ANOMALY

EECH)

BY

IKE, CHIEMERIE GIBSON
GOU/U22/CSC/860

DEPARTMENT OF COMPUTER SCIENCE,
FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY,

GODFREY OKOYE UNIVERSITY, ENUGU STATE.
IN PARTIAL FULFILLMENT OF THE AWARD OF BACHELOR OF SCIENCE (B.SC), DEG REE IN COMPUTER SCIENCE.

SUPERVISOR: DR. FRANK OKEBANAMA

JUNE, 2026

APPROVAL

This study titled "**ANOMALY PREDICTION MODEL FOR CHILD DEVELOPMENT (DELAYED SPEECH) USING MACHINE LEARNING**" has been reviewed and approved by the Computer Science Departments Committees in the Faculty of Computing and Information Technology, Godfrey Okoye University, Enugu, Nigeria.

Dr. Frank Okebanama Supervisor	.... Signature	.... Date
External Examiner External Examiner	.... Signature	.... Date
Dr. Frank Okebanama HOD, Department of Computer Science and Mathematics	.... Signature	.... Date
Prof. F.I Ajah Dean, Faculty of Computing and Information Technology	.... Signature	.... Date

CERTIFICATION

I hereby declare that the above-mentioned research work has been done solely by me, and the findings have been obtained by me through my personal study and observati on. Accordingly, I shall bear full responsibility for any adverse consequences of cheati ng as per the University policy.

____________________ ____________________

IKE, CHIEMERIE GIBSON **DATE**
GOU/U22/CSC/860

DEDICATION

This study, above all, is offered in thanksgiving to the All Powerful God for His blessings, wisdom, might, and guidance which enabled me to overcome all obstacles that came my way in doing this research and in my studies.

This is also my gratitude to my loving and supportive parents without whom I wouldn’t be here today, for their unwavering love, sacrifices, encouragement, and foundation of my growth and success.

In particular, I would like to dedicate the paper to my brother **Sean**, who had difficulties with speech development and became a role model in understanding the problems that children with speech difficulties face. Observing how my brother suffered from an inability to convey his desires and feelings and all the difficulties he had to face until he learned to speak when he was nine helped me comprehend the importance of early speech development interventions.

In conclusion, this study is also dedicated to all the children in Nigeria facing speech and language development issues and their respective families for supporting them throughout the process. It is my sincere wish that this project makes a slight contribution in making sure that there is increased awareness and proper treatment provided to these children so that they may reach their full potential in life.

ACKNOWLEDGMENTS

First and foremost, I would like to thank **Almighty God** for all the grace, inspiration, and strength He gave me throughout the entire period of carrying out this project. None of this would have been possible without Him.

My supervisor, **Dr. Frank Okebanama**, deserves my deepest gratitude for the guidance, criticism, and encouragement he showed me throughout this study. His input as well as his commitment to academic excellence have been invaluable in helping me complete this study.

It is my pleasure to express my gratitude to the Head of the Department, **Dr. Frank Okebanama,** my Dean**, Prof. Ifenyiwa Ajah** and all the lecturers in the Department of Computer Science, Godfrey Okoye University, for their devotion to achieving academic excellence and for providing me with the knowledge and skills that have made this study possible.

My heartfelt thanks go to my parents, **Mr and Mrs Ifeanyi Nwankwo** and family for all the encouragement they have provided through their unselfish financial support and constant words of support and love.

Finally, I thank all my colleagues and friends who offered assistance, shared ideas, and provided moral support during the course of this project. Your contributions, large and small, are deeply appreciated.

