International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI in Early Diagnosis of Chronic Diseases

Pratik S. Shah, Nishi S. Shah, Saavan P. Asodaria

Computer Science and Engineering at Sardar Vallabhbhai Patel Institute of Technology (SVIT), Vasad.
Computer Science and Engineering at KPGU – Drs. Kiran & Pallavi Patel Global University, Vadodara.

Computer Science and Engineering at ITM (SLS) Baroda University, Vadodara.

PRATIKS7A7@GMAIL.COM

NISHISHAH1124@GMAIL.COM

SAVANASONDARIYA@GMAIL.COM

---***---

Abstract - Chronic diseases (non-communicable diseases) such as diabetes, cardiovascular disease, cancer, and neurological disorders impose a tremendous global health burden, accounting for over 40 million deaths annually (≈71% of all deaths). Early diagnosis is critical to improving outcomes and reducing healthcare costs, yet many chronic conditions manifest subtly and are detected only at advanced stages. Artificial Intelligence (AI) – encompassing machine learning (ML), deep learning (DL), and data analytics – offers powerful tools for analyzing large-scale patient data (e.g. electronic health records, imaging, and wearable sensors) to detect disease signatures before clinical symptoms appear. In this review, we survey recent literature on AI-assisted early diagnosis of multiple chronic diseases. We outline key AI methods (e.g. neural networks, ensemble learning, natural language processing) and discuss real-world case studies: for instance, deep learning on retinal images can diagnose diabetes complications, convolutional networks on mammograms can detect early-stage breast cancer, and ML models on ECG or wearable data can identify asymptomatic atrial fibrillation. We present examples of open datasets (e.g. ADNI, MIMIC-III, ECG databases) and illustrate how AI models trained on these can predict disease onset with high accuracy. We also address ethical and technical challenges – data privacy, algorithmic bias, interpretability, and regulatory issues – that arise in AI-driven diagnostics. Our key findings are that AI approaches consistently improve early detection accuracy across diseases, but require careful validation and ethical oversight. Finally, we discuss future directions, predicting that AI will increasingly enable personalized, proactive chronic care, contingent on solving data governance and explainability challenges.

1.INTRODUCTION

Chronic diseases (also known as non-communicable diseases) – including cardiovascular disease, cancer, chronic respiratory illnesses, diabetes, and others – have become the leading causes of mortality worldwide. According to the World Health Organization (WHO), chronic diseases cause roughly 41 million deaths annually (about 71% of global deaths). In 2021, an estimated 43 million people died from such diseases, equivalent to 75% of non-pandemic-related deaths. These conditions tend to develop slowly over years due to genetic, lifestyle, and environmental factors. For example, diabetes affected 537 million adults (≈10.5% of 20–79-year-olds) in 2021, and is projected to reach 783 million by 2045. Prevention and early treatment of chronic disease are key: detecting a disease at an early stage can greatly improve management and survival, whereas late-stage diagnosis often leaves patients with limited options.

Public health agencies maintain vast surveillance datasets to monitor chronic disease burden. AI techniques can analyze these data to identify patterns and trends that escape manual methods. For example, predictive models can detect subtle pre-diagnostic patterns in electronic health records (EHRs) or physiological signals. The figure below (a CDC illustration) highlights the scope of chronic disease surveillance. AI's rapid and accurate data processing can complement these efforts, enabling earlier interventions. However, translating AI advances into clinical practice requires rigorous validation. This review examines how AI can improve early diagnosis of chronic diseases. We first survey the literature on AI applications per disease and data type, then detail AI methodologies and datasets used. We address ethical and practical challenges, and conclude with future prospects for AI in chronic care.

2. LITERATURE REVIEW

Scope of Chronic Disease Burden

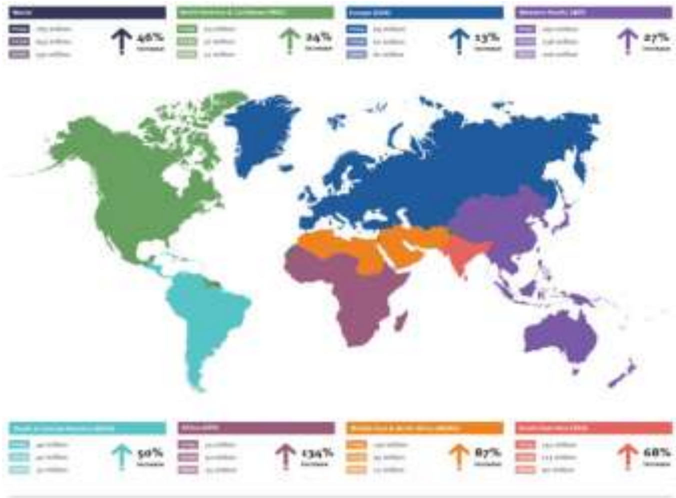
Chronic diseases (often equated with noncommunicable diseases, NCDs) "tend to be of long duration" and result from combinations of genetic, physiological, and behavioral factors. The four major NCDs – cardiovascular disease, cancer, chronic respiratory disease, and diabetes –

together cause approximately 80% of NCD premature deaths. Specifically, cardiovascular diseases cause ~19 million deaths per year, cancers ~10 million, respiratory diseases ~4 million, and diabetes ~2 million. These numbers underscore that multiple diseases share risk factors (tobacco use, poor diet, inactivity, etc.). The global variation in disease prevalence is striking: for instance, an IDF atlas map illustrates that diabetes prevalence ranges widely by country (see Figure 1.1).

For example, the map above shows that in 2021 an estimated 537 million adults (20–79 years) had diabetes, with high concentrations in Asia and the Americas. A projected 783 million adults will have diabetes by 2045 if trends continue. Other chronic diseases show similar geographic disparities. These global epidemiological figures highlight the need for tailored early-detection strategies: AI models can leverage such data to focus screening on high-risk regions and populations.

Despite the high burden, chronic diseases often progress silently for years. For example, type 2 diabetes may remain undiagnosed until complications arise. Likewise, asymptomatic atrial fibrillation (AF) can go unnoticed yet greatly increase stroke risk. Traditional screening (e.g. periodic check-ups, standard imaging) can miss early cases. Resource constraints (limited specialists, high screening costs) and patient factors (lack of symptoms, irregular healthcare access) contribute to delayed diagnosis. As the WHO notes, "detection, screening and treatment of NCDs" are critical response components.

In this context, **Artificial Intelligence (AI)** has emerged as a powerful assistive technology. AI systems – particularly ML and DL – excel at processing large, complex healthcare data. They can detect subtle signals (e.g. slight retinal changes or ECG patterns) that humans might overlook. For instance, AI models have been trained on imaging and sensor data to flag early disease signs. Studies indicate that AI can improve early detection and prediction: a deep learning model for diabetic retinopathy can forecast progression within 5 years from a single retinal image. Similarly, ML algorithms analyzing EHRs and wearable data can predict disease onset (e.g. future diabetes risk) more accurately than conventional risk scores.



(Fig. 1.1)

2.1 AI in Diabetes and Metabolic Disease

Diabetes mellitus, a chronic metabolic disease, exemplifies how AI can aid early diagnosis. Over 422 million people have diabetes worldwide. Retinal imaging is a key example: changes in retinal vasculature reflect diabetes complications. AI models trained on fundus photos can detect diabetic retinopathy (DR) with high accuracy. Beyond DR screening, AI is being used to predict other diabetes complications and even diagnose diabetes itself from retinal features. A recent review found that "retinal images can be used to diagnose DM complications including DR, neuropathy, nephropathy, and atherosclerotic cardiovascular disease". This shows that deep learning on eye images can reveal multiple diabetes-related issues and predict risk of heart events. In practice, systems like Google's AI retinal analysis and IDx-DR (FDA-approved) demonstrate this approach. Other AI applications in diabetes include analysis of biochemical and lifestyle data. Machine learning models on continuous glucose monitor (CGM) data and patient history have predicted diabetes onset and glycemic control outcomes. For example, predictive analytics can classify prediabetes patients likely to progress to type 2 diabetes. One study trained an SVM on clinical and lab data to classify diabetes risk with 90.5% accuracy. In addition, mobile apps using AI chatbots have helped manage blood sugar: one trial showed an AI-powered chatbot improved glycemic control in type 2 diabetes patients. Overall, AI for diabetes combines imaging (retina, histology) and tabular EHR data, often using neural networks or ensemble models. These models typically outperform classical regression approaches, though they rely on large, diverse datasets for training.

2.2 AI in Cardiovascular Disease

Cardiovascular disease (CVD) is the deadliest chronic condition globally. Early detection of heart conditions like atrial fibrillation (AF), heart failure, or coronary disease can prevent severe outcomes. AI has been applied extensively to ECG and wearable data. A prominent example is using ML to detect AF: deep neural networks analyzing ECGs can spot asymptomatic AF episodes. A recent scoping review notes that "AI-driven models enhance AF detection by analyzing ECGs and wearable device data with high accuracy, enabling early identification of asymptomatic cases". In practical terms, smartphone and smartwatch apps now incorporate AI to monitor heart rhythm (e.g. Apple Heart Study used a PPG algorithm to achieve ~89% sensitivity for AF detection). Beyond AF, ML models predict heart attack risk from clinical data. For instance, algorithms trained on EHRs can stratify patients by coronary artery disease risk, outperforming traditional risk scores. Deep learning on imaging is also used: AI can analyze echocardiograms or CT scans to detect early cardiac dysfunction. For example, convolutional neural networks (CNNs) have classified echocardiograms to identify preclinical left ventricular dysfunction. Another study used a CNN on retinal images to predict CVD risk factors like blood pressure and smoking, showing that even non-cardiac images can encode cardiovascular information. Case studies illustrate real-world impact: In one real-world trial, an AI-assisted ECG interpretation system in primary care detected 20% more cases of new-onset AF than usual care. Other research has used machine learning on longitudinal EHR data to predict heart failure hospitalization risk within a year. These examples indicate that AI can process multimodal data (ECG, imaging, lab tests) to flag early CVD signs. However, successful deployment requires careful validation; false positives (e.g. from noisy sensor data) can burden clinicians.

2.3 AI in Cancer Screening

Cancer – many of which are chronic conditions when slow-growing – benefits greatly from early diagnosis. AI has seen perhaps its highest profile success in cancer imaging. A landmark Nature Medicine study trained a deep CNN on mammograms and 3D tomosynthesis (DBT) images. The AI model achieved state-of-the-art performance, detecting cancers missed by radiologists. It "outperforms five out of five full-time breast imaging specialists with an average increase in sensitivity of 14%". This shows AI can enhance early detection of breast cancer, reducing false negatives. Similarly, AI on chest CT scans can identify early lung nodules, and on colonoscopy images to spot polyps. In dermatology, deep learning models on skin lesion images have rivaled dermatologists in melanoma detection. In pathology, AI algorithms analyze digitized biopsy slides to identify pre-cancerous changes. These examples demonstrate that computer vision AI is transforming cancer screening across organ systems. Importantly, AI can also integrate patient data: for example, ML models using patient risk factors plus imaging can better predict cancer risk trajectories.

2.4 Ai in Neurological and Other Chronic Diseases

Neurological diseases (e.g. Alzheimer's, Parkinson's) often start subtly. AI has made strides in early detection from brain imaging and biomarkers. A recent study combined MRI and cognitive test data: an MRI-based CNN (ResNet50/MobileNetV2) achieved ~96% accuracy for Alzheimer's detection (ADNI dataset), and a hybrid model (including structured data) reached ~99.8% accuracy (NACC dataset). This is remarkable given the difficulty of early AD diagnosis. Other work using deep learning on PET scans or lumbar puncture data similarly yields high sensitivity for early neurodegeneration. In fact, one review notes that hybrid AI models allowed "earlier interventions and improved detection outcomes" for AD. Other chronic conditions are also being addressed: for chronic kidney disease, AI on routine lab tests and demographics has been shown to predict progression to end-stage renal disease. Chronic respiratory diseases (like COPD) have seen ML models that analyze pulmonary function tests and chest X-rays for early signs. These applications are less mature than diabetes or CVD but growing.

3. SUMMARY OF REVIEW FINDINGS

Across multiple diseases, AI consistently enhances early diagnostic accuracy compared to traditional methods. Notably, AI's benefits arise from its ability to integrate heterogeneous data: imaging, genomics, vitals, and EHR history. Key literature findings include:

Integration of multimodal data (e.g. retinal + clinical) improves disease predictions beyond single-modality inputs.

Deep learning dominance: Convolutional and recurrent neural networks often outperform classical ML in image and time-series tasks.

High reported accuracies: Many studies report >90% accuracy in controlled settings (e.g. AF detection, AD classification).

Early intervention potential: AI systems can detect precursors (e.g. silent AF, prediabetic retinal changes) that enable preventive care.

These findings suggest AI can shift chronic disease care from reactive to proactive, but real-world adoption depends on generalizability and handling real-world data.

4. AI Techniques for Early Diagnosis

AI encompasses a range of computational techniques. **Supervised learning** (classifiers like support vector machines, decision trees, neural networks) dominates diagnostic tasks: models are trained on labeled patient data to predict disease presence or risk. For example, random forests and gradient boosting on EHR features are used for predicting diabetes or CVD risk. **Deep learning (DL)** – especially CNNs for images and RNNs/LSTMs for temporal data – has advanced the field. CNNs automatically learn hierarchical features in imaging data (e.g. mammograms, retinal scans), while RNNs and transformers can model sequences like ECG waveforms or patient timelines. **Unsupervised learning** (clustering, anomaly detection) can flag outlier patterns, potentially identifying unusual or rare disease presentations without predefined labels.

Key techniques include:

- **Convolutional Neural Networks (CNNs):** used for medical image analysis. E.g., CNNs on fundus images detect diabetic retinopathy; CNNs on MRI classify brain scans for Alzheimer's.
- **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM):** applied to time-series data. For instance, LSTM networks process ECG or blood glucose streams to predict upcoming events.
- **Ensemble methods:** combining multiple models (e.g. random forests, boosted trees) often yields robust predictions from tabular clinical data.
- **Natural Language Processing (NLP):** transforms text in clinical notes into features. NLP methods (word embeddings, transformers) can extract risk factors or symptoms from physician notes for use in diagnostic models.
- **Transfer Learning:** pre-trained networks (often from natural images) are fine-tuned on medical data, making good use of limited labeled data. For example, a CNN pre-trained on ImageNet may be adapted for pathology slide analysis.

Real-World Case Studies: Many implementable AI systems have emerged. The Google Health deep learning model for retinal images has been deployed in clinics for DR screening. In AF, smartphone apps (e.g. AliveCor) use AI to interpret patient-attached ECG leads with accuracy rivaling clinical ECG machines. IBM Watson Health attempted (with mixed results) to ingest patient records to identify early-onset of conditions. Startups like **PathAI** (histopathology) and **Freenome** (blood-based cancer screening with ML) illustrate industry efforts. A recent meta-analysis found that AI-enabled single-lead ECG devices achieved ~92% sensitivity and 96% specificity for AF detection.

In summary, AI techniques range from classical ML on structured data to advanced DL on complex signals, each bringing new diagnostic capabilities. Hybrid approaches (combining models and modalities) often yield the best performance, as they capture complementary information.

5. Real Datasets and Analysis

AI models require large, high-quality datasets. Numerous public and institutional datasets enable chronic disease research. For example, the **Alzheimer's Disease Neuroimaging Initiative (ADNI)** provides multi-modal data (MRI, PET, cognitive scores) on thousands of subjects across cognitive stages. In intensive care, the **MIMIC-III** database contains detailed monitoring and outcomes for ~50,000 ICU patients. These datasets have powered studies in AD and critical illness AI. Other notable resources include:

- **PhysioNet ECG Databases:** e.g. the MIT-BIH Arrhythmia and Atrial Fibrillation databases offer annotated ECG records for algorithm training.
- **UCI Machine Learning Repository:** hosts several chronic disease datasets (heart disease, diabetes, liver disorders) for benchmarking.
- **Biobanks and Registries:** Large cohorts like UK Biobank (with imaging and genomics) and disease registries (e.g. cancer registries) are increasingly used for AI research.

