Context-Aware Assistive Differential Diagnosis System

A project report submitted in partial fulfillment of the requirements
for the award of the degree of

B.Tech in

Computer Science and Artificial Intelligence

By

Avneet Singh Bedi(**2021UCA1815**)

Sneha Gupta (**2021UCA1858**)

Balvinder Singh(**2021UCA1845**)

Under the supervision of

Dr. Rudresh Dwivedi

Computer Science(CS)

DIVISION OF COMPUTER SCIENCE
NETAJI SUBHAS UNIVERSITY OF TECHNOLOGY
DELHI-110078

2025

CERTIFICATE

This is to certify that the project titled **Context-Aware Assistive Differential Diagnosis System** is a bonafide record of the work done by

Avneet Singh Bedi(**2021UCA1815**)

Sneha Gupta (**2021UCA1858**)

Balvinder Singh(**2021UCA1845**)

under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of **Bachelors of Technology** in **Computer Science and Artificial Intelligence** of the **Netaji Subhas University of Technology, DELHI-110078**, during the year 2024-2025.

Their work is genuine and has not been submitted for the award of any other degree to the best of my knowledge.

DATE:

Dr. Rudresh Dwivedi

Assistant Professor

Division of Computer Science

Netaji Subhas University of Technology

DECLARATION

This is to certify that the work which is being hereby presented by us in this project titled "**Context-Aware Assistive Differential Diagnosis System**" in partial fulfilment of the award of the Bachelor of Engineering submitted at the Department of Information Technology, Netaji Subhas Institute of Technology, University of Delhi, New Delhi, is a genuine account of our work carried out during the period from August 2020 to December 2021 under the guidance of Dr. Rudresh Dwivedi, Department of Computer Science, Netaji Subhas University of Technology, New Delhi.

The matter embodied in the project report to the best of our knowledge has not been submitted for the award of any other degree elsewhere.

DATE:

Avneet Singh Bedi	**Sneha Gupta**	**Balvinder Singh**
2021UCA1815	2021UCA1859	2021UCA1845

ACKNOWLEDGEMENT

We would like to take this opportunity to acknowledge the support of **Dr. Rudresh Dwivedi**, our project guide, for helping us and guiding us in the course of this project . **Dr. Manoj Kumar**, the Head of the Department, Department of Computer Science.

Our internal reviewers, **REVIEWER NAME 1** , **REVIEWER NAME 2** , **REVIEWER NAME 3** for their insight and advice provided during the review sessions.

We would also like to thank our individual parents and friends for their constant support.

ABSTRACT

This thesis explores the development of a Context-Aware Assistive Differential Diagnosis System, which leverages advanced machine learning (ML) and natural language processing (NLP) techniques to revolutionize medical data management and diagnosis support. The system converts unstructured medical reports into medically relevant tokens, enabling precise data retrieval and visualization through a text-to-vitals conversion tool that generates health metrics based on user prompts.

For healthcare professionals, the platform introduces a context-sensitive AI-driven differential diagnosis assistant, enhancing diagnostic accuracy by providing real-time insights based on patient history and symptoms. Additionally, a 24/7 AI medical support chatbot allows for continuous patient engagement and automatic database updates.

This project addresses significant challenges in medical data tokenization, ML model integration, and efficient report compression. By combining these innovations, the system aims to drastically improve the speed and accuracy of diagnoses, marking a transformative step towards AI-enhanced, patient-centered healthcare.It represents a significant advancement in AI-driven healthcare, blending state-of-the-art LLMs, NLP, and data visualization technologies to create a smarter, more efficient medical ecosystem.