ABSTRACT

Childhood developmental anomalies such as speech delay, ASD, and ADHD account for an estimated 10–15% of children worldwide, but in Nigeria, most cases go unnoticed until the affected child reaches school age, owing to the critical lack of skilled developmental experts and excessive dependence on clinical expertise rather than diagnostic tools. The current study sought to design a machine learning-based intelligent model called the Anomaly Prediction Model for Child Development that is able to predict childhood developmental anomalies within the age group of zero to five years based on simple clinical and demographic features. After an exhaustive analysis of seven peer-reviewed journals within the years 2016 to 2025, it was determined that Random Forests, Support Vector Machines, and Logistic Regression have proven their effectiveness in structured development prediction problems with questionnaires providing accuracy rates of 70%-90%. A two-step design approach of Object-Oriented Analysis and Design (OOAD) in conjunction with CRISP-DM (Cross Industry Standard Process for Data Mining) was utilized where OOAD was responsible for application layer design and UML modelling, and CRISP-DM provided structure for the 6 step ML pipeline.Publicly available data sets from the developmental milestone database from the CDC, standards for child development from the World Health Organization, and data sets found in published research studies have been obtained and pre-processed using the programming language Python (using Pandas) to form a representative training data set. Data processing involved feature engineering, handling missing values, normalization, and an 80/20 split between training and test sets. Several supervised machine learning algorithms were then trained using Scikit-learn including Random Forest, SVM, and Logistic Regression. Random Forest method generated improved performance results on classification accuracy measures which outperform all benchmarks created by other researchers. Feature selection identified developmental milestone deviation, medical history of patients, socio-demographic factors, and behavioral factors as critical factors for diagnosis. The system architecture design involved the creation of a framework which included a presentation layer, application layer, and database layer based on SQL based relational database management systems. This study shows that the construction of a reliable, affordable, and scalable prediction system for developmental disorders is possible through use of freely available information and open source software applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

The emergence of speech abilities is one of the key aspects of early childhoo d . Starting from birth up until five years of age, children are supposed to pass the wa y from instinctive utterances and baby-talk towards learning first words, increasing ex pressive vocabulary, and creating phrases and sentences. All of these milestones are vital in developing the communicative skills necessary for children to be socially and cognitively competent. In the case where a child is unable to attain all of these speech milestones in the stipulated period, this results in what is known as delayed speech, the inability to speak words and sound as much as one's peer group in accordance to age. As noted by Rupert, Hughes, and Schoenherr (2023), delayed speech is among the most commonly encountered developmental problems in family practice settings, and parent's worries and milestone assessment serve as the foundation for identification.

Late speech development occurs in about 5-10% of preschool children worldwide, which means that this disorder represents the most widespread neurodevelopmental issue among young patients (Borovsky, Thal and Leonard, 2021). The adverse impact of late speech development is considerable; indeed, if not diagnosed and treated in a timely manner, children suffering from this condition face an increased chance of experiencing speech problems, low literacy rate, lack of social skills, poor academic results, and inability to find a job in the future. Unfortunately, although its clinical implications are quite serious, the approach to detecting such disorders remains time-consuming and inefficient due to its heavy reliance on subjective observations.

For Nigeria, these issues become more serious by leaps and bounds. Nigeria boasts of over 220 million people, including a very high population of children aged less than five years in Africa, but there is a severe lack of speech-language pathologists in the country, and most of the specialists are found only in a few urban tertiary hospitals in the country. Research has shown that delayed speech formation is responsible for 21 percent of all language disorders in one particular group of Nigerians in North Central Nigeria. Most Nigerian children who experience speech delay fail to get detect ed until their poor academic performance at school alerts anyone to this problem, whi ch happens after the important time of neuroplasticity when they can actually learn h ow to speak.

Machine learning, a field within artificial intelligence where computer algorith ms are employed to discover predictive patterns from data, provides a paradigm-shifti ng chance to solve this problem. According to Gasparini et al. (2023), who published t heir findings in the Journal of Child Psychology and Psychiatry, a Random Forest mo del could successfully predict delayed speech development at the age of 11 with a se nsitivity and specificity of more than 70 percent based on seven parent-reported item s. Moreover, Borovsky et al. (2021), who published their results in Nature Scientific Re ports, reported an AUC of 0.73. The following paper will develop an anomaly predictio n system known as ChildSpeak, which is a web-based application designed using the programming languages of JavaScript and Next.js framework along with a bespoke machine learning classifier algorithm referred to as the random forest with training da tasets of 2,408 samples including 1,054 Kaggle toddler samples, 1,054 Kaggle childre n samples, and 300 WHO/CDC synthetic samples.

1.2 Statement of the Problem

However, despite all the established outcomes associated with undiagnosed d elayed speech development in childhood, the present means through which potential patients can be identified in Nigeria are clearly insufficient. There are at least four iss ues that have inspired the need for the present research. Firstly, there is a clear scarcit y of speech therapists in Nigeria, whereby only a few tertiary-level urban hospitals pro vide their services. Secondly, the present method of identification is highly subjective and involves inter-rater variability. Third, late diagnosis remains the order of the day: Most Nigerian children that are affected are diagnosed when the important neuroplas ticity period of the first three years has passed, limiting the efficacy of speech therapy in these cases. Fourth, presently, there is no web-based machine learning solution ai med at diagnosing abnormalities in the development of children's speech abilities for use by Nigerians without sophisticated equipