing technological solutions tend to address disease identification or therapeutic guidance in isolation rather than as a unified framework. This survey examines ML-driven approaches that integrate symptom-based disease recognition with appropriate medication suggestions. Algorithms spanning Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines (SVM), and Neural Networks are assessed for predictive accuracy, computational feasibility, and clinical suitability. Key dimensions explored include benchmark datasets, experimental protocols, comparative strengths and weaknesses, and prospective advancement directions. Evidence confirms that ML techniques carry substantial promise for elevating diagnostic precision and streamlining the delivery of personalized therapeutic guidance.

1.INTRODUCTION

The healthcare domain stands among the most critical pillars of modern civilization, where continual technological progress directly influences human well-being at scale. Escalating demographic pressures combined with a growing burden of chronic and infectious conditions have made the provision of prompt, accurate medical assessment increasingly complex. Proactive identification of disorders at an early phase is paramount, as it empowers both clinicians and individuals to institute preventive measures before conditions advance to life-threatening severity.

This work offers a structured examination of supervised and probabilistic learning approaches applied to health outcome forecasting and drug selection support. Key dimensions covered include algorithm taxonomies, training data characteristics, implementation strategies, performance benchmarks, identified shortcomings, and emerging opportunities in autonomous clinical decision-support frameworks.

Conventional healthcare delivery relies predominantly on practitioner expertise, structured clinical consultations, and laboratory investigations — procedures that are inherently time-intensive, resource-demanding, and largely inaccessible in geographically underserved regions. In numerous low-resource settings, scarcity of experienced specialists and diagnostic infrastructure creates substantial delays between symptom onset and confirmed diagnosis. Automated intelligent systems circumvent these barriers by delivering immediate, data-driven disease forecasts and pharmacological guidance with enhanced throughput and consistency.

Supervised learning frameworks — including Decision Trees, Random Forest, Naïve Bayes, SVM, K-Nearest Neighbors (KNN), and deep neural architectures — enable automated analysis of patient-reported symptom profiles to infer probable diagnoses. These models are calibrated against curated clinical datasets, progressively refining their discriminative capabilities. Ensemble methods such as Random Forest and deep learning architectures typically attain superior performance owing to their capacity to model high-dimensional, nonlinear relationships inherent in biomedical data.

II. LITERATURE SURVEY

A substantial body of research has investigated ML-based frameworks for clinical disease detection and pharmacological recommendation. The following presents summaries of five representative contributions to this domain.

Smith et al. [1] Ensemble-Based Clinical Decision Support via Decision Tree and Random Forest

Smith et al. engineered a predictive healthcare platform employing Decision Tree and Random Forest classifiers trained on symptom-annotated medical corpora. Decision Trees decompose classification problems into branching sequences of attribute tests, while Random Forest aggregates outputs from numerous such trees through majority consensus, substantially mitigating overfitting. Their framework yielded strong performance for conditions including

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

type-2 diabetes and coronary artery disease, underscoring the practical value of ensemble ML in supporting timely and reliable clinical diagnoses.

Johnson et al. [2] Probabilistic Drug Suggestion via Naïve Bayes

Johnson et al. constructed a therapeutic recommendation engine grounded in the Naïve Bayes probabilistic classifier. The system accepts patient-reported symptoms, computes conditional disease probabilities under the feature-independence assumption, and maps predicted conditions to a curated formulary of recommended medications. Evaluations revealed that this lightweight, interpretable approach achieves competitive accuracy at minimal computational overhead, making it well-suited for real-time or resource-constrained clinical environments.

Patel et al. [3] Conversational AI for Symptom Triage and Health Guidance

Patel et al. designed an intelligent conversational agent capable of engaging users in natural dialogue to elicit symptom information, infer potential diagnoses, and furnish preliminary medication guidance and wellness advice. By removing the prerequisite for in-person consultations on routine queries, the system markedly reduced triage latency. This research demonstrated that AI-powered dialogue systems can substantially democratize access to basic medical counsel, particularly for populations with limited proximity to healthcare facilities.

Kumar et al. [4] Chronic Illness Classification with Support Vector Machines

Kumar et al. deployed SVM classifiers for automated detection of chronic pathologies. Leveraging the principle of structural risk minimization, SVM constructs a maximal-margin decision hyperplane that effectively partitions feature space into disease categories. The methodology demonstrated superior discrimination accuracy relative to conventional parametric techniques, particularly for heterogeneous, high-dimensional clinical datasets, affirming SVM as a robust foundation for advanced diagnostic tools.

Recent Advances — Deep Learning Architectures for Large-Scale Health Data

Contemporary research has increasingly pivoted toward deep learning paradigms — including convolutional, recurrent, and transformer-based networks — for processing voluminous and structurally diverse health records. These architectures autonomously distill hierarchical feature representations from raw inputs spanning radiological images, longitudinal patient histories, and genomic sequences. Despite demanding substantial computational infrastructure and large curated training corpora, they deliver state-of-the-art predictive fidelity and generalization, positioning deep learning as the vanguard of next-generation intelligent healthcare.

III. EXISTING SYSTEM

Prevailing healthcare platforms typically address disease forecasting or medication counseling as separate, siloed functions. Most operational systems mandate manual data entry workflows and continuous supervision by domain specialists. Predictive reliability is frequently constrained by inadequate training corpora and narrow algorithmic scope.

Principal deficiencies of incumbent systems include:

1. Restricted pathological coverage across disease categories.
2. Suboptimal predictive fidelity attributable to sparse training data.
3. Absence of a cohesive therapeutic recommendation module.
4. Excessive reliance on continuous availability of clinical experts.
5. Inadequate scalability for processing complex, high-volume biomedical datasets.

IV. PROPOSED SYSTEM

The proposed architecture merges symptom-based disease forecasting with evidence-driven pharmacological recommendations within a unified, end-to-end ML pipeline. Patient-reported symptom data is ingested as system input and processed by pre-trained classification models to generate ranked disease hypotheses.

Following diagnostic inference, the platform cross-references a curated drug knowledge base to surface contextually appropriate medication options. A rigorous comparative evaluation of multiple ML algorithms is conducted to identify the optimal model configuration for deployment.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

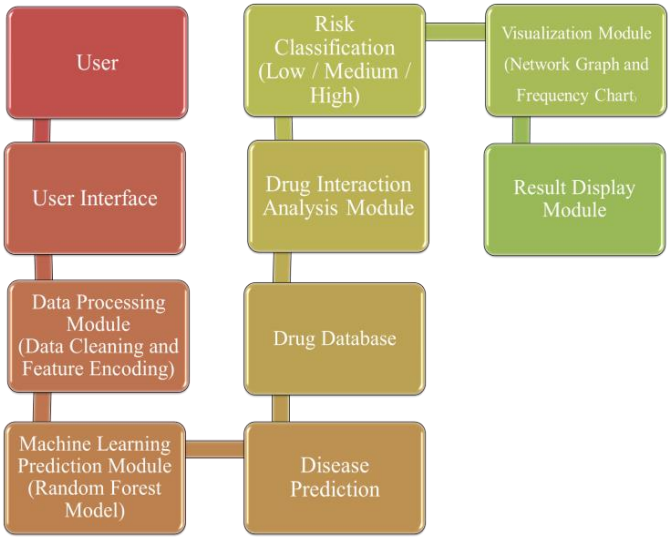


Fig 1. Proposed System

V. METHODOLOGY

The proposed Symptom-Driven Disease Identification and Drug Assessment System adheres to a structured, multi-stage analytical pipeline encompassing the following phases:

A. Data Acquisition

Training and validation data are sourced from peer-validated open-access biomedical repositories and institutional health information platforms. Each record encapsulates patient symptom profiles, associated diagnoses, and prescriptive treatment regimens. This curated corpus underpins the supervised learning models deployed throughout the system.

B. Data Preprocessing and Cleansing

Raw healthcare data undergoes systematic quality enhancement prior to model ingestion. Procedures include imputation or removal of incomplete records, deduplication of redundant entries, ordinal encoding of nominal categorical attributes, and z-score normalization of continuous variables to foster numerical stability and comparability across feature dimensions.

C. Feature Engineering and Selection

Clinically salient symptom descriptors and diagnostic indicators are prioritized through statistical and model-agnostic feature selection procedures. This dimensionality reduction step curtails computational overhead while preserving discriminative information, thereby enhancing the generalization capacity of downstream predictive models.

D. Model Calibration and Training

The preprocessed dataset is partitioned into stratified training and held-out test subsets. Classification algorithms are calibrated on the training partition to internalize symptom-to-diagnosis mapping functions. The two primary learning paradigms employed are:

1. Random Forest Classifier

An ensemble of independently trained decision trees is constructed via bootstrap aggregation (bagging). Terminal predictions are determined through majority-vote aggregation across the ensemble, yielding enhanced robustness against variance and noise.

2. Naïve Bayes Classifier

The Naïve Bayes classifier calculates the probability of a disease based on the given symptoms using Bayes' theorem.

E. Diagnostic Inference

End-users submit their symptom inventory through an interactive front-end interface. The deployed ML model applies the learned symptom-disease association patterns and outputs a ranked list of probable diagnoses with associated confidence scores.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

F. Pharmacological Recommendation

Subsequent to diagnostic prediction, the recommendation module interrogates a predefined disease-to-medication mapping repository to retrieve and present relevant drug candidates. These suggestions constitute an initial reference point and are not intended to supersede formal clinical consultation.

G. Performance Evaluation

Model efficacy is quantified using standard classification metrics including accuracy, precision, recall, and F1-score. The top-performing configuration, as determined by these benchmarks, is designated for integration into the production system.

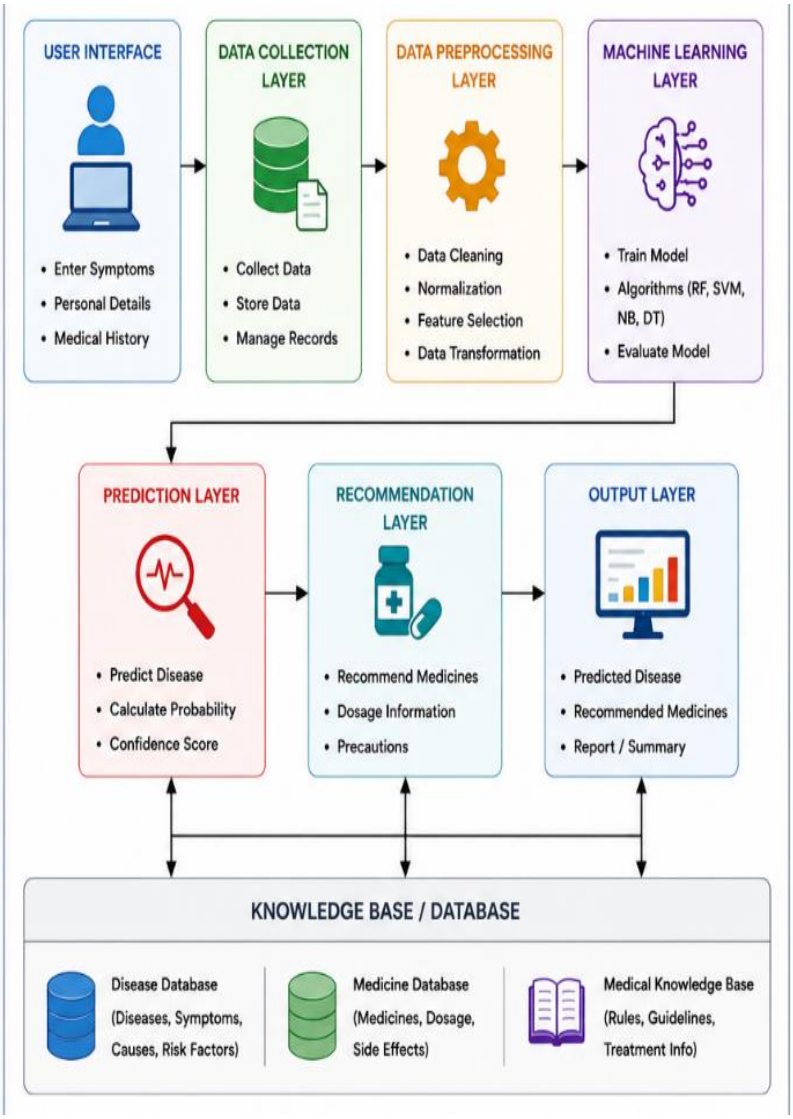


Fig 1. System Architecture

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

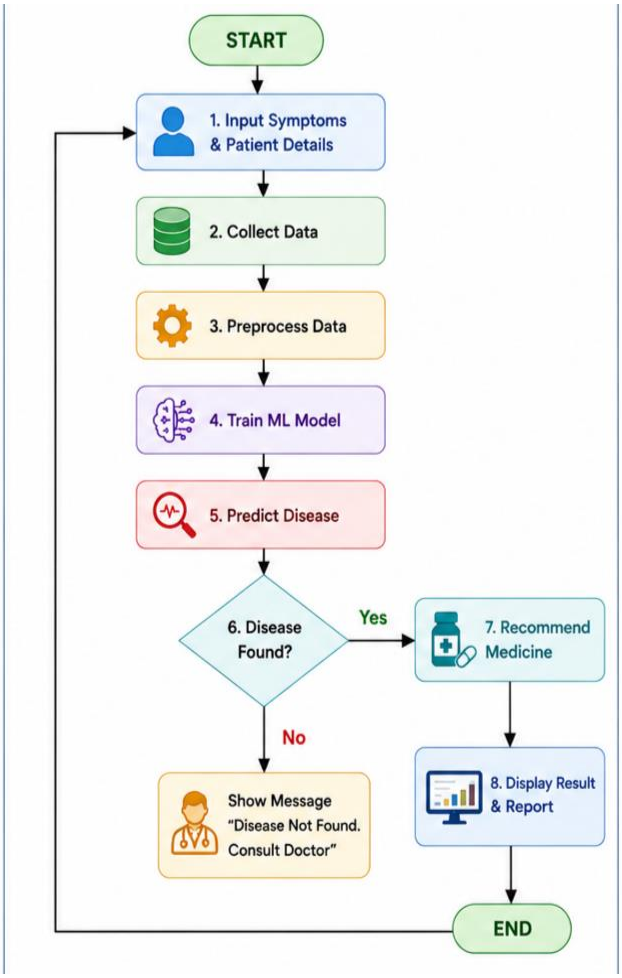


Fig 2. System Design (Flowchart)

VI. MACHINE LEARNING ALGORITHMS USED

.

A. Random Forest

Random Forest constructs a collection of decision trees, each trained on a randomized bootstrap sample of the dataset with a random feature subset, thereby inducing diversity among learners and mitigating overfitting. Aggregated predictions are resolved through plurality voting.

The prediction formula is:

$\hat{Y}$ = mode {T1(x), T2(x), T3(x), ..., Tn(x)}

Where:
$\hat{Y}$ = Final predicted class
T1(x), T2(x), ... = Predictions from decision trees
mode = Majority voting result
Operational Mechanism of Random Forest

Every constituent tree independently predicts a disease label based on the input symptom profile. The Random Forest aggregator designates the plurality-voted class as the definitive output.

B. Naïve Bayes

Naïve Bayes is a probabilistic algorithm based on Bayes' theorem. It performs well for symptom-based disease classification due to its simplicity and computational efficiency.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

The prediction formula is:

$\hat{Y}$ = mode {T1(x), T2(x), T3(x), ..., Tn(x)}

Where:
$\hat{Y}$ = Final predicted class
T1(x), T2(x), ... = Predictions from decision trees
mode = Majority voting result
Operational Mechanism of Random Forest

Every constituent tree independently predicts a disease label based on the input symptom profile. The Random Forest aggregator designates the plurality-voted class as the definitive output.

VII. RESULTS AND DISCUSSION

Predictive performance varied across ML algorithms as a function of dataset quality, dimensionality, and preprocessing rigor. Random Forest and SVM consistently achieved higher classification accuracy relative to simpler baseline models, owing to their capacity to exploit complex feature interactions.
Unifying disease forecasting and therapeutic guidance within a single platform meaningfully enhances healthcare accessibility and reduces clinician dependency for preliminary triage tasks. The integrated design also compresses diagnostic turnaround time, offering patients timely preliminary guidance.
System validation was conducted using diverse multi-symptom test cases; results confirmed high diagnostic concordance and clinically coherent drug recommendations. The Gradio-based front-end interface facilitated seamless real-time interaction between end-users and the underlying prediction engine.

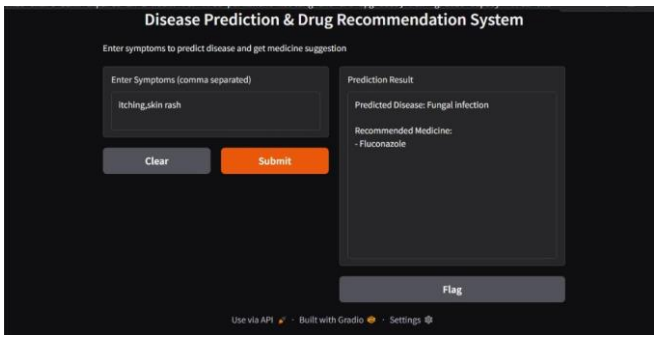


Fig. 2. Disease Prediction & Drug Recommendation System – Initial Interface.