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE AND ABBREVIATIONS

Abbreviation	Meaning
LLMs	Large Language Models
AI	Artificial Intelligence
ML	Machine Learning
DDx	Differential Diagnosis
EHRs	Electronic Health Records
NLP	Natural Language Processing
BERT	Bidirectional Encoder Representations from Transformers
BM25	Best Matching 25 (Ranking Function)e
ColBERT	Contextualized Late Interaction over BERT
SciSpacy	Scientific Text Processing with SpaCy
MedPaLM	Medical Pre-trained Language Model
PubMedBERT	Biomedical BERT Pre-trained on PubMed Data
BART	Bidirectional and Auto-Regressive Transformers
T5	Text-to-Text Transfer Transformer

Table 1: Nomenclature and Abbreviations

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Motivation

In the evolving healthcare landscape, a pressing challenge remains: the absence of an integrated application that consolidates various medical systems into a cohesive solution. While specialized tools like health-related Large Language Models (LLMs), diagnostic systems, and 24/7 support platforms exist, they often function in isolation, compromising patient care. My goal is to develop a comprehensive medical software solution that harmonizes these elements, facilitating seamless workflows for both patients and healthcare providers.

A critical driver for this initiative is the alarming number of medical cases adversely affected by incomplete information. Misdiagnoses often arise from gaps in patient history, particularly when crucial details are overlooked. Many patients, unaware of the significance of their symptoms, may skip sharing relevant information. For instance, a patient with recent stomach issues may not mention them when presenting new breathing problems, yet this context is vital for accurate diagnosis.

Additionally, there is no central hub for medical reports, making it difficult for doctors to access a comprehensive view of their patients' health. By integrating a 24/7 chat system that logs medical events, doctors can gain a fuller picture of patient conditions.

This tool will serve as a perfect memory assistant, complementing the critical skills of healthcare professionals. By harnessing advanced technologies like AI and machine learning, we can enhance diagnostic accuracy and improve patient engagement, ultimately transforming the healthcare experience for all stakeholders involved.

1.2 Key Challenges

Developing a comprehensive medical software application has presented various challenges, both technical and conceptual. Initially, the high-level system design posed significant hurdles as we sought to integrate multiple functionalities—such as patient records, diagnostic tools, and 24/7 support—into a cohesive platform. Crafting a user interface that balances complexity with usability required careful planning and user feedback, ensuring that it caters to both patients and healthcare providers.

On the technical side, building the frontend using NEXT.js, alongside advanced tools like Redux Toolkit, Tailwind, and TypeScript, demanded a steep learning curve. Integrating these technologies required precise coordination to maintain state management effectively and ensure a responsive user experience. Additionally, implementing authentication through Node.js with JWT and Nodemailer for OTP verification proved challenging, particularly in maintaining security while ensuring a seamless user flow.

Furthermore, handling data within the application adds another layer of complexity. Creating a robust backend capable of managing sensitive medical data while adhering to privacy regulations is paramount. The challenge lies in designing a system that not only stores data efficiently but also allows for real-time updates, particularly from the 24/7 chat interface.

Looking ahead, we anticipate challenges in scaling the system as user demand increases, ensuring data integrity and security during this process. Additionally, integrating machine learning components to enhance diagnostic capabilities will require careful planning and resource allocation to navigate complexities in model training and deployment. Balancing these factors will be essential as we strive to create an innovative, reliable healthcare solution.

1.3 Problems Addressed

Healthcare is a complex field, where accurate diagnosis and timely treatment are crucial for patient outcomes. However, one of the persistent challenges in the medical domain is the fragmentation of critical healthcare services. Currently, there are several tools available for individual aspects of healthcare, such as diagnostics, 24/7 medical support, and patient management, but they often function in isolation. This disjointed approach leaves patients and healthcare providers navigating multiple platforms, often leading to incomplete or inefficient care. The lack of an integrated solution that can simultaneously handle patient records, diagnostic tools, and 24/7 support contributes to delays, missed diagnoses, and suboptimal patient experiences.

Another major issue is the lack of a centralized hub for storing and managing medical reports. While electronic health records (EHRs) exist, they are often fragmented across various healthcare providers, leading to incomplete or inconsistent patient histories. This fragmentation can result in critical medical information being overlooked, especially during emergencies or when patients switch between healthcare providers. The absence of a single system where medical reports, doctor consultations, and daily health updates can be seamlessly stored and accessed by authorized healthcare professionals leads to gaps in care. This is particularly concerning when the lack of comprehensive patient data impedes a physician's ability to make informed, accurate diagnoses.