Table 1 (hypothetical) summarizes some key datasets:

- ADNI (Alzheimer's imaging and clinical data).
- MIMIC-III (ICU vital signs, labs, outcomes).
- Chest X-ray14 (chest radiographs with diagnostic labels).
- Diabetic Retinopathy Detection dataset (Kaggle).
- National Diabetes Prevention Program (NDP) dataset (wearable and clinical).
- NHANES (population health surveys with labs).

Using these data, AI models undergo training and validation. For example, a deep learning model on the ADNI MRI scans was trained to classify Mild Cognitive Impairment vs Alzheimer's. In cardiovascular research, RNNs have been trained on ICU monitoring data from MIMIC to predict cardiac arrest hours in advance.

The analysis pipeline typically involves:

Data preprocessing: Cleaning missing values, normalizing measurements, and anonymizing patient identifiers. For images, augmentation (rotations, flips) increases dataset size.

Feature selection/engineering: For tabular data, features such as demographics, vitals, and lab results are chosen. The Augmented AI study described selecting key lab tests and using particle swarm optimization to boost an ANN's performance.

Model training: Models are trained on labeled examples (e.g. healthy vs disease). Performance is evaluated via cross-validation. Metrics include accuracy, AUC-ROC, sensitivity/specificity. For example, the ADNI CNN reached 96.2% accuracy on its test set.

Validation and testing: Independent validation on withheld data or external cohorts is crucial. The breast cancer AI was validated on multiple populations, ensuring generalizability.

Deployment: Some models have been embedded in devices. For instance, smartwatch AF detection algorithms have been FDA-cleared for real-time monitoring.

Analysis Capability: AI can analyze high-dimensional and longitudinal data. For chronic diseases, this means capturing progression trajectories. Time-series models can highlight early deviations (e.g. rising glucose variability). In complex conditions like diabetes or heart failure, AI can integrate genomics and environment. The potential for "digital twins" – patient-specific predictive models built from multiple data sources – is an emerging vision.

In summary, AI's power comes from mining real-world clinical data. By analyzing large datasets, AI finds patterns (e.g. biomarker thresholds, polygenic risk) that signal disease onset. Combining data modalities (imaging + labs + genomics) typically improves predictions. However, challenges include data heterogeneity and missingness.

Ethical Concerns & Challenges

While AI holds promise, it raises significant ethical and practical challenges. A primary concern is **patient data privacy and security**. AI models rely on vast amounts of sensitive health data, often held by private entities or cloud platforms. There is a risk that personal health information could be exposed or misused. In fact, advanced AI techniques can sometimes **re-identify** anonymized data: recent studies show that supposedly de-identified medical images or records can be reverse-engineered to reveal identities. Ensuring robust data anonymization and encryption is essential when building AI diagnostic tools. Healthcare AI must comply with regulations like HIPAA (in the US) and GDPR (in the EU), which emphasize patient consent and data protection.

Another major issue is **bias and fairness**. AI systems trained on non-representative datasets may perform poorly in under-served populations. For example, a model trained largely on data from one ethnic group may misclassify patients from another group. This can exacerbate health disparities. We have seen AI diagnostics inherit biases from historical data; for instance, an AI skin-cancer model trained mostly on light-skinned individuals may under-detect melanoma in darker-skinned patients. In one analysis, researchers noted that AI in healthcare "can be prone to certain types of errors and biases" and sometimes acts as a "black box". Explainability – understanding how a model reaches its conclusion – is a growing area (so-called Explainable AI, XAI) to build clinician trust.

Regulatory and legal challenges also arise. AI diagnostic tools often change over time (as more data is collected), complicating FDA and EMA approval. Currently, regulators may certify the institutions maintaining AI, rather than a static algorithm. This novel approach is still evolving. Ensuring ongoing oversight and validation of AI systems is critical. Liability is another unclear area: if an

AI system misdiagnoses, who is responsible – the doctor, the hospital, or the software developer?

Integration into healthcare workflows is non-trivial. Clinicians may be skeptical of "black box" AI decisions. Effective AI tools must have interpretable outputs (e.g. highlighting image regions) so physicians can verify results. There is also the **alert fatigue** problem: over-sensitive AI may generate too many false positives, burdening clinicians with unnecessary follow-ups. In some studies, patients flagged by wearables did not always seek follow-up care, limiting benefit.