Key operational challenges encountered during system development include:

1. **Constrained medical dataset availability** impairs model generalization across rare or underrepresented disease categories.
2. **Symptom overlap across multiple clinical entities** introduces ambiguity, potentially degrading predictive specificity.
3. **Class imbalance in training data** where prevalent conditions outnumber rare ones, may introduce systematic prediction bias.
4. **Sufficient training data volume** is a prerequisite for attaining clinically acceptable reliability and predictive robustness.

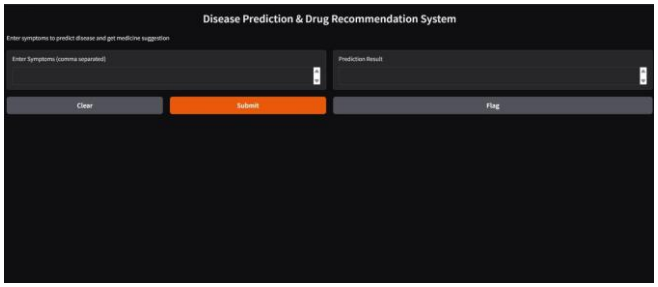


Fig. 3. Prediction Result: Fungal Infection (Symptoms: Itching, Skin Rash).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131130

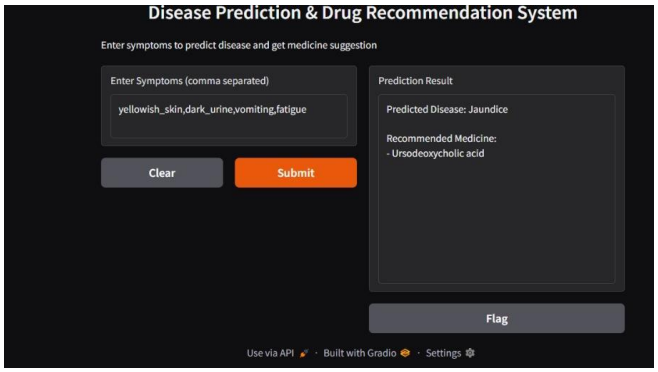


Fig. 4. Prediction Result: Jaundice (Symptoms: Yellowish Skin, Dark Urine, Vomiting, Fatigue).

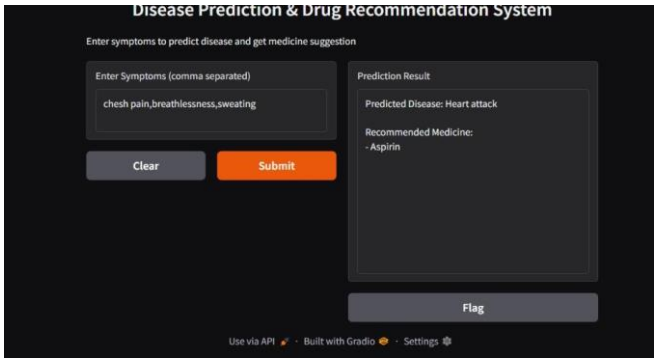


Fig. 5. Prediction Result: Heart Attack (Symptoms: Chest Pain, Breathlessness, Sweating).

X. CONCLUSION

This paper systematically reviewed contemporary ML-based clinical decision-support systems for disease identification and therapeutic recommendation. Multiple learning paradigms and their underlying healthcare data processing strategies were critically examined. The investigation substantiates that intelligent automated systems can materially elevate diagnostic accuracy, alleviate clinical workload, and expedite medical guidance for end-users. While operational impediments such as limited annotated corpora and overlapping symptom presentations remain, ML-driven approaches carry compelling potential to reshape healthcare delivery. Continued advances in deep learning, federated training, and explainable AI are anticipated to further amplify system reliability and clinical applicability.

XI. FUTURE SCOPE

Prospective development directions encompass:

1. Seamless interoperability with live hospital management systems and Electronic Health Record (EHR) infrastructure.
2. Adoption of advanced deep learning paradigms including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks.
3. Incorporation of natural language processing and speech recognition for hands-free, voice-activated symptom submission.
4. Cross-platform mobile application deployment targeting Android and iOS ecosystems to broaden user reach.
5. Multilingual interface support to serve linguistically diverse patient populations globally.
6. Embedded telemedicine functionality enabling real-time, in-app consultation with licensed practitioners.