Moreover, patients often fail to recognize the significance of certain health events, which might seem irrelevant to them but are critical for doctors to consider. For instance, a patient who experienced stomach pain two weeks ago may not connect it to a current issue of shortness of breath. As a result, they may neglect to inform their doctor of the earlier symptom, assuming the two are unrelated. This selective reporting can lead to a partial picture of the patient's health, making diagnosis more challenging and increasing the likelihood of medical errors. Doctors, although highly skilled in diagnostics and treatment, cannot account for missing pieces of information that could otherwise provide a holistic view of the patient's condition. The absence of such data can result in misdiagnosis, delays in treatment, and even poor patient outcomes.

Another significant issue in healthcare is the high rate of diagnostic errors due to incomplete or inaccessible information. A doctor's expertise and intuition are indispensable, but they are limited by the information at hand. Many cases of misdiagnosis are not a result of poor medical judgment but rather a lack of comprehensive patient data. The unavailability of a patient's full medical history, ongoing health concerns, and previous diagnostic results makes it more likely that critical aspects of their condition are overlooked.

Additionally, with the increasing demand for real-time medical support, especially in remote or underserved regions, patients often find it challenging to access continuous, reliable healthcare. Existing telemedicine services may provide consultations but are rarely integrated into the patient's broader health records, leaving doctors without crucial contextual information that could aid in diagnosis and treatment. The absence of a system that continuously tracks and updates a patient's health data as they interact with healthcare professionals over time is a significant gap in current healthcare technology.

Lastly, doctors face their own set of challenges when trying to manage multiple patients simultaneously. The sheer volume of information they need to process—ranging from patient histories to current symptoms—makes it difficult for them to retain every detail. This often leads to reliance on memory, which can be fallible, especially under pressure or when managing a large caseload. While doctors are highly trained to diagnose based on available information, the absence of a reliable system to keep track of all relevant patient data introduces unnecessary risks into the healthcare process.

In sum, the current landscape of healthcare suffers from a lack of integration, fragmented data, incomplete patient reporting, and the unavailability of continuous, real-time support—all of which contribute to diagnostic errors, delayed treatments, and inefficient care. These are the primary problems we aim to address through our project.

1.4 Approach to the Problem

To address the pressing challenges in the healthcare domain—fragmented systems, incomplete data access, missed diagnostic opportunities, and lack of centralized support—we aim to develop an integrated healthcare system that unites various essential components into a cohesive platform. Our approach is grounded in harnessing the power of AI, machine learning, and scalable technologies to build a tool that supports both patients and doctors, enhancing the efficiency and accuracy of medical diagnoses and ongoing healthcare management.

1. **Creating a Unified System for Healthcare Management:** The core idea of our approach is to design a single, comprehensive application that consolidates patient management, diagnostic tools, medical records, and real-time 24/7 support into one cohesive platform. This platform will allow healthcare providers and patients to interact with a unified system rather than a series of disjointed tools, which often operate in isolation. Instead of relying on multiple apps for booking doctor appointments, managing health reports, and seeking medical advice, our system integrates all of these functionalities into a single interface. This approach addresses the lack of interoperability between systems, which has long been a pain point in the medical field. With a centralized system, patients no longer need to manage disparate tools or struggle with lost data, and doctors gain immediate access to a full spectrum of medical information.

   To ensure scalability and accessibility, we have chosen to build the application using Next.js on the frontend and Node.js on the backend, with a MongoDB database for securely storing patient records, medical reports, and health updates. By utilizing modern frameworks like JWT for authentication and Nodemailer for secure OTP generation, we ensure robust and secure access to the system while maintaining user-friendly interfaces.

2. **Centralized Hub for Medical Reports:**

   A major part of our approach revolves around the creation of a centralized hub for medical records. Rather than having patients and doctors rely on different sources to gather information, our system will act as the one-stop repository for all patient medical reports. Medical records, including scanned PDFs of test results and doctor notes, will be stored in an encrypted format and easily accessible to authorized personnel.

   Additionally, our 24/7 healthcare chat will continuously update this database with real-time patient interactions and ongoing medical events. For instance, if a patient has been experiencing headaches for a few days and reports this in the chat, the information is automatically logged into their profile. This ensures that when the doctor reviews the patient's medical history, they have a full picture of both major health events and smaller, seemingly unrelated symptoms that patients may forget to mention during in-person visits.

