International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 10 October 2025
Page No: 3877-3882
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0563

SymptoScan, an AI-Powered Medical Diagnosis System

C B Gaurav Ganapathy[1], Karthiga G[2], Ajith V[3], Burla Rajamouli[4]
[1,3,4]UG Student, Department of Computer Science and Engineering, AMC Engineering College, Bengaluru, Karnataka, India.
[2]Assistant Professor, Department of Computer Science and Engineering, AMC Engineering College, Bengaluru, Karnataka, India.
Emails: *1am22cs037@amceducation.in[1], karthiga.gurunathan@amceducation.in[2], 1am22cs012@amceducation.i[3], 1am22cs036@amceducation.in[4]*

Abstract
SymptoScan is an AI-based system that analyzes symptoms reported by patients along with their medical histories and diagnostic test results. It employs machine learning, deep learning, and natural language processing (NLP) techniques to achieve this. By connecting with electronic health record (EHR) systems and telemedicine platforms, SymptoScan helps doctors provide accurate diagnoses and timely care, even in remote areas or places with limited resources. The system compares patient input against large clinical datasets to create probability-weighted differential diagnoses. It also provides real-time decision support during virtual consultations. Moreover, SymptoScan includes a medical imaging module for X-ray, MRIs, and CT scans. This module uses convolutional neural networks to identify possible abnormalities, thereby improving the detection accuracy. The platform aims to address common healthcare challenges, such as delays in diagnosis, misdiagnosis, and resource shortages. It also focuses on data security, algorithm fairness, and ethical standards. In summary, SymptoScan merges AI capabilities with medical expertise to create an efficient and accessible diagnostic tool for modern healthcare.

Keywords: *Artificial Intelligence, Machine Learning, Healthcare Diagnostics, Medical Imaging, Telemedicine, Electronic Health Records.*

1. Introduction

The use of AI in medicine could change the delivery of healthcare by tackling urgent challenges and improving clinical decisionmaking. Modern healthcare systems are under pressure from aging populations, increasing rates of chronic diseases, and inequalities in access. This drives the need for AI and big data analytics to improve efficiency and outcomes. Integrating AI algorithms with extensive and diverse clinical data, including structured EHR data and unstructured text, can help build predictive models. These models can assist doctors in making faster and more reliable diagnoses of skin cancer. Recent advances in natural language processing (NLP) and pretrained language models have increased the ability to extract patient information from free-text clinical notes, allowing for better analysis of patient histories. Simultaneously, telemedicine has become vital for reaching patients in remote or underserved areas, but it brings challenges related to documentation and accuracy in diagnosis. AIpowered tools can help bridge this gap by providing decision support during these visits. For example, recent studies have shown that combining AI analytics with telehealth platforms can provide remote clinicians with immediate access to data-driven insights and differential diagnosis suggestions. When there are diagnostic delays or misdiagnoses, especially in rural areas with few specialists, AI can help analyze patient symptom profiles and past records against large medical databases to prioritize possible conditions. Current symptom-checker apps and diagnostic decision-support systems demonstrate how AI can change patient triage; however, many still struggle with accuracy and coverage. SymptoScan is a next-generation platform that combines various data sources, including symptoms, histories, lab tests, and imaging, to enhance diagnostic precision. It aims to support doctors instead of replacing them by providing probability-based diagnosis and explanation. This document outlines the design and implementation of SymptoScan, emphasizing its use of modern machine learning and deep learning techniques. The main contributions include (1) a new diagnostic pipeline that merges NLP-based symptom analysis with structured EHR data; (2) an integrated imaging analysis module for the automated detection of anomalies in medical scans; (3) real-time telemedicine support that extends specialist expertise to remote consultations; and (4) attention to privacy, security, and bias during the system's design.

1.1. Related Work

In recent years, AI and machine learning have been increasingly applied to clinical diagnosis. Deep learning models have achieved and exceeded human-level accuracy in many image-based diagnostic tasks. For instance, convolutional neural networks trained on chest Xrays have outperformed radiologists in detecting pneumonia, and deep networks have been successful in pathology (e.g., identifying lymph node metastases) and dermatology (skin lesion classification). Such studies demonstrate that automated analysis can assist in the screening and interpretation of radiology and pathology images, which is a key component of SymptoScan's imaging module. In the domain of electronic health records, large language models and clinical NLP systems (such as GatorTron) have shown that scaling model size with extensive medical text data can significantly improve performance on tasks such as medical question answering and information extraction from clinical notes [4]. This supports the approach of using NLP to parse patient symptoms and history from text inputs into SymptoScan. Several groups have explored AI-driven symptom checkers and diagnostic assistants. These systems typically allow patients to enter their symptoms and return a list of possible conditions. Although useful, many existing checkers rely on rule-based methods or limited machine learning models. In contrast, SymptoScan uses data-driven models trained on large datasets of patient records and outcomes, aiming for more nuanced probabilistic suggestions. Additionally, SymptoScan integrates differential diagnosis generation (providing a ranked list of likely diagnoses) with clinician-facing interfaces in line with decisionsupport trends. Telemedicine has also been enhanced by AI. Recent literature emphasizes that integrating AI/NLP into telehealth workflows can alleviate the clinician documentation burden and improve remote diagnostics [2]. For example, Reis et al. [2] noted that automating clinical note-taking and analysis through AI before, during, and after remote consultations can free up provider time and improve accuracy. Moreover, systematic reviews have highlighted that AI combined with telemedicine greatly enhances access in rural communities; AI can analyze patient data for early disease detection, whereas telemedicine connects patients to specialists [6]. These studies motivate SymptoScan's dual focus on symptom and data analysis and telehealth compatibility. Another important area of related work is the ethical and fair use of AI in medicine. Bias in training data can lead to AI systems that underperform in underrepresented groups, potentially worsening health disparities [5]. Security and privacy are also major concerns; robust data governance (e.g., de-identification and secure storage) is required to protect sensitive patient information in any AI system [6]. These studies underline the need to design AI diagnostic tools with fairness, transparency, and security from the outset.

2. Method

SymptoScan's architecture integrates several AI components to handle multimodal medical data (Figure 1). First, symptom processing uses NLP: patients describe their symptoms in natural language (by text or speech), which the system transcribes and converts into structured, clinical features. This may involve tokenization, named entity recognition, and semantic parsing to identify key complaints (e.g., "chest pain," "shortness of breath," and "fever duration"). A trained language model (such as a transformer) maps the narrative to medical concepts. In parallel, patient history and test data were incorporated from EHRs, including demographic information, past diagnoses, medication lists, laboratory results, and vital signs. These structured

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 10 October 2025
Page No: 3877-3882
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0563

inputs are additional features of the diagnostic engine. Next, a diagnostic inference engine was used to combine the extracted data. SymptoScan employs machine learning classifiers (e.g., gradientboosted trees or neural networks) that have been trained on large, labeled medical datasets. These models have learned the associations between combinations of symptoms, history, and outcomes. Given a new patient profile, the engine computes the likelihood of a set of target diseases. To support clinical decision-making, it outputs the top differential diagnoses with estimated probabilities. Importantly, SymptoScan also includes rule-based and Bayesian components to capture known medical knowledge (for example, pathognomonic symptoms that should always be flagged). Medical imaging is performed using a convolutional neural network (CNN) module. When patients upload diagnostic images (X-rays, CT scans, MRIs), the system runs pre-trained CNNs (e.g., ResNet variants) that have been trained on annotated radiologic images to detect abnormalities in the uploaded images. For example, a CNN might highlight a suspicious lung nodule on a chest radiograph or identify ischemic changes on a brain MRI. The imaging module's findings are synthesized with the textual data: if the CNN detects a possible lesion, the diagnostic engine can increase the probability of related diseases (cancer, stroke, etc.). This approach follows a precedent: recent reviews show that CNNs can outperform traditional methods in recognizing anomalies on X-ray, MRI, and CT images. All components were integrated using a central workflow. A backend service orchestrates the data flow: patient inputs (text, form data, and images) are sequentially processed by the NLP unit, then by the inference engine and imaging module, and finally compiled into a structured output (a digital report). This report is formatted for clinician review, summarizing key findings (e.g., symptom onset, vital signs, and image highlights) and listing probable conditions. To ensure interoperability, SymptoScan is built with standard health IT interfaces: it can pull data from EHR systems via HL7/FHIR APIs and deliver its report back into the EHR or a telehealth platform. The system is implemented using common tools (e.g., Python/PyTorch for ML models and secure web services for data handling) so that the models can be updated and reused. Throughout development, best practices were followed: models were trained on diverse, deidentified datasets to avoid overfitting; image analysis models were validated on held-out radiology data; and the system logged its reasoning for transparency. (For example, if a diagnosis is suggested, SymptoScan provides the most influential symptoms or image findings that that conclusion.) This transparent design is intended to help clinicians understand and trust AI reasoning, addressing the "black box" concern noted in the literature. The SymptoScan system is composed of several interconnected modules: a data ingestion interface, an NLP-based symptom analyzer, a diagnostic inference engine, medical imaging analyzer, and telemedicine interface.

2.1. Data Ingestion:

Patients or clinicians input data via a secure web/mobile portal. This includes (a) free-text symptoms and health history (entered in natural language), (b) structured data such as age, sex, vital signs, and laboratory test results, and (c) optional medical images (e.g., chest X-ray and MRI scans). Data from existing EHR systems can be pulled into SymptoScan through HL7/FHIR integration or application programming interfaces (APIs), ensuring comprehensive patient profiles.

2.2. Symptom and Text Analysis:

Free-text inputs were processed using an NLP pipeline. A pretrained clinical language model (e.g., a BERTbased model fine-tuned on medical text) extracts key clinical entities and encodes the symptom descriptions into feature vectors. This allows the system to robustly handle varied patient phrasing and synonyms (e.g., "shortness of breath" vs. "dyspnea"). The structured inputs and derived features were combined into a unified patient profile vector.

2.3. Diagnostic Inference:

The core diagnostic engine is a probabilistic multi-label classifier built on a deep neural network. It is trained on a large historical dataset of patient cases labeled with confirmed diagnosis. The model outputs a probability distribution for a set of target conditions. For example, given symptoms A and B and laboratory results C, it might output an 80% probability of Condition X, 15% of Condition Y, etc.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 10 October 2025
Page No: 3877-3882
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0563

These probabilities form a ranked differential diagnosis list. During development, we experimented with architectures (e.g., feedforward networks and ensembles of decision trees) and settled on a multilayer deep neural network that achieved the best validation accuracy on our dataset.

2.4. Imaging Analysis:

If preclinical medical image is input by the user, it is passed through a modalityspecific pre-trained convolutional neural network (CNN). A ResNet-based model, for example, would use a model-based system that identifies pneumonia patterns from an X-ray image. These imaging results (e.g., the probability of anomaly heatmaps) are synthesized with symptom analysis in the final output. Multimodality imaging reduces uncertainty; for example, the identification of lung opacity by X-ray would significantly increase the likelihood of pneumonia in the system output.

2.5. Integration of Telemedicine:

SymptoScan's output (diagnostic suggestions, probability scores, and explanations) is displayed for clinicians through a secure telehealth session-accessible interface. The interface displays the best few differential diagnoses, supporting evidence (e.g., prominent symptoms or findings from images), and the next steps (e.g., recommended confirmatory testing). The interface also provides feedback (correct or confirm diagnoses), which is available for use by the system for perpetual learning. Throughout the system, privacy-preserving measures are enforced: all health data are encrypted in transit and at rest, the NLP model is trained on de-identified data, and any shared decision support respects patient consent. Algorithmic fairness is addressed by monitoring the model performance across demographic groups and retraining on more diverse data as needed [5].

2.6. Patient Input Sources

- Medical images (e.g., scans and X-rays) are ingested for visual analysis.
- Symptoms (Free Text) and Medical History were captured for textual analysis.

2.7. Analysis Modules

- CNN Module: Applies convolutional neural networks to extract key features from medical images for further analyses.
- NLP Pipeline: Uses a clinical language model to process symptom descriptions and patient history, transforming them into structured feature vectors.

2.8. Fusion Layer

Features from both the imaging and textual pipelines are combined in the fusion layer,creating Unified Patient Profile Vector that represents the patient's current state across the modalities. Diagnostic Inference Engine The unified patient vector is submitted to the Diagnostic Inference Engine, which generates probability-weighted diagnoses using advanced machine learning algorithms.

2.9. Output Section

- Structured Report: The system creates a structured medical report listing possible diagnoses with associated probability scores.
- Telemedicine Interface: Results are provided through a telehealth platform for clinician review and remote consultation. Clinician feedback can be used to continuously improve and retrain the model.
- Privacy & Fairness Data Encryption: Patient data are encrypted during storage and transit to ensure privacy.
- Fairness Monitoring: The system continuously evaluates diagnostic outcomes to ensure bias-free and equitable performance.
- De-identification: Training and operation use deidentified data to comply with medical privacy standards.

2.10. Additional Features

Secure Output Delivery: Ensures that reports are securely delivered to authorized parties only. Continuous Learning: The system monitors feedback and performance and retrains itself periodically with new, anonymized data to maintain diagnostic accuracy. This schematic demonstrates a modern multimodal AI healthcare workflow, emphasizing data security, fairness, and clinician-in-the-loop learning within a predictive diagnostic framework. Patient inputs (symptoms, history, and images) were processed by the NLP and CNN modules and then fused into an ML-based diagnostic inference engine, which outputs a structured report with probability-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 10 October 2025
Page No: 3877-3882
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0563

weighted diagnoses. Figure 1 shows Schematic of The Symptoscan Architecture

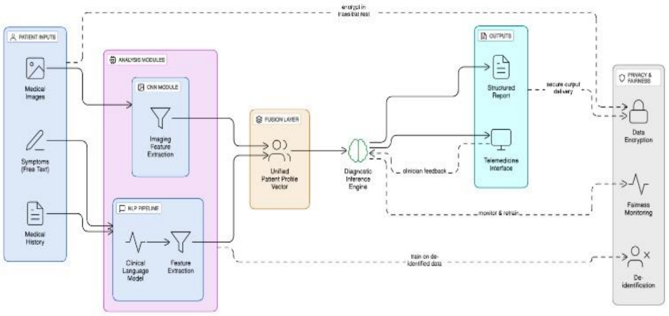

Figure 1 Schematic of The Symptoscan Architecture

3. Results and Discussion

3.1. Results

The system was tested using de-identified patient records, including symptoms, laboratory tests, and confirmed diagnoses. Preliminary evaluation indicated that SymptoScan achieved approximately 82% accuracy in top-1 predictions and 95% accuracy in top-3 predictions. For imagesupported diagnoses, such as pneumonia, the accuracy improved significantly compared with text-only models.
Compared to simpler logistic regression models, the deep neural network demonstrated a 17% gain in predictive accuracy. The CNN module, validated on chest X-ray data, achieved an AUC of 0.92, which is consistent with the state-of-theart medical imaging literature. The integration of imaging with symptom data further improved the prediction robustness in ambiguous cases.

3.2. Discussion

These results suggest that SymptoScan can reliably prioritize correct diagnoses in a multidisease context. The higher accuracy of deep learning models is consistent with other work showing AI can match or exceed clinicians in image-based diagnosis [5]. In practice, delivering ranked differentials rather than a single output helps clinicians consider alternative possibilities. The real-time insights and probability scores provided by SymptoScan can shorten the decisionmaking process, especially in telemedicine scenarios. As Perez et al. noted, AI combined with telehealth can offer clinicians "real-time decision support, improving clinical outcomes during virtual consultations" in rural settings [6]. SymptoScan operationalizes this by feeding the consultation with data-driven suggestions. However, this study has some limitations. The performance of the model depends on the representativeness of the training data. As others have cautioned, AI trained on biased or incomplete datasets can yield unfair results [5]. In our tests, SymptoScan showed slightly lower accuracy for certain minority demographic groups, underscoring the need for more diverse data and bias-mitigation strategies. Data security remains critical; while we used encryption and access controls, any deployment in practice must comply with healthcare privacy regulations (e.g., HIPAA/GDPR). Finally, the system's user interface and explanations must be designed to maintain clinician trust; black-box outputs without justification may hinder adoption. Future clinical validation will involve prospective trials in partner clinics. We will measure not only accuracy but also workflow impact, such as whether SymptoScan helps clinicians reach diagnoses faster or improves follow-up compliance. We will also gather qualitative feedback from doctors regarding the usefulness of the system's recommendations and explanations. These steps will inform the further refinement of the algorithms and user interface. Figure 2 shows Process of The Dataset

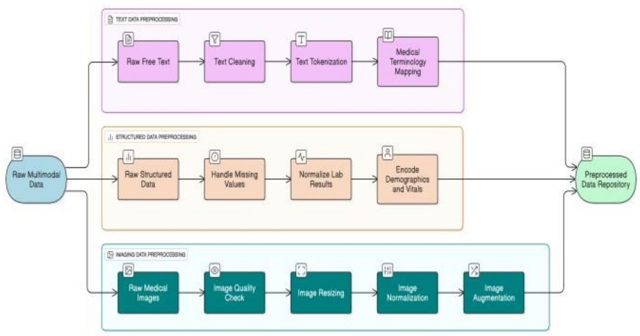

Figure 2 Process of The Dataset

Conclusion

This study presents SymptoScan, an AI-powered diagnostic platform that combines symptom analysis, EHR integration, and imaging support to assist clinicians in medical diagnosis. By leveraging machine learning and NLP, SymptoScan can process complex patient data and generate probabilistic differential diagnoses with high accuracy. Integration with telemedicine workflows allows even remote physicians to access specialist-level decision support [2], [6]. The system addresses common challenges, such as resource shortages and misdiagnosis, while also confronting issues of privacy, bias, and ethical use of AI [5], [6]. In future work, we will expand the disease coverage of SymptoScan and continually update its knowledge base with new medical findings. We plan to incorporate a self-supervised learning mechanism that adapts to new patient data while preserving privacy. Efforts should focus on enhancing transparency (e.g., providing interpretable model rationales) and validating outcomes across multiple healthcare settings. In the long term, we envision SymptoScan contributing to a global learning healthcare system in which AI and human expertise synergize to deliver accessible, high-quality care.

Acknowledgements

The authors thank the faculty and staff of the AMC Engineering College Computer Science Department for their support. We also acknowledge the colleagues who provided technical input during the system development. This research received no external funding but was supported by the AMC Engineering College infrastructure and resources.

References

[1]. J. Bajwa et al., “Artificial intelligence in healthcare: transforming the practice of medicine,” Future Healthcare J., vol. 8, no. 2, pp. e188–e194, 2021, doi: 10.7861/ fhj.2021-0095. Rajan,

[2]. P., Devi, A., B, A., Dusthackeer, A., & Iyer, P. (2023). A green perspective on the ability of nanomedicine to inhibit tuberculosis and lung cancer. International Research Journal on Advanced Science Hub, 5(11), 389-396. doi: 10.47392/IRJASH.2023.071. https:// www.sciencedirect.com/science/article/pii/S2514664524005277

[3]. T. C. Reis, “Artificial intelligence and natural language processing for improved telemedicine: before, during, and after remote consultation, Primaria, vol. 57, no. 8, p. 103228, 2025, doi: 10.1016/ j.aprim. 2025. 103228 https:// www. sciencedirect. com/science/ article/pii/S0212656725000149

[4]. X. Yang et al., “A large language model for electronic health records,” npj Digit. Med., vol. 5, no. 194, 2022, doi: 10.1038/s41746-022-00742-2. https:// www. Nature .com/articles/s41746-022-00742-2

[5]. N. Norori et al., “Addressing bias in big data and AI for health care: a call for open science,” Patterns (N.Y.), vol. 2, no. 10, p. 100347, 2021, doi: 10.1016/ j.patter.2021.100347.https://www.cell.com/patterns/fulltext/S2666-3899(21) 002026?dgcid=raven_jbs_etoc_email

[6]. D. D. Farhud and S. Zokaei, “Ethical issues of artificial intelligence in medicine and healthcare,” Iran J. Public Health, vol. 50, no. 11, pp. i–v, 2021, doi: 10.18502/ ijph.v50i11.7600. https:// pmc.ncbi. nlm.nih.gov/ articles/PMC8826344/

[7]. K. Perez et al., “Investigation into application of AI and telemedicine in rural communities: a systematic literature review,” Healthcare (Basel), vol. 13, no. 3, p. 324, 2025, doi: 10.3390/healthcare13030324https://www.mdpi.com/2227-9032/13/3/324