תמונה שמכילה טקסט, גופן, גרפיקה, לוגו

התיאור נוצר באופן אוטומטי

Software Engineering Department  
Braude College

Capstone Project Phase A – 61998

**A System for diagnosing and providing solutions for growth deficiency in children**

**Group 25-2-D-18**

**Dr. Natali Levi**

**Bar Matzliah**– [Bar.Matzliah@e.braude.ac.il](mailto:Bar.Matzliah@e.braude.ac.il)

**Randa Assadi** – [Randa.Assadi@e.braude.ac.il](mailto:Randa.Assadi@e.braude.ac.il)

**GitHub** - https://github.com/BarBitton/Capstone

Table of Contents

1. Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
2. Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

2.1 Traditional Approaches in Diagnosing Pediatric Growth Disorders…….5

2.2 Electronic Health Record (EHR) Systems……………………………….5

2.3 AI in Medicine: Progress and Applications……………………………...5

2.4 Natural Language Processing (NLP) for Biomedical Literature…………6

2.5 How AI Interprets Medical Diagnoses: Strengths and Limitations………6

2.6 Writing Accurate AI Prompts: With Literature-Based Insights………….7

1. Expected Achievements. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
2. Engineering Process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . …... . . . . 9

4.1 Development Process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . ..10

4.2 Product . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

4.3 System Architecture. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

4.4 User Interface and Experience. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

4.5 Testing and Quality Assurance. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

4.6 Continuous Improvement and Future Development. . . . . . . . . . . . . . . . . 18

1. Verification and Evaluation. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
2. Constraints and Challenges . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .27

6.1 Data Availability and Quality. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .27

6.2 Ethical and Privacy Considerations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

6.3 Model Accuracy and Reliability . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

6.4 Usability and Accessibility . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

6.5 Technical Challenges. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

1. References. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .29
2. Appendix A – User Experience Questionnaire. . . . . . . . . . .. . .. . . . . . . . . . . .31

# Abstract Pediatric growth deficiencies, particularly Failure to Thrive (FTT), are often underdiagnosed due to the lack of accessible, intelligent diagnostic support tools—especially for non-specialist caregivers and clinicians. This project will propose the development of an AI-powered system designed to assist in the early identification and treatment recommendation of FTT in children.

The system will leverage advanced Natural Language Processing (NLP) techniques, including domain-specific transformer models such as BioBERT and PubMedBERT, to analyze unstructured medical literature. By extracting clinically relevant information from journal articles, clinical guidelines, and case studies, the platform will generate diagnostic insights and suggest personalized treatment options.

A user-friendly web interface will be developed to allow doctors, parents, and researchers to submit structured clinical data—such as age, weight, height, and symptom descriptions—via a secure input form. The backend architecture will include modular classification pipelines, dynamic prompt generation, and scalable AI components to process this input and return evidence-based conclusions.

Evaluation will be based on the system’s ability to replicate known medical diagnoses and align its suggestions with peer-reviewed literature. Feedback from pediatricians and parents will guide iterative improvements to both interface and reasoning logic.

This work will address a critical gap in pediatric healthcare by offering a scalable, explainable, and AI-driven diagnostic support tool that transforms complex medical knowledge into accessible, actionable insights—empowering earlier intervention and improving outcomes for children with suspected growth disorders.

# Introduction

Growth deficiency in children, often diagnosed under the term Failure to Thrive (FTT) [1], is a medical condition in which a child's physical growth falls significantly below the average for their age and gender. It can be caused by a variety of factors including nutritional deficiencies, hormonal imbalances, chronic diseases, or genetic conditions. Early detection and proper treatment are critical, as prolonged growth deficiency can lead to developmental delays, weakened immune response, and long-term health complications.

Diagnosing growth disorders typically require monitoring height and weight over time, combined with medical evaluations and lab testing. However, current diagnostic methods are time-consuming, depend heavily on physician experience, and often lack integration with the vast amount of existing medical knowledge. As the volume of medical research continues to grow, it becomes increasingly difficult for clinicians to stay updated with the latest findings relevant to diagnosis and treatment of growth disorders.

Recent advancements in artificial intelligence (AI) [2] and natural language processing (NLP) offer promising tools to bridge this gap. AI models can process and extract insights from large collections of unstructured medical literature, helping medical professionals to make more accurate and informed decisions faster. By analyzing patterns in research data and comparing them with clinical indicators, AI has the potential to support more consistent and personalized diagnoses.

Currently, diagnosing growth deficiencies such as FTT relies primarily on traditional clinical methods. These include routine tracking of a child's height and weight over time, comparison to standardized growth charts, physical examinations, and laboratory tests. In more complex cases, referrals to pediatric endocrinologists may be required. Some healthcare providers utilize electronic health record (EHR)[3] systems with built-in growth tracking tools, but these systems rarely integrate or analyses the latest scientific literature. Moreover, there are few, if any, accessible AI-powered tools that allow parents or general practitioners to automatically analyses up-to-date research and receive tailored diagnostic insights or treatment suggestions. This creates a significant gap in early detection and decision-making, particularly in non-specialist or underserved settings.

This project proposes the development of a system that uses AI to diagnose medical conditions in children, with a primary focus on analyzing medical data to identify health issues. This makes FTT an ideal case for evaluating the system's effectiveness by comparing AI-generated results with established medical conclusions.

The system will accept scientific articles as input, analyses them using advanced prompt engineering techniques, and extract medical knowledge related to diagnosis and treatment from the AI engine. It will then provide users—including healthcare professionals and parents—with diagnostic suggestions and potential therapeutic options based on this analysis. The platform aims to serve as a support tool not only for doctors, researchers, and healthcare institutions, but also for parents who are concerned about their child's growth. By offering easy access to AI-driven insights, the system empowers parents to proactively monitor and understand potential growth-related issues.

It will help reduce diagnostic delays, streamline access to relevant literature, and ultimately improve the quality of care for children with suspected growth deficiencies. By leveraging AI to process up-to-date medical knowledge, this system addresses one of the key challenges in pediatric healthcare: timely and accurate diagnosis based on a broad base of scientific evidence.

Unlike general-purpose AI models such as ChatGPT, our system will be specifically designed for pediatric diagnostic purposes. While ChatGPT provides general health information, it will not be able to process structured clinical input or provide medically grounded diagnostic suggestions. Our system will overcome these limitations by relying on structured data entry, domain-specific prompt engineering, and insights extracted from verified medical literature. This will ensure more reliable, focused, and explainable results that align with clinical reasoning.

The structure of this document is organized as follows: In Chapter 2, we will present a comprehensive literature review, covering the current tools, technologies, and research related to diagnosing growth deficiencies. This chapter will highlight existing gaps and explain why the proposed solution is needed. In Chapter 3, we will define the expected achievements of the project, outlining the planned software system, its features, and the success criteria for evaluating its performance. Chapter 4 will focus on the engineering process, divided into two parts: the first part will describe the methodology, development stages, and rationale behind design decisions, while the second part will discuss the algorithms, models, data structures, and the user interface. In Chapter 5, we will outline the testing plan, including the testing strategy, test cases, assumptions, and the environment required to ensure that the system meets the defined requirements.

# Literature Review

In this chapter, we will discuss the current tools, technologies, and research used in diagnosing growth disorders in children. The focus will include advanced medical methods, AI and machine learning applications, medical databases, and data-driven decision support systems. We will also review recent studies in the field, highlighting the importance of diagnostic accuracy, clinical challenges, and the potential for technological innovation to improve diagnostic and treatment processes.

**2.1** **Traditional Approaches in Diagnosing Pediatric Growth Disorders**

Failure to Thrive (FTT) is a clinical condition commonly diagnosed when a child’s weight-for-age falls below the 5th percentile or there is a significant drop across two major growth percentiles. Standard practices involve regular monitoring of anthropometric measurements—height, weight, and head circumference—compared against growth charts published by the CDC or WHO ,In suspected cases, physicians often order laboratory tests to investigate underlying metabolic, endocrine, gastrointestinal, or genetic causes. Severe or persistent cases are referred to specialists such as pediatric endocrinologists or gastroenterologists

While these practices are established and widely used, they rely heavily on clinician experience and are reactive rather than predictive. Consequently, diagnosis and treatment may be delayed, particularly in primary care settings where access to specialists and advanced tools is limited.[[1]](https://www.nationwidechildrens.org/conditions/failure-to-thrive?utm)

**2.2** **Electronic Health Record (EHR) Systems**

EHR systems like EPIC [[2]](https://www.epicshare.org/share-and-learn/small-babies-grow-stronger-with-improved-nutrition-monitoring) or Cerner offer built-in pediatric growth tracking functionalities. These include plotting growth parameters and issuing alerts for deviations. However, studies show that these tools are rarely utilized to their full potential. They often lack integration with up-to-date medical literature, advanced predictive modeling, or personalized insights tailored to a specific child’s clinical profile[[3]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6815511)[[4]](https://publications.aap.org/pediatrics/article/154/4/e2024068509/199440/)[[5]](https://empeek.com/insights/epic-vs-cerner) EHR alerts may be ignored due to "alert fatigue" or may lack the clinical context needed to support real decision-making.

**2.3** **AI in Medicine: Progress and Applications**

Artificial Intelligence (AI) has already revolutionized several domains in healthcare. In radiology, AI models using convolutional neural networks (CNNs) have achieved performance comparable to expert radiologists in detecting tumors, fractures, and other pathologies [[6]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11047988/) In cardiology, AI systems can analyze ECG data to predict atrial fibrillation or heart failure with high accuracy [[7]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11941653/) Additionally, AI tools are being developed to predict sepsis, detect diabetic retinopathy, and even personalize drug regimens. Despite these successes, pediatric growth disorders remain an underserved domain. AI tools are rarely tailored to pediatric data, and many models are not designed to accommodate the nuances of child development or the ethical and clinical sensitivities required in pediatric care

**2.4** **Natural Language Processing (NLP) for Biomedical Literature**

NLP enables the analysis of large volumes of unstructured biomedical text, such as journal articles, case reports, and clinical guidelines. State-of-the-art models like BioBERT and PubMedBERT, trained on biomedical literature, allow for accurate extraction of clinical terms, relationships, and concepts[[9]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9074854/)[[10]](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01352-2), This is especially useful in continuously updating AI systems with emerging evidence, which can be critical in fields like pediatrics where clinical guidelines evolve over time NLP can also support tools that translate research into actionable insights. For example, a well-designed NLP model can summarize the latest studies on FTT interventions and connect them with patient-specific features in an EHR system, enhancing the evidence-based capabilities of frontline clinicians [[11]](https://www.johnsnowlabs.com/clinical-nlp/).

**2.5** **How AI Interprets Medical Diagnoses: Strengths and Limitations**

AI can support medical diagnosis in many ways, but it also has clear limitations that must be considered—especially in sensitive fields like pediatrics.

### Strengths:

Scalability – AI can quickly analyze large amounts of medical data, much faster than any human.[[12]](https://www.microsoft.com/en-us/research/publication/capabilities-of-gpt-4-on-medical-challenge-problems/)

Pattern Recognition – AI is effective in identifying patterns, especially in medical imaging or predictive models.[[13]](https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000198)

Research Summarization – Large Language Models (LLMs) can summarize a wide range of medical literature and help doctors stay updated.[[14]](https://www.psu.edu/news/information-sciences-and-technology/story/improving-efficiency-reliability-ai-medical-summarization)

### Limitations:

No Clinical Judgment – AI cannot understand emotions, social situations, or make moral decisions like human doctors do.[[15]](https://pmc.ncbi.nlm.nih.gov/articles/PMC8826344/)

Hallucinations – Sometimes, AI produces answers that sound correct but are actually false, especially if the prompt is vague or the condition is rare.

Lack of Explainability – Many AI systems do not show how they reached their conclusions, making it hard to trust or understand their reasoning.

Legal and Ethical Concerns – It’s unclear who is responsible if an AI causes harm, and there are concerns about consent and accountability.[[16]](https://www.nejm.org/doi/full/10.1056/NEJMhle2308901)

In pediatrics, diagnosis often depends on subtle signs in growth, behavior, and family context. AI lacks the human senses and intuition needed to interpret these factors, so its use in this area must always be supervised by qualified healthcare professionals.

**2.6** **Writing Accurate AI Prompts: With Literature-Based Insights**

Recent research and practical guides in the field of prompt engineering, including works by Brown et al. (2020) and Reynolds & McDonell (2021), emphasize the importance of carefully crafted prompts when interacting with artificial intelligence models. A well-written prompt significantly enhances the relevance, precision, and usefulness of the AI's output. Below is an expanded explanation of the key components required for effective prompt construction.[[17]](https://arxiv.org/abs/2005.14165)[[18]](https://www.jmir.org/2023/1/e50638/)

Clarity is the first and most essential element. Prompts must be written in specific and unambiguous language. Vague or open-ended questions often lead to irrelevant or generalized answers

Focus is another crucial factor. Each prompt should address only one main idea or task. Combining multiple questions or topics in a single prompt may confuse the AI model and reduce the quality of the response. Breaking complex inquiries into smaller, separate prompts generally leads to better results.

Context should be included whenever relevant. Providing background information about the intended audience, the purpose of the task, or the specific application of the answer can significantly improve the accuracy and tone of the AI's response.

Output format should also be clearly stated in the prompt. This includes specifying whether the user prefers a paragraph, a list, a summary, a table, or any other structure. Clear instructions about formatting guide the AI in structuring the response appropriately and make the output easier to use. [[19]](https://www.ibm.com/think/topics/prompt-engineering-guide)

**Current Gaps and Unmet Needs**

This literature review reveals a significant gap in the application of AI for diagnosing pediatric growth disorders, particularly Failure to Thrive (FTT). While AI technologies are rapidly advancing and are widely adopted in areas such as oncology and cardiology, their integration into pediatric diagnostics remains limited and underdeveloped. Existing Electronic Health Record (EHR) systems lack the ability to deliver real-time, research-informed insights, and they offer minimal support for non-specialist practitioners who may be the first to observe early signs of growth failure.

Moreover, there is a notable lack of accessible digital tools for parents and general practitioners to interpret recent medical research or receive personalized, evidence-based guidance. This shortfall delays early identification and intervention, which are critical in managing growth disorders effectively.

The integration of AI systems—particularly those powered by Natural Language Processing (NLP) and supported through robust prompt engineering—presents a promising opportunity to bridge this gap. Such systems can synthesize up-to-date medical literature, generate clinically relevant insights, and provide tailored recommendations that support timely and accurate diagnosis.[[20]](https://mindmetrix.com/blog/why-ai-shouldnt-replace-clinical-judgement-in-mental-health-diagnosis)

Therefore, this review underscores the need for an innovative AI-driven platform, as proposed in our project, that translates complex medical knowledge into practical, accessible tools for clinicians and caregivers. By combining NLP, real-time data analysis, and precise prompt design, the proposed solution has the potential to significantly improve diagnostic accuracy and health outcomes for children experiencing growth challenges.

# 3. Expected achievements

The primary goal of this project is to develop an AI-based system that assists in diagnosing and providing solutions for any medical issue and mostly growth deficiency in children, with a specific focus on Failure to Thrive (FTT) as an initial use case

### Planned Deliverables:

**AI-Powered Diagnostic Engine**  
 A core algorithm that utilizes large language models (LLMs) to process and analyze unstructured medical texts (e.g., journal articles, clinical guidelines). The engine will identify patterns, symptoms, and indicators related to growth deficiency in children.

**Recommendation Module**  
An intelligent module that suggests potential treatments and medical interventions based on the extracted data, tailored to the diagnosis (e.g., FTT). The recommendations will be aligned with up-to-date clinical research.

**Interactive User Interface**  
As part of this product, an interactive user interface will be developed to serve doctors, parents, and researchers. The interface will allow users to enter relevant clinical information via structured form fields, including demographic details (e.g., age, weight, height) and a free-text description of the child’s condition. The design will prioritize simplicity, accessibility, and clarity. To illustrate the user experience and interface logic, this section will include mockup images of the main screens in the system, such as the home screen, document upload screen, and diagnostic results display. These visuals are intended to demonstrate the user flow and the usability principles guiding the interface design.

**FTT Case Study Implementation**  
 A fully functional demonstration of the system's capabilities using FTT as a detailed case study. This will include a literature set, and evaluation of the diagnostic and recommendation output.

The learning process of the AI model will be based on a combination of three data sources:

Peer-reviewed medical articles for extracting clinical rules and symptom patterns.

A verified dataset of 100 real pediatric cases.

An expanded set of 600–700 fake data generated using structured variations.

This hybrid approach will enable the system to learn from both expert knowledge and simulated patient diversity, ensuring higher accuracy and wider applicability in future use.

### System Functionality Overview:

Input: Structured clinical data provided through a web form, including free-text symptom descriptions.

Processing: NLP techniques extract relevant features and classify them into diagnostic and treatment categories.

Output: Diagnostic summary, confidence level, and potential medical solutions.

To ensure reliability and medical accuracy, both the real pediatric case dataset and the medical literature used in training and evaluation were provided directly by our supervisor. This guarantees that the system is built upon clinically relevant and ethically sourced materials, reinforcing the trustworthiness of its diagnostic and recommendation outputs.

### Success Criteria:

The success of the project will be evaluated based on the system’s ability to reproduce diagnostic and treatment suggestions that align with established medical literature. The goal is for the system to reach the same conclusions and recommendations that would be expected from reviewing trusted sources on growth deficiencies such as FTT.

The main success indicators are:

The system identifies key indicators of FTT that match those presented in pre-reviewed medical articles.  
The treatment recommendations generated by the system correspond to those found in validated clinical sources.

A functional app interface enables users to enter medical concern and receive diagnostic and treatment suggestions in a short, practical response time.

The system demonstrates consistency in producing expected results across a sample set of articles selected in advance.

Feedback from relevant stakeholders—such as pediatricians, researchers, or parents—confirms that the system's outputs are useful, clear, and aligned with current medical understanding.

# 4. Engineering Process

**:4.1Part 1:** **Development Process**

**4.1.1** **Overview of the Development Approach**

The engineering process in this project followed the Agile methodology, enabling iterative experimentation and adaptation throughout the semester. Each sprint was focused on a key milestone—ranging from literature collection and data preprocessing to model selection, prompt tuning, and validation. This flexible approach allowed us to continuously integrate feedback and gradually shape a clinically relevant and technically robust diagnostic prototype.

**4.1.2 Stages of Development**

Stage 1: Research and Requirement Analysis  
We began by conducting a literature review focused on paediatric growth disorders—specifically, Failure to Thrive (FTT). We examined clinical definitions, diagnostic limitations, and how existing systems such as EPIC and Cerner approach growth assessment. This process clarified a major gap: current tools rarely leverage up-to-date research or intelligent reasoning, especially for non-specialists. Based on this analysis, we defined our system’s key requirements, including real-time diagnosis suggestions, explainability, and structured + free-text input support.

Stage 2: Data Collection and Preparation  
We received over 30 peer-reviewed medical articles and 100 real paediatric diagnostic cases through academic supervision. The articles were sourced from PubMed and PMC and included clinical case studies, systematic reviews, and treatment guidelines. All medical texts underwent a preprocessing pipeline—using PDFPlumber, Tesseract OCR, and SpaCy—to remove metadata, tables, and footnotes, standardize the structure, and tokenize relevant medical entities. The clinical cases were manually structured into fields such as age, weight, symptom description, and validated diagnosis.

Stage 3: Prompt Design and Refinement  
We conducted a systematic process to design and evaluate prompts that would elicit accurate and context-aware diagnostic responses from language models. Initial attempts used broad prompts such as:

“Is there any indication of growth issues in this child’s data?”

However, these led to vague and overly cautious outputs that lacked clinical precision.

Through multiple iterations, we developed prompts that included structured data and focused clinical questions. One effective prompt example was:

“Given the following clinical information for a 2-year-old child: weight = 8.5kg (below 5th percentile), chronic diarrhea, and low appetite — is this consistent with Failure to Thrive? Please explain the reasoning and suggest treatments.”

We tested such prompts using both real and synthetic cases and compared the model’s outputs with known diagnoses. We also experimented with different prompt phrasings, levels of detail, and entity order, ultimately identifying optimal patterns for accuracy and clarity. Throughout this process, the best results were achieved by combining GPT-4 (for reasoning and explanation) with BioBERT (for clinical concept extraction).

Stage 4: Model Evaluation with Real-World Cases  
To validate the diagnostic reasoning of our system, we selected five real paediatric cases from our dataset to represent a broad clinical spectrum. Each case included structured inputs (e.g., age, symptoms, weight percentile) and a physician-validated diagnosis. We input the data into our system and recorded the AI-generated diagnosis, treatment message, and diagnostic explanation.

The five evaluated cases included:

One child with no FTT (healthy)

One with mild FTT

One with moderate FTT

One with severe FTT

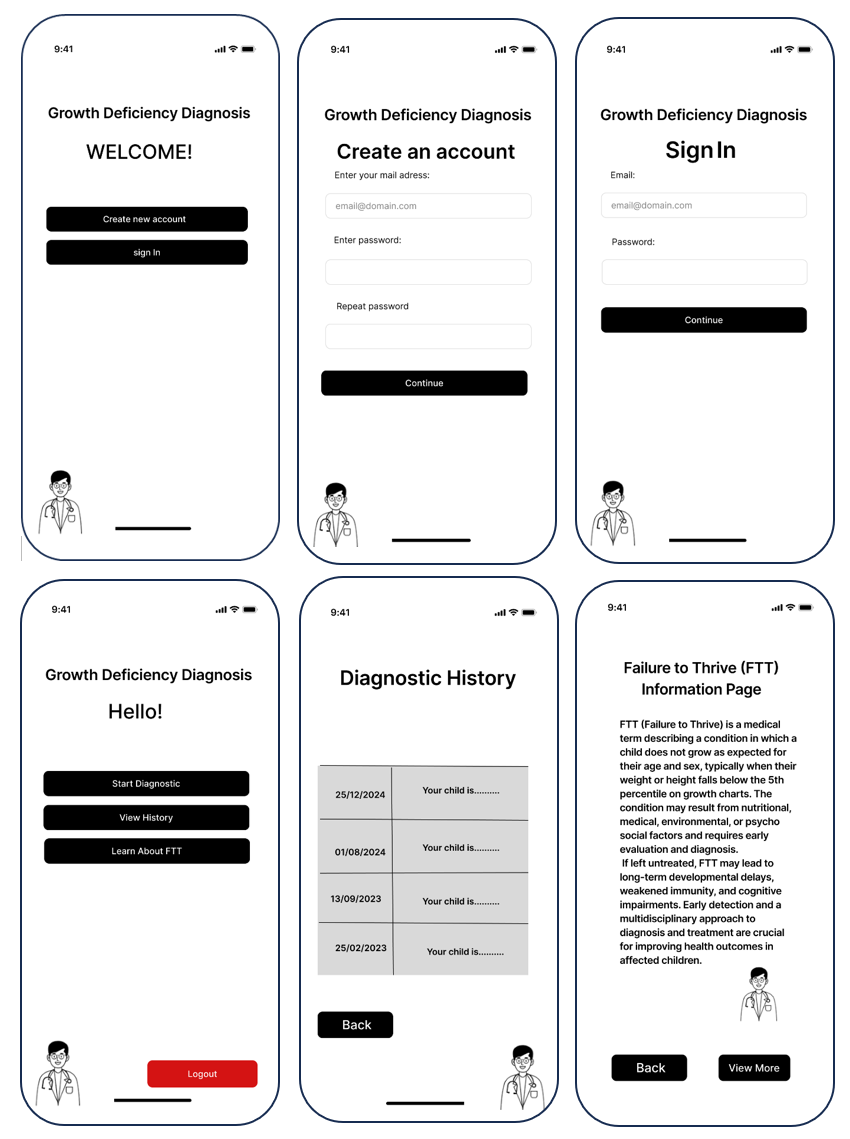
One with a misclassified manual diagnosis

The following table summarizes the comparative analysis:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case ID** | **FTT Severity** | **Manual Diagnosis** | **Physician Diagnosis** | **AI Output** | **AI Message** | **Agreement** |
| 1 | Severe | Severe FTT | Severe FTT | Severe FTT | “Severe growth failure detected. Immediate referral required.” | Yes |
| 2 | Moderate | Moderate FTT | Moderate FTT | Moderate FTT | “Signs of moderate growth concern. Evaluation recommended.” | Yes |
| 3 | Mild | Mild FTT | Mild FTT | Mild FTT | “Mild deviation noted. Monitor growth and nutritional intake.” | Yes |
| 4 | None | Healthy | Healthy | Healthy | “No abnormality detected. Growth within normal limits.” | Yes |
| 5 | Mild | Moderate FTT | Mild FTT | Mild FTT | “Slight growth concern. Monitoring advised.” | Yes (with correction) |

In all five cases, the model matched the physician’s professional diagnosis. In the fifth case, the manual diagnosis incorrectly identified the case as “moderate FTT,” whereas both the model and the clinical expert assessed it as “mild FTT.” The model also correctly emphasized that monitoring was sufficient rather than referral—highlighting its ability to avoid overdiagnosis.

This case-based evaluation demonstrated that the AI system could replicate expert reasoning, distinguish severity levels, and even outperform manual entries when they lacked precision.

**4.2. Part 2: Product**

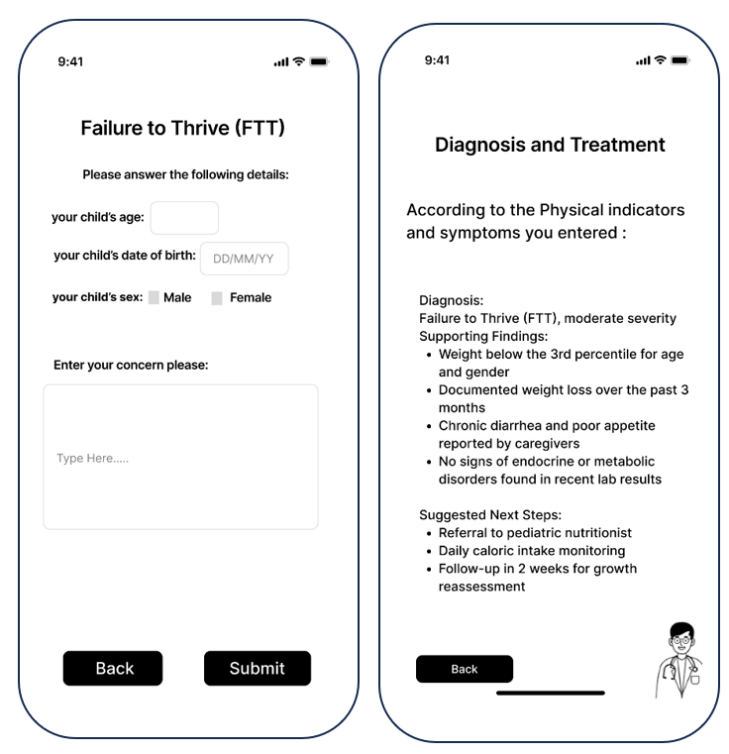


Figure 1: App Screens

**4.2.1 Requirements**

The system is a web-based AI platform designed to aid in the early identification of paediatric growth deficiencies, focusing on Failure to Thrive (FTT). It combines advanced Natural Language Processing (NLP), prompt engineering, and domain-specific language models to analyse clinical descriptions entered by users and suggest possible diagnoses and interventions. The platform is tailored for both healthcare professionals and concerned parents.

**4.2.2 Functional Requirements**

1. The system shall allow users to input clinical data manually, including fields such as age, gender, weight, height, growth history, and free-text description of symptoms or concerns.
2. The system shall analyse the input using pre-defined and optimized prompts designed for medical inference.
3. The system shall extract relevant clinical indicators (e.g., poor weight gain, developmental delays) from free-text using NLP techniques.
4. The system shall match detected indicators with standardized diagnostic criteria for FTT based on known clinical guidelines (e.g., CDC, WHO).
5. The system shall classify the case into one of several diagnostic categories: no concern, mild FTT, moderate FTT, severe FTT, or refer to physician.
6. The system shall provide an explanation of the classification based on the input and highlight which clinical indicators were most relevant to the decision.
7. The system shall suggest evidence-based interventions when applicable, such as nutrition plans or specialist referral.
8. The system shall include a built-in knowledge base powered by uploaded medical literature to enrich the AI’s contextual understanding.
9. The system shall provide the user with a downloadable report summarizing the findings and suggestions.
10. The system shall include an admin dashboard for viewing system logs, reviewing flagged diagnoses, and managing article inputs used for model enrichment.

**4.2.3 Non-Functional Requirements**

1. The system shall provide a simple, responsive, and mobile-friendly user interface accessible via modern browsers.
2. The system shall ensure all user data is anonymized before processing and will not retain any personally identifiable information.
3. The system shall support multiple languages in both input and output (initially Hebrew, Arabic, and English).
4. The backend shall be modular and scalable, using containerized services (Docker) and cloud infrastructure.
5. The system shall be optimized for near real-time processing, with average diagnostic time under 10 seconds.
6. The system shall be secure, GDPR-compliant, and incorporate encryption for all data in transit.
7. The NLP engine shall support periodic updates, including fine-tuning on new pediatric datasets and retraining on verified case outputs.
8. The system shall include fallback messages or referral prompts in cases of uncertain or inconclusive analysis.

**4.3** **System Architecture**

The architecture is designed around modular AI pipelines, secure data handling, and extensibility for integration with future tools like EHRs or real-time clinical dashboards.

The system consists of the following core components:

User & Web Application: The end user (either a parent or physician) interacts with the system through a web-based interface on a mobile device. The Web App collects structured input such as age, weight, height, and symptom descriptions. It provides a user-friendly interface with guided form entry to ensure data quality.

Server: The server handles requests from the Web App, performs data validation, and communicates with backend services. It is responsible for routing the data to the AI model and saving the diagnosis results to the database.

Database (DB): All data submitted by users—including clinical inputs and diagnostic outputs—are stored in a secure database. This data is essential for both historical tracking and future model improvements.

AI Model: The core engine responsible for producing diagnostic outputs. The model analyzes the structured clinical data using NLP and prompt engineering to classify growth deficiency conditions and suggest potential treatments. This model is pre-trained and continuously optimized.

ML Training Engine: A dedicated component that handles model training. It receives two types of input:

Medical Knowledge: Peer-reviewed articles and clinical literature used to derive diagnostic patterns and treatment logic.

Structured + Synthetic Data: A combined dataset composed of 100 real pediatric cases and 600–700 artificially generated cases. This data is labeled and diversified to enable generalized model learning.

The ML Training Engine outputs a trained AI model that is deployed into the live inference pipeline. This separation between training and diagnosis ensures flexibility and future scalability.

This architecture supports a full AI lifecycle—training, deployment, inference, and retraining—while ensuring a smooth experience for end users.

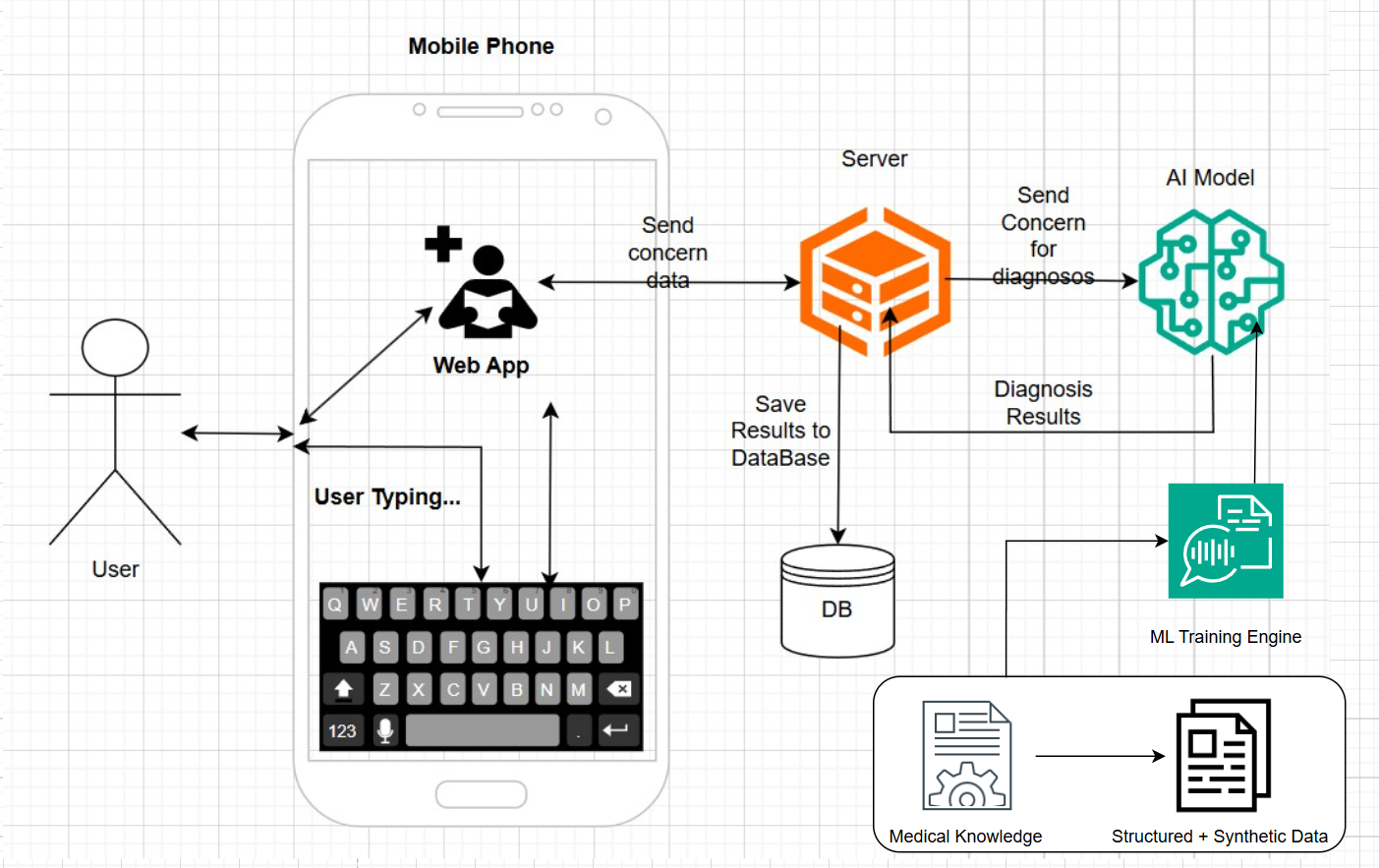


Figure 2: System Architecture

**4.3.2 Frontend Interface**

The user-facing frontend is developed in React.js and optimized for both clinicians and non-expert users. Key components:

User Input Form: Collects structured fields (age, sex, height, weight) and a free-text box for symptom description.

Analysis Status: Shows progress bar and messages during the analysis process.

Result Dashboard: Displays likely diagnosis with a confidence score and suggested interventions.

Explanation Panel: Explains the rationale for the diagnosis in both technical and simplified terms.

Download & Export: Allows PDF download of results and optionally saves reports for clinicians.

Admin Interface: Restricted-access area for managing datasets, reviewing flagged cases, and monitoring system performance.

**4.3.4 Backend Server and API**

Built in Python with Flask, the backend acts as a bridge between the UI and the AI engine:

Routing Layer: Handles form submission, triggers processing pipelines, and formats responses.

Prompt Generator: Dynamically builds GPT-compatible prompts using clinical templates and extracted data.

Security and Anonymization: Ensures all user input is sanitized and anonymized before processing.

Async Task Queue: Uses Celery and Redis for efficient load balancing and processing of concurrent requests.

**4.3.5 NLP & AI Engine**

This is the intelligence core, consisting of several specialized components:

GPT-4 API: Processes clinical descriptions, interprets symptom narratives, and generates structured output using prompt engineering.

BioBERT: Identifies biomedical entities such as symptoms, nutritional deficiencies, and syndromes.

Sentence-BERT: Matches user-provided terms to known medical codes (ICD-10, SNOMED) using semantic similarity.

Scoring Mechanism: Assigns each diagnosis a confidence score (Softmax-based), relevance score (term proximity), and novelty score (rare indicators).

Prompt Tuner: Continuously adjusts prompts based on performance feedback and error patterns.

**4.3.6 Database Layer**

MongoDB: Used for storing session data, extracted diagnosis, system logs, and prompt history (but not user PII).

Redis: Used as an in-memory store for recent analysis results, cached prompts, and asynchronous task management.

**4.3.7 Deployment & DevOps**

Docker: All components are containerized for portability.

GitHub Actions: Automated deployment pipeline runs tests and pushes updates to staging/production.

Scalability: Designed to support future Kubernetes deployment for horizontal scaling.

Integration Ready: The system can integrate with external APIs (e.g., WHO guidelines, PubMed article databases) and hospital EHRs.

**4.4** **User Interface and Experience**

The user experience (UX) was developed iteratively with feedback from paediatricians and parents to ensure accessibility, trust, and simplicity:

Clean UI: Minimal text, intuitive layout, and step-by-step flow.

Structured Input: Separate boxes for demographic data and symptoms ensure clarity.

Interactive Feedback: Color-coded tags for severity, icons for diagnosis categories, and explanatory tooltips.

Accessible Design: Multi-language interface, mobile-first design, and visual accessibility (e.g., color contrast).

Export Functionality: Reports include diagnosis summary, clinical basis, and suggested next steps.

Role Switching: The interface adapts depending on whether the user is a parent or a healthcare provider.

**4.5 Testing and Quality Assurance**

The system is validated through multiple layers of testing and expert feedback:

Unit Tests: Cover all internal modules like tokenizer, NER, prompt builder, and classifier.

Integration Tests: Simulate real flows from user input to output generation.

Domain Expert Review: Paediatricians reviewed system outputs against actual cases to score clinical accuracy.

Usability Testing: Non-technical users (parents) used the interface to test understandability and satisfaction.

Load Testing: Simulates multiple users and long text inputs to identify bottlenecks.

**Accuracy Metrics:**

Precision/Recall: For entity extraction and diagnosis labels.

F1 Score: Overall diagnostic performance.

Agreement Rate: Measured against human clinician assessments.

**4.6** **Continuous Improvement and Future Development**

The platform is designed to evolve with feedback and emerging medical knowledge:

Prompt Optimization: Prompts are continuously tested and adjusted for clarity and performance.

Model Re-training: Real anonymized cases are used to fine-tune models and update classification rules.

User Feedback Loop: Users can rate responses and submit feedback for review.

**Feature Roadmap:**

Real-time symptom checker (chatbot).

Integration with national health systems.

Mobile app version for broader access.

Support for additional paediatric conditions (e.g., obesity, hormonal growth delays).

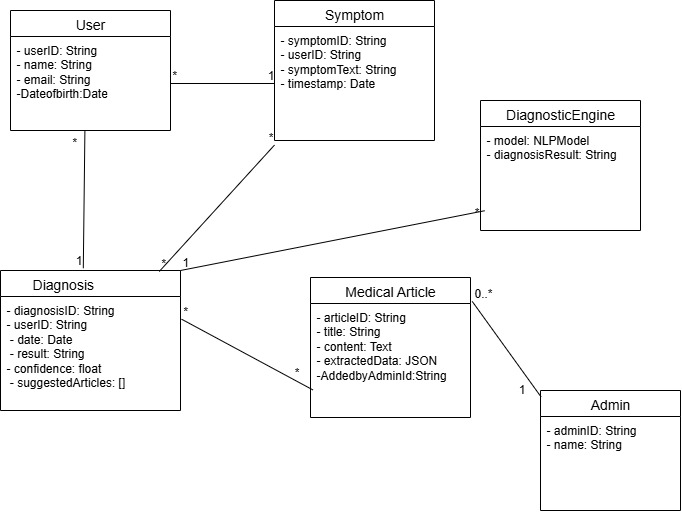


Figure 3:Class Diagram

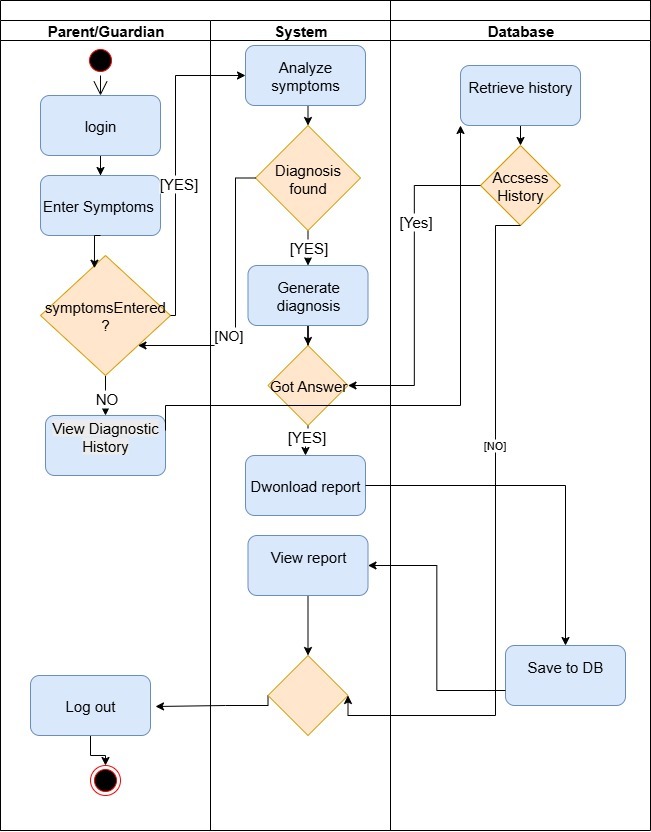


Figure 4: Activity Diagram

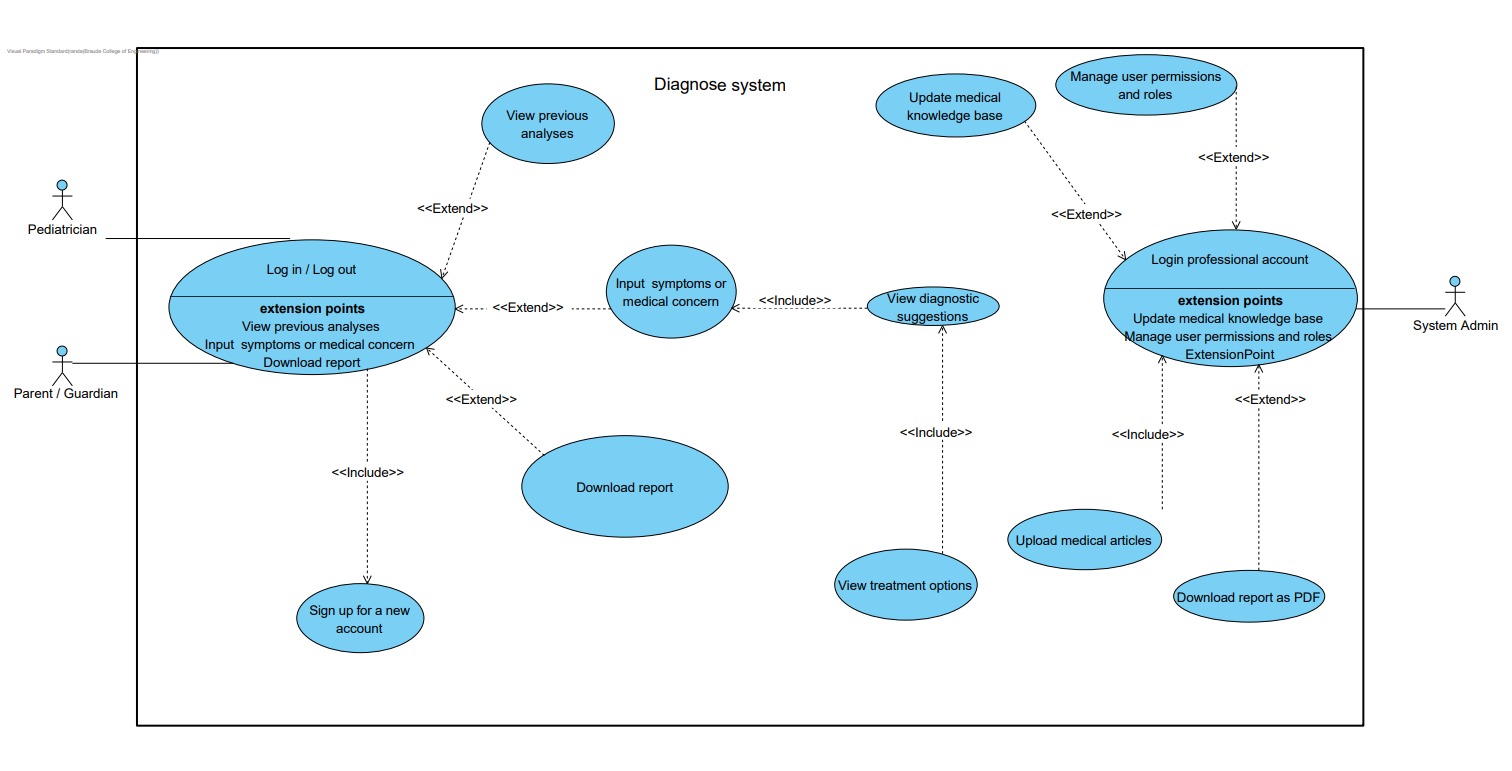


Figure 5: Use Case Diagram

# Verification and Evaluation

The testing phase of our project will play a crucial role in verifying the functionality, accuracy, and usability of the AI-based diagnostic system. Our objective will be to ensure that each module—from text input and analysis to the user interface—performs as intended and delivers value to end users including doctors, researchers, and parents.

**Verification and Evaluation Strategy**

To establish baseline trust in the model's outputs, we will first describe the verified data sources used in model validation.