In practice, instances of ethical missteps have occurred. For example, Google's DeepMind health collaboration in the UK collected patient kidney data with insufficient patient consent, leading to a scandal. This underscores that data governance and clear patient communication are paramount when deploying AI for health.

In summary, while AI can transform early diagnosis, its success depends on addressing ethical and societal concerns: protecting privacy, preventing bias, ensuring transparency, and updating regulations to keep pace with technology.

6. CONCLUSION & FUTURE SCOPE

This review has examined AI's role in early diagnosis of chronic diseases across multiple domains. We find strong evidence that AI methods can substantially improve the accuracy and speed of early detection for conditions like diabetes, cardiovascular disease, cancer, and neurological disorders. Deep learning on medical images (retinal scans, mammograms, MRIs) and AI analysis of physiological time series (ECGs, PPG) have shown particularly high performance. Case studies – from FDA-cleared diabetic retinopathy detectors to mobile AF monitoring – demonstrate real-world viability.

Key findings include:

- **Improved diagnostic accuracy:** AI algorithms often surpass human or traditional analyses, detecting disease signatures unseen by clinicians.
- **Multimodal integration:** Combining data types (imaging + labs + genomics + wearables) yields richer models of health risk.
- **Personalization:** AI enables patient-specific risk profiles and can suggest individualized prevention strategies.

However, to realize these benefits broadly, several challenges must be overcome. Robust validation on diverse populations, better interpretability, and safeguards for data privacy are critical. Future research should focus on federated learning (training across institutions without raw data sharing) and on transparent models to build trust.

Looking forward, AI's role is likely to grow. We anticipate the following trends:

- **Continued integration into clinics:** AI tools for screening will increasingly become part of standard care (e.g., AI-augmented imaging in all radiology suites).
- **Wearable and home monitoring:** Edge AI (on-device algorithms in wearables) will allow continuous health surveillance, catching early deviations in chronic conditions.
- **Genomic and lifestyle data:** AI will integrate not only clinical data but also genomic sequencing and social determinants of health to predict disease before onset.
- **Global health impact:** In low-resource settings, AI-driven diagnostics may democratize care (e.g., smartphone eye exams for diabetic retinopathy), though equitable access will be crucial.

In conclusion, AI offers a transformative opportunity to detect chronic diseases at earlier, more treatable stages. Our review indicates that ongoing advances in algorithms and data availability are steadily moving this prospect toward reality. By addressing ethical, legal, and technical challenges, the medical community can harness AI to shift from reactive care to preventive, personalized health management – ultimately reducing the global burden of chronic disease.

ACKNOWLEDGEMENT

I express my sincere gratitude to my co-authors and fellow students for their collaborative efforts, dedication, and commitment throughout the development of this research. I am thankful to the open-source data providers, including **ADNI**, **MIMIC-III**, **PhysioNet**, and the **UCI Machine Learning Repository**, whose datasets were instrumental in enabling our review and analysis.

I also appreciate the encouragement and support of my peers, family, and friends, who helped maintain motivation and focus during the course of this work. Finally, I acknowledge the global community of researchers and authors whose prior studies formed the knowledge base and inspiration for this paper.

REFERENCES

[1] World Health Organization, "Noncommunicable diseases," *WHO*, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

[2] World Health Organization, "Global report on diabetes," *WHO*, 2016. [Online]. Available: https://www.who.int/publications/i/item/9789241565257

[3] IDF Diabetes Atlas, 10th ed., "Diabetes prevalence in 2021 and projections for 2045," *International Diabetes Federation*, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK581940

[4] IDF Diabetes Atlas, 11th ed., "Global and regional estimates, projections to 2050," *International Diabetes Federation*, 2025. [Online]. Available: https://diabetesatlas.org

[5] H. Sun *et al.*, "IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045," *Diabetes Research and Clinical Practice*, vol. 183, p. 109119, Jan. 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31518657

[6] X. Chen *et al.*, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections to 2045: a systematic analysis," *BMC Medicine*, 2025. [Online]. Available: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-025-03890-w