This comprehensive storage and retrieval system directly addresses one of the key problems in healthcare: fragmented patient histories that lead to diagnostic oversights. By giving doctors a more complete overview of each patient's medical background, they can make more informed and accurate diagnoses.

3. **AI-Assisted Differential Diagnosis and Patient History Summarization:**

   One of the most revolutionary aspects of our project is the development of an AI-based differential diagnosis tool. Using a large language model (LLM) fine-tuned on medical data, we aim to provide doctors with a context-aware assistive tool that helps them consider multiple diagnostic possibilities for a given patient based on their medical history, current symptoms, and test results.

   The LLM will be trained to analyze the comprehensive patient history stored in the database, recognize patterns, and suggest possible diagnoses based on similar cases and symptoms. This tool will not only provide diagnosis suggestions but also explain the reasoning behind its suggestions, helping doctors make decisions with more clarity.

   The differential diagnosis tool will function in conjunction with a patient history summarization model, which will use AI to convert vast amounts of medical data into concise summaries. This helps overcome the challenge of information overload, where doctors are forced to sift through long, disjointed medical records. By summarizing relevant medical history and current conditions, the AI acts as a memory assistant for the doctor, ensuring that critical information is always available at the time of diagnosis.

4. **Improved Patient-Doctor Communication and Continuous Monitoring:**

   A crucial part of our approach focuses on improving communication between patients and healthcare providers. Our 24/7 healthcare support chat is designed to bridge the gap between patients and doctors, ensuring continuous, real-time medical support. Unlike traditional telemedicine systems that only provide consultations, our system integrates this chat into the patient's broader medical profile.

   Often, patients fail to recognize the importance of seemingly minor symptoms or events. For example, a patient might think their stomach ache from two weeks ago is unrelated to their current breathing difficulties. However, this information could be crucial in determining the correct diagnosis. Our system ensures that every interaction between the patient and the chat tool is logged and made available to the doctor, preventing potentially important details from slipping through the cracks.

This continuous monitoring not only enhances the quality of patient care but also allows doctors to gain a real-time understanding of the patient's health, making it easier to track long-term conditions or identify the onset of new symptoms.

Chapter 2

Literature Review

2.1 Introduction

In the realm of healthcare, the integration of Artificial Intelligence (AI) and Machine Learning (ML) has the potential to significantly enhance medical diagnostics and treatment outcomes. This project seeks to address key challenges in the field by developing a unified system combining medical tokenization, 24/7 healthcare support, and assistive differential diagnosis (DDx) tools for doctors. In this summary, we review research across relevant fields to understand the scientific basis behind the proposed system.

2.2 Medical Tokenization

Medical tokenization involves transforming unstructured medical data, such as electronic health records (EHRs), into structured tokens or entities that can be used for analysis, information retrieval, and machine learning applications. The challenge of tokenization lies in the domain-specific nature of medical data, which requires sophisticated techniques for entity recognition, disambiguation, and context-sensitive understanding.

Recent research in this field has seen the emergence of transformer-based models like BioBERT [1] and ClinicalBERT [2], which are specialized for biomedical text. These models are capable of tokenizing complex medical narratives, such as diagnosis notes, and turning them into usable structured data. Additionally, the development of tokenization pipelines using tools like SciSpacy [3] has enabled the efficient parsing of clinical notes, facilitating information retrieval and entity extraction.

2.3 Text-to-Data Point Search Models

Effective search functionality in medical data is essential for enabling healthcare professionals to retrieve relevant information quickly. Text-to-data point search models aim to bridge the gap between free-text medical notes and structured data points (e.g., lab results, medications). Traditional search engines fall short in this context because medical language often includes abbreviations, domain-specific terms, and implicit references to medical conditions.

Research has shown the effectiveness of using models like BM25 [4] and ColBERT [5] for creating high-accuracy search systems in the medical domain. These models rely on embeddings from transformers to capture the semantics of search queries, allowing for more relevant results in response to complex natural language queries.

2.4 24/7 Medical Large Language Models (LLM)

Large Language Models (LLMs) have the capacity to transform healthcare by providing real-time, intelligent support for patients. In particular, a 24/7 medical LLM would provide continuous assistance to patients, helping them manage their health and understand symptoms. Several models, such as MedPaLM [6], have shown promise in this area, enabling patient interactions that are both informative and contextually aware.

LLMs have also been integrated into chatbot systems, like Rasa [7], to provide personalized healthcare advice, thereby filling gaps in patient care during off-hours. One of the challenges here is fine-tuning these models to ensure safety and relevance in a medical setting while also addressing concerns related to patient privacy and data security.

2.5 Tokens to Medical Summarization Models

A key requirement for doctors is to access summarized medical histories quickly. Token-to-text summarization models seek to convert tokenized medical data into readable, concise summaries that help doctors make informed decisions. Transformer models such as BART [8] and T5 [9] have been applied to medical summarization tasks, offering a way to distill complex medical histories into actionable summaries. These models can reduce the time doctors spend sifting through reports, while ensuring that no critical information is missed.

2.6 Context-Aware Assistive Differential Diagnosis (DDx)

Differential diagnosis (DDx) is a critical task in medicine, involving the identification of a disease or condition from a set of symptoms. Context-aware models, such as those based on transformer architectures, provide doctors with AI assistance in this task. For instance, PubMedBERT [10] has been trained specifically on medical data to support diagnosis tasks by offering contextually appropriate suggestions.

Recent research highlights the potential of incorporating knowledge graphs into the process, where medical knowledge is stored in graph structures to provide doctors with context-aware diagnostic suggestions, greatly enhancing the reliability of the system.

2.7 Conclusion

By leveraging cutting-edge research in NLP, LLMs, and tokenization, this project aims to create a comprehensive, AI-powered medical system that enhances both patient and doctor experiences. The ultimate goal is to create a centralized, intelligent system that reduces diagnostic errors and improves patient outcomes.

Bibliography

[1] L. Zheng, X. Liu, Z. Sun, and Y. He, "Improving medical ocr information extraction with integrated bert and layoutxlm models," in *Health Information Processing. Evaluation Track Papers*, B. Tang, Q. Chen, H. Lin, F. Wu, L. Liu, T. Hao, Y. Wang, H. Wang, J. Lei, Z. Li, and H. Zong, Eds. Singapore: Springer Nature Singapore, 2023, pp. 164–173.

[2] K. Huang, J. Altosaar, and R. Ranganath, "Clinicalbert: Modeling clinical notes and predicting hospital readmission," 2020. [Online]. Available: https://arxiv.org/abs/1904.05342

[3] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, "The curious case of neural text degeneration," 2020. [Online]. Available: https://arxiv.org/abs/1904.09751

[4] M. Barczy, F. K. Nedényi, and G. Pap, "On simultaneous limits for aggregation of stationary randomized inar(1) processes with poisson innovations," 2020. [Online]. Available: https://arxiv.org/abs/2001.07127

[5] J. E. Sanches, L. S. Ricco, Y. Marques, W. N. Mizobata, M. de Souza, I. A. Shelykh, and A. C. Seridonio, "Majorana molecules and their spectral fingerprints," *Physical Review B*, vol. 102, no. 7, Aug. 2020. [Online]. Available: http://dx.doi.org/10.1103/PhysRevB.102.075128

[6] A. Demlow and M. Neilan, "A tangential and penalty-free finite element method for the surface stokes problem," 2023. [Online]. Available: https://arxiv.org/abs/2307.01435

[7] M. Patwari, R. Gutjahr, R. Raupach, and A. Maier, *JBFnet - Low Dose CT Denoising by Trainable Joint Bilateral Filtering*. Springer International Publishing, 2020, p. 506–515. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-59713-9$_4$9

[8] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," 2019. [Online]. Available: https://arxiv.org/abs/1910.13461

[9] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," 2023. [Online]. Available: https://arxiv.org/abs/1910.10683

[10] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon, "Domain-specific language model pretraining for biomedical natural language processing," *ACM Transactions on Computing for Healthcare*, vol. 3, no. 1, p. 1–23, Oct. 2021. [Online]. Available: http://dx.doi.org/10.1145/3458754