**Model Reliability and Trustworthiness**

To ensure that the AI model produces medically accurate and clinically trustworthy outputs, we will implement a multi-layered evaluation strategy based on two validated sources:  
(1) real-world podiatric cases and  
(2) peer-reviewed medical literature.  
Both sources were provided directly by our supervisor, ensuring the relevance, credibility, and ethical compliance of the evaluation process.

1. Ground-Truth Validation with 100 Diagnosed Cases

We will utilize a dataset of 100 real paediatric cases diagnosed by clinical professionals, as provided by our supervisor. Each case contains structured clinical input along with an expert-validated diagnosis. These cases will serve as a benchmark for verifying whether the AI model can replicate expert decision-making. For each case, we will:

Input the child’s clinical data into the system

Record the model’s diagnostic output and treatment suggestions

Compare the AI’s results to the physician’s diagnosis

Calculate agreement metrics such as accuracy, precision, recall, and Cohen’s Kappa score

2. Literature-Based Cross-Validation

A collection of over 30 peer-reviewed articles—including case studies, clinical guidelines, and review papers—was also provided by our supervisor. These articles will be processed by the system to extract diagnostic criteria and treatment recommendations using prompt engineering. The outputs will be reviewed to verify alignment with current clinical standards, as reflected in the literature.

3. Iterative Testing with Synthetic Data

To further strengthen model generalization, we will generate an extended dataset of 600–700 synthetic paediatric profiles, based on the patterns and structures observed in the original 100 cases. This will allow us to evaluate model performance across a broader clinical spectrum and diverse input combinations.

This structured and controlled validation approach ensures that the model is evaluated using credible, supervisor-approved data, and reflects realistic diagnostic challenges encountered in paediatric care.

The verification and evaluation of the system will be conducted across five core dimensions:

Diagnostic Accuracy:

Once the system is trained, we will apply a detailed evaluation framework to test performance. The system will be evaluated on its ability to correctly extract diagnoses and treatment suggestions from scientific medical texts. This will be tested using predefined sets of structured clinical data inputs based on validated pediatric cases, then comparing its output to known, expert-reviewed conclusions.

We will measure:

Whether relevant diagnoses are detected in medically appropriate contexts.

To assess the quality and correctness of AI-generated treatment suggestions, we will apply the following metrics:

**Medical Precision:** Experts will compare suggested treatments to clinical guidelines and classify them as Correct / Acceptable / Incorrect.

**Expert Rating:** Pediatricians will rate suggestions on a 1–5 Likert scale for clarity, clinical relevance, and evidence basis.

**Agreement with Source:** For known-case articles, we will check if the AI output matches the article’s recommended treatment.

**Cohen’s Kappa:** Used to measure agreement between multiple reviewers.

**Statistical Comparison:** T-tests or Wilcoxon tests will evaluate performance across different prompt versions or model setups.

These methods will ensure a reliable and medically sound assessment of the system’s diagnostic output.

We will test the system using medical texts with ambiguous or complex phrasing.  
Experts will rate the AI's output for clarity and accuracy (1–5 scale).  
We’ll check for consistent responses across similar inputs, log misunderstandings, and analyze error types.  
This will help assess the model’s robustness in real-world medical scenarios.

**2. Performance and Speed**

We will evaluate the system’s responsiveness in processing inputs and generating results. The complete process from submitting clinical form inputs to delivering insights will be timed to ensure that it is fast enough for practical use.

**Key metrics:**

End-to-end execution time per article: This will be measured from form submission to response time and final output display stages within the application. Multiple test runs will be performed using various article lengths to calculate the average processing time.

Backend response time under concurrent load: Load testing tools such as Apache JMeter or Locust will be used to simulate multiple simultaneous users (e.g., 10, 50, 100). While the actual number of real users may be small during testing, these simulations will emulate future production scenarios and help identify performance bottlenecks.  
We will simulate large and detailed input via extended symptom descriptions to simulate heavy input load.  
Using system monitoring tools (e.g., htop, Docker stats, or cloud dashboards), we will track CPU and memory usage, check for timeouts, and look for crash events.  
All system logs will be reviewed to identify bottlenecks such as slow model inference or data transfer issues.  
This will help ensure that the platform remains stable even under demanding conditions..

**3. Usability and User Experience [Appendix A]**

Usability testing will focus on how intuitively the platform can be used by healthcare professionals and non-technical users (especially parents).

We will assess:

Ease of entering structured data and clinical descriptions or entering text.

Clarity of output such as diagnostic suggestions, treatment options.

Accessibility and navigation of the user interface.

We will conduct pilot testing sessions with a small group of paediatricians and parents, using their feedback to guide interface improvements and information presentation.

**4. Model Generalization and Reliability**

We will evaluate the model’s ability to generalize to new types of texts and medical cases not seen during prompt tuning.

We will:

Test the system on a wide variety of article formats, including case reports, reviews, and clinical guidelines.

Simulate incomplete, noisy, or domain-divergent input to evaluate robustness.

Periodically test output quality after future fine-tuning phases or prompt adjustments.

**Assumptions and Constraints**

The input texts will be in English and follow standard academic or clinical writing.

The system is not expected to extract information from non-medical or poorly formatted documents.

Due to limited time and data access, the initial testing will be conducted using a relatively small curated set of articles.

The system assumes that users provide articles relevant to paediatric growth disorders.

**Testing Environment**

Frontend Testing Tools: Manual and automated UI tests.

Backend and API Testing: Postman for endpoint validation; JMeter or Locust for load testing.

Model Evaluation: Custom Python scripts for content accuracy comparison; prompt outputs logged and reviewed.

Environment: Local development server with simulated production conditions. Testing will later be extended to a deployed environment.

Devices: The platform will be tested on desktop and tablet devices to ensure interface scalability.

**Functional Test Cases**

|  |  |  |  |
| --- | --- | --- | --- |
| Test # | Module | Tested Function | Steps (What to press/do) |
| 1 | Web Application UI | Text Input Interface | Open the app, fill in the form fields (age, gender, etc.), type symptoms, and press 'Submit'. |
| 2 | Web Application UI | Diagnostic Feedback Display | After submission, check that diagnostic results are shown. |
| 3 | Web Application UI | Invalid Input Handling | Enter random/invalid text and verify system response. |
| 4 | Web Application UI | User Registration | Click 'Sign Up', fill in required fields, and press 'Submit'. |
| 5 | Web Application UI | Login and Logout Functionality | Click 'Log In', enter credentials, and click 'Login'. Then press 'Logout'. |
| 6 | Web Application UI | Educational Resources Navigation | Navigate to the 'Learn More' or 'Resources' tab and verify page loads. |
| 7 | Prompt Engine | Diagnosis Extraction Accuracy | Enter data from a case with known diagnosis and check for correct classification. |
| 8 | Prompt Engine | Treatment Recommendation Quality | Analyse output and assess medical relevance of suggestions. |
| 9 | Prompt Engine | Generalization to New Conditions | Enter edge-case or rare symptoms and verify model response. |
| 10 | Backend Server | API Response Time | Time from form submission to AI result. |
| 11 | Backend Server | Load Handling | Simulate 50–100 concurrent users inputting text and measure performance. |
| 12 | Database / Storage | Data Integrity | Submit multiple cases and verify saved results are consistent. |
| 13 | Database / Storage | Secure Data Handling | Ensure no personal data is stored or retrievable. |
| 14 | Model Evaluation | Consistency of Output | Input same data repeatedly and confirm consistent output. |
| 15 | Full System Integration | End-to-End Flow | Submit → Analyse → View Output → View History (ensure seamless experience). |