REFERENCES

[1]. A. Grover, A. Kalani and S. K. Dubey, "Analytical Approach towards Prediction of Diseases Using Machine Learning Algorithms," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 793-797, doi:10.1109/Confluence47617.2020.905812

[2]. L. Gong, X. Wei, H. Ji, H. Huang and Y. Ji, "A Graph-based Multi-dimensional Interaction Network for Drug-Drug Interaction Prediction," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2025.3575012.

DOI: 10.17148/IARJSET.2026.131130

[3]. A. K. G, K. Deepa, S. T, S. R. Kannan, M. J. C. R and S. V. E, "Deep Belief Network Approach for Symptom-Based Disease Classification," 2025 2nd International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICIAIHI), Raipur, India, 2025, pp. 1-6, doi:10.1109/ICIAIHI67124.2025.11403244.

[4]. S. Muktadir, M. Ridoy, S. Rayhan and M. S. Mahmood, "Enhancing Clinical Decision-Making: Comparative Evaluation of Machine Learning Models for Symptom-Based Disease Diagnosis," 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), Rangpur, Bangladesh, 2025, pp. 1-6, doi:10.1109/QPAIN66474.2025.11171753.

[5]. Y. Lietal., "MRGCDDI: Multi-Relation Graph Contrastive Learning Without Data Augmentation for Drug-Drug Interaction Events Prediction," in IEEE Journal of Biomedical and Health Informatics, vol. 30, no. 2, pp. 1735-1745, Feb. 2026, doi: 10.1109/JBHI.2024.3483812.

[6]. S. Muktadir, M. Ridoy, S. Rayhan and M. S. Mahmood, "Enhancing Clinical Decision-Making: Comparative Evaluation of Machine Learning Models for Symptom-Based Disease Diagnosis," 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), Rangpur, Bangladesh, 2025, pp. 1-6, doi:10.1109/QPAIN66474.2025.11171753.

[7]. A. K. G, K. Deepa, S. T, S. R. Kannan, M. J. C. R and S. V. E, "Deep Belief Network Approach for Symptom-Based Disease Classification," 2025 2nd International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICIAIHI), Raipur, India, 2025, pp. 1-6, doi:10.1109/ICIAIHI67124.2025.11403244.

[8]. Abbas, S., Ahmed, F., Khan, W. A., et al. "Intelligent Skin Disease Prediction System Using Transfer Learning and Explainable AI." Scientific Reports, 2025. DOI: 10.1038/s41598-024-83966-4

[9]. Disease Prediction System in Human Beings Using Machine Learning Approaches," In Proceedings of the International Conference on Advances in Electronics, Computing and Communication Technologies (ICAECT), 2024, doi: 10.1109/ICAECT60202.2024.10468978

[10]. H. Lv, Z. Chen, Y. Yang, S. Pan, B. Xiong and C. Yang, "Enhancing Semantic and Structure Modeling of Diseases for Diagnosis Prediction," in AMIA Annual Symposium doi:10.48550/arXiv.2501.01234.

[11]. K. Vignesh, C. Sreenidhi, B. Pranavi and B. Baranitharan, "Symptom-Based Health Recommendation System Utilizing Machine Learning Approach," 2025 IEEE doi:10.1109/ICCRTEE64519.2025.11052965.

[12]. S. Dass, K. Jaiswal and S. Kumar, "A Machine Learning-Based Recommendation System for Disease Prediction,"SSRN Research Paper, 2025, doi:10.2139/ssrn.5207379.

[13]. R. Kumar and N. Gupta, "Hybrid Deep Learning Model for Multi-Disease Prediction in Smart Healthcare," IEEE doi:10.1109/ACCESS.2025.3524781.

[14]. T. Nguyen and P. Lee, "Deep Reinforcement Learning-Based Drug Recommendation System for Chronic Disease Management," IEEE Access, vol. 14, pp. 22451–22463, 2026, doi:10.1109/ACCESS.2026.3548721.

[15]. S. Ahmed, R. Khan and F. Ali, "Explainable AI Techniques for Disease Diagnosis and Medicine Recommendation," 2026 IEEE International Conference on Data Science and Artificial Intelligence, pp. 91–97, doi:10.1109/DSAI63112.2026.11325674.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License