WEB BASED EARLY DISEASE DETECTION USING AI AND ML

Poonam B. Awari
College Of Management And Computer Scince ,Yavatmal
poonamawari54@gmail.com

Abstract
Early detection of diseases plays a crucial role in improving patient outcomes and reducing healthcare costs. However, in many regions, limited access to timely diagnosis and medical expertise leads to delays in disease identification. This research presents the design and development of a web-based early disease detection system that leverages artificial intelligence and machine learning algorithms to predict possible diseases based on user-entered symptoms and health parameters. The system provides an accessible online platform where users can input symptoms through a secure web interface. A central prediction engine processes the input using a trained machine learning model—integrating algorithms such as Naïve Bayes, Decision Trees, and Random Forest—to generate probable disease outcomes along with confidence scores. In addition, the platform offers health recommendations and directs users to appropriate medical specialists for further evaluation. The system aims to assist in preliminary screening, particularly in rural and under-served areas, by providing real-time, data-driven insights without requiring in-person consultations. Experimental results demonstrate high prediction accuracy and user satisfaction in pilot testing.

Keywords: *Artificial intelligence (AI) in healthcare , Machine learning (ML) , Predictive analytics, Symptom-based diagnosis , Health informatics ,Medical decision support system ,Online health platform , Remote diagnosis ,Telemedicine , Data-driven healthcare ,Healthcare web application*

Introduction

Early detection of diseases is a critical factor in improving patient outcomes and reducing the burden on healthcare systems. Timely diagnosis enables effective treatment, limits disease progression, and lowers mortality rates. However, in many developing and rural regions, access to healthcare facilities and medical experts remains limited, often leading to delayed diagnosis and inadequate treatment. With the rapid advancement of digital technologies and the widespread availability of the Internet, web-based health systems have emerged as powerful tools for bridging this gap and enhancing preventive healthcare delivery.

A web-based early disease detection system integrates the principles of artificial intelligence (AI), machine learning (ML), and data analytics to predict potential diseases based on user-reported symptoms, physiological parameters, or clinical data. By allowing individuals to enter their symptoms through an online interface, such systems can process the information using trained ML algorithms—such as Decision Trees, Naïve Bayes, or Random Forest—to identify probable diseases and provide initial medical recommendations. These systems can further assist healthcare providers by acting as a preliminary screening mechanism and facilitating remote consultations.

Existing healthcare applications primarily focus on disease management rather than prevention. In contrast, a web-based early disease detection system emphasizes preventive healthcare, enabling users to identify risk factors before severe symptoms appear. Moreover, these systems can be integrated with electronic health records (EHRs), wearable sensors, and IoT devices to continuously monitor health status, detect anomalies, and send alerts for early intervention. The COVID-19 pandemic has further highlighted the importance of web-based tools in enabling remote health assessment, reducing hospital congestion, and supporting telemedicine initiatives.

Literature Review

Early detection of disease has become a central goal in both clinical and public-health domains, driven by the need to reduce morbidity, mortality and health-system burden. With the proliferation of internet access, digital data sources and web-based platforms, there is growing interest in using web-based solutions and online data for early disease detection. The literature can be grouped into three overlapping streams: (1) web-based surveillance systems for infectious or outbreak diseases; (2) web or online patient-facing portals and symptom-based disease prediction systems; and (3) machine-learning/AI-enabled systems using web/data interfaces for early diagnosis of non-infectious conditions.

1. Web-based surveillance & outbreak detection

Several systematic reviews show that web-based data sources (e.g., search engine queries, social media posts, news feeds) and web-applications now play a supplementary role to traditional disease surveillance. For example, the study by Choi et al. in *BMC Public Health* reviewed web-based

infectious disease surveillance systems from 2000–2015 and found that such systems are intuitive, adaptable and real-time, offering advantages over traditional surveillance systems, although they still face limitations in accuracy and interpretation of web-data.

Likewise, O'Shea (2017) reviewed event-based internet "biosurveillance" systems and identified 50 internet-based systems, showing that internet data sources are increasingly diverse and that these systems are best used to supplement traditional surveillance rather than completely replace it.

2. Web-based portals / symptom-reporting systems for individual-level disease detection

A somewhat different stream of work focuses on portals or web-interfaces where users input symptoms or health-parameters and receive a prediction or recommendation. For instance, Mazumder et al. (2013) developed a web-based disease detection system using the ID3 algorithm: the user enters symptoms on a website and the backend predicts probable disease.

More recently, Veeraiah et al. (2023) described a system combining machine-learning and Internet of Things (IoT) data for diagnosing disease using a web application.

Such systems show promise as accessible tools for preliminary screening especially in low-resource settings. However, limitations include: reliance solely on self-reported symptoms (which may be inaccurate); limited validation in clinical settings; and a lack of richness in data (versus hospital/clinical data).

3. Machine-learning / AI in early disease detection using web or online data

Another significant body of work explores how machine-learning (ML) and deep-learning (DL) techniques can use online data (web, search queries, user behaviour) for early disease detection or prediction. For example, Yin et al. (2019) reviewed how machine-learning has been applied to user-generated online personal-health data (UGC: blogs, forums, social media) and noted both the promise and data-quality challenges inherent in such online data.