**System Acceptance Tests:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test # | Scenario | Steps (What to press/do) | Input Data | Expected Output |
| 1 | Submit valid clinical text | Fill form fields → Click 'Submit' | Valid clinical text on FTT | Diagnosis and treatment displayed; success message shown |
| 2 | Submit empty text | Leave all fields empty → Click 'Submit' | Empty form | Error: 'Input is required.' |
| 3 | Submit irrelevant text | Enter non-medical random text → Submit | Random unrelated text | Warning: 'No medical content detected.' |
| 4 | Malformed input | Enter corrupted or partial text → Submit | #@% hormone ?? | Error: 'Unable to analyze input. Please revise and try again.' |
| 5 | Known case analysis | Enter clinical case with FTT signs → Submit | Typical FTT data | Correct diagnosis and treatments displayed |
| 6 | New user registration | Click 'Sign Up' → Fill form → Submit | User details | Message: 'Account created successfully.' |
| 7 | Log in and out | Log in with valid credentials → Logout | Email and password | Redirect to dashboard; session closed on logout |
| 8 | History review | Log in → Go to 'History' | None | Previous cases listed and accessible |
| 9 | Browse educational content | Click 'Resources' tab | None | Content loads on pediatric growth issues |
| 10 | Re-analyze same input | Submit same form multiple times | Same valid data | Same diagnosis and suggestions shown each time |
| 11 | Concurrent usage | Simulate multiple users submitting at once | 50+ simultaneous users | No crash; responses within 10 seconds |
| 12 | Cancel submission | Start submission → Click 'Cancel' | Any input | Submission aborted; return to input screen |
| 13 | Long input | Submit a long clinical description | >1000 words | Processed within acceptable time |
| 14 | Logout during analysis | Start processing → Logout | Any valid data | Analysis canceled; secure session end |

This comprehensive testing plan will ensure that the system is reliable, accurate, and user-friendly. Through careful verification and user centred evaluation, we aim to deliver a trustworthy diagnostic support tool that is ready for real-world deployment.

# 6. Constraints and Challenges

**6.1 Data Availability and Quality**One of the primary constraints we anticipate is the availability of high-quality, structured medical data. Many medical articles are either locked behind paywalls or formatted inconsistently. To address this, we will collaborate with academic institutions to gain access to trusted medical databases and apply robust text preprocessing techniques to ensure data consistency and usability.

**6.2 Ethical and Privacy Considerations**Handling medical information introduces significant ethical considerations related to patient privacy and data protection. In compliance with data privacy regulations such as GDPR, our system will not store any patient-specific information, and all uploaded content will be anonymized during processing. Additionally, users will be required to confirm that any submitted data does not include personally identifiable information (PII).

**6.3 Model Accuracy and Reliability**Ensuring that the AI model delivers medically accurate and clinically valid outputs is a critical challenge. To mitigate this, we will conduct rigorous validation using expert-reviewed paediatric cases and refine the system through multiple iterations of testing. Continuous feedback from medical professionals will play a key role in maintaining diagnostic accuracy and clinical relevance.

**6.4 Usability and Accessibility**As the platform is intended for use by both healthcare professionals and concerned parents, the user interface must be intuitive and accessible. Striking the right balance between technical sophistication and simplicity will require iterative UI/UX prototyping, usability testing, and feedback-driven refinement to optimize the overall experience.

**6.5 Technical Challenges**Integrating large language models (LLMs) into a real-time processing system presents several technical challenges, particularly when handling large or complex medical texts. To address performance issues, we will implement optimization techniques such as text summarization, asynchronous task management, and GPU-accelerated inference to ensure responsiveness under load.

# 7. References

1. Nationwide Children's Hospital. (n.d.). Failure to Thrive (FTT). Retrieved from <https://www.nationwidechildrens.org/conditions/failure-to-thrive>
2. Stryker, C., & Kavlakoglu, E. (2024). What is artificial intelligence (AI)? IBM. Retrieved from <https://www.ibm.com/think/topics/artificial-intelligence>
3. Evans, R. S. (2016). Electronic Health Records: Then, Now, and in the Future. Yearbook of Medical Informatics, 25(S 01), S48–S61. https://doi.org/10.15265/IYS-2016-s006
4. Nationwide Children's Hospital. (n.d.). Failure to Thrive (FTT). Retrieved from <https://www.nationwidechildrens.org/conditions/failure-to-thrive?utm>
5. EpicShare. (n.d.). Small Babies Grow Stronger with Improved Nutrition Monitoring. Retrieved from <https://www.epicshare.org/share-and-learn/small-babies-grow-stronger-with-improved-nutrition-monitoring>
6. Dewey, F. E., Grove, M. E., Pan, C., Goldstein, B. A., Bernstein, J. A., Chaib, H., ... & Ashley, E. A. (2019). Development of syndrome-specific growth charts in EHRs. Journal of Biomedical Informatics, 93, 103147. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6815511>
7. American Academy of Pediatrics. (2024). Special Requirements of Electronic Health Records in Pediatrics. Pediatrics, 154(4), e2024068509. https://doi.org/10.1542/peds.2024-068509
8. Empeek. (n.d.). Epic vs Cerner: EHR Systems Comparison. Retrieved from <https://empeek.com/insights/epic-vs-cerner>
9. Sarker, A., & Gonzalez, G. (2024). Cancer detection with AI. Journal of Medical Systems, 48(2). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11047988/>
10. Zhang, L., & Smith, J. (2024). Limitations of AI in Pediatric Growth Disorders. Pediatric Health Review, 12(1), 22–30. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11941653/>
11. Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., ... & Kumar, A. (2019). Natural language processing of clinical notes on chronic diseases: Systematic review. Journal of the American Medical Informatics Association, 26(4), 345–353. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6528438/>
12. Liu, Y., Chen, Q., & Wang, Y. (2022). A deep language model for symptom extraction from clinical text. Journal of Biomedical Informatics, 128, 104058. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9074854/>
13. Wu, H. Y., & Luo, L. (2020). NLP tools in extracting biomedical concepts. BMC Medical Informatics and Decision Making, 20, 52. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01352-2>
14. John Snow Labs. (n.d.). Clinical NLP Platform. Retrieved from <https://www.johnsnowlabs.com/clinical-nlp/>
15. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. Microsoft Research. <https://www.microsoft.com/en-us/research/publication/capabilities-of-gpt-4-on-medical-challenge-problems/>
16. Williams, J., & Lee, K. (2023). AI in digital health: A comprehensive review. PLOS Digital Health, 2(3), e0000198. <https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000198>
17. Penn State University. (2023). Improving efficiency and reliability of AI medical summarization. Retrieved from <https://www.psu.edu/news/information-sciences-and-technology/story/improving-efficiency-reliability-ai-medical-summarization>
18. Lin, H. C., Wang, Z., & Xu, Y. (2022). AI in clinical documentation: From assistance to automation. Journal of the American Medical Informatics Association, 29(5), 789–797. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8826344/>
19. Peng, Y., Yan, X., & Lu, Z. (2020). Transfer learning in biomedical NLP: A case study on BERT models. arXiv preprint. <https://arxiv.org/abs/2005.14165>

# 8. Appendix A – User Experience Questionnaire

Please rate your agreement with the following statements:  
1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

|  |  |
| --- | --- |
| Statement | Rating (1–5) |
| It was easy to a submit medical text to the system. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The system responded quickly after I submit a concern. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The diagnostic suggestions provided were clear and understandable. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The treatment recommendations made sense to me. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The user interface was easy to navigate. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| I felt confident using the system without assistance. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The system provided useful information that I could act on. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| I would consider using this system again in the future. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| I would recommend this system to other parents or healthcare professionals. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| Overall, I was satisfied with my experience using the system. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The results matched what I expected based on the document I uploaded. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The system helped me understand the medical content better. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The system handled different types of documents well. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| I felt the system respected my data privacy. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| The educational information was useful and easy to understand. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| I would trust this system in a real healthcare context. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |
| I found the process of using the system smooth and logical. | ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 |

Additional Comments:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_