Similarly, a 2023 review looked at deep-learning techniques for detection and prediction of pandemic diseases (IDs) and found significant progress but also gaps in robustness, generalisability

The "internet-based surveillance + AI" stream thus points toward increasing sophistication (AI, big data, web data fusion) but also emphasises the need for rigorous evaluation, real-world deployment, and attention to bias, privacy and transparency.

4. Gaps and limitations in the literature

From the above streams key limitations emerge that are directly relevant to web-based early disease detection systems:

- Data quality and representativeness: Web-based systems often rely on non-clinical data (search queries, forums, symptom input) that may not be validated or representative of the full population. For example, O'Shea notes that event-based internet systems should "supplement" not replace traditional surveillance.
- Accuracy / validation issues: Many systems report prototype or pilot-level findings; fewer are tested in large-scale real-world clinical settings. Mazumder's 2013 system is quite dated and uses simple decision-tree (ID3) without modern ML.
- Privacy, ethics and trust: Web-based data sources raise concerns of individual privacy, informed consent, and trust in recommendations. The 2016 systematic review flagged privacy as a key limitation.
- Integration with clinical workflows: Many works operate outside of health-system workflows (e.g., stand-alone web portals). Integration with EHRs, clinician input, and real-time decision support remains limited.
- Context-specific adaptation: Systems often developed in high-income settings or with certain diseases; adaptation to low- and middle-income countries, multi-lingual settings, or lesser-known diseases remains a gap. For example, the digital technology study in neglected tropical diseases (NTDs) shows limited electronic-technology work in certain disease domains.
- Explainability, user interface & adoption: For web-based early detection tools meant for lay users (symptom input, suggestions) the usability, trust, language/localisation, and explanation of predictions are often not addressed.
- Focus on non-infectious diseases or chronic conditions is less common: Many web-based detection systems focus on infectious/outbreak surveillance; fewer robust systems target chronic disease early detection via web/symptom-input portals.

5. Implications for current research

Given the reviewed literature, a web-based early disease detection system that aims to combine user-symptom/health-parameter input, intelligent ML backend, and real-time web interface must pay special attention to:

- Selecting appropriate data sources (symptoms, user input, possibly wearable/sensor data) and establishing accuracy/validation metrics.

- Ensuring model generalisability and avoiding bias (especially when targeting underserved populations or non-traditional disease domains).
- Designing the web interface/UX for non-expert users, with clarity, trust and transparency.
- Addressing privacy/security: user-data protection, anonymisation, secure transmission/storage.
- Considering integration or at least alignment with clinical workflows or health-system referral mechanisms (so that an early detection tool can lead to actionable medical follow-up).
- Filling an identified gap: for example, targeting a disease domain not well covered by existing web-based systems (e.g., chronic disease in low-resource settings), or combining early detection with personalised recommendations (e.g., tests, further referrals) through the web interface.

Conclusion

Web-based early disease detection systems represent a significant advancement in preventive healthcare, offering timely, accessible, and cost-effective solutions for populations with limited access to medical facilities. By integrating machine learning algorithms with user-friendly web interfaces, these systems can analyze symptoms and health parameters to provide preliminary disease predictions, risk assessments, and personalized health recommendations. The review of current literature highlights that while existing solutions demonstrate potential in improving early diagnosis and patient engagement, challenges remain in terms of data accuracy, privacy, clinical validation, and user trust.In conclusion, web-based early disease detection has the potential to transform preventive healthcare by empowering individuals with timely insights, facilitating early interventions, and enhancing overall health outcomes. Future work should focus on expanding system accuracy through larger and more diverse datasets, integrating wearable and IoT health data, and ensuring ethical deployment with strong privacy protections, ultimately moving towards a holistic, reliable, and universally accessible digital health solution.

References

1. J. Choi, Y. Cho, E. Shim & H. Woo, "Web-based infectious disease surveillance systems and public health perspectives: a systematic review," *BMC Public Health*, vol. 16, art. no. 1238, Dec. 2016.
2. N. Quinn, K. H. Hsiao, I. Maitland-Scott, M. Gomez, M. T. Baysari, Z. Najjar & L. Gupta, "Web-Based Apps for Responding to Acute Infectious Disease Outbreaks in the Community: Systematic Review," *JMIR Public Health and Surveillance*, vol. 7, no. 4, Apr. 2021, e24330.
3. S. Doan, Q.-H. Ngo, A. Kawazoe & N. Collier, "Global Health Monitor: A Web-based System for Detecting and Mapping Infectious Diseases," preprint, Nov. 2019.
4. H. Rajora, N. S. Punn, S. K. Sonbhadra& S. Agarwal, "Machine learning equipped web based disease prediction and recommender system," arXiv preprint, Jun. 2021.
5. "Establishing a web-based integrated surveillance system for early detection of infectious disease epidemic in rural China: a field experimental study," *BMC Medical Informatics and Decision Making*, vol. 12, art. no. 4, Feb. 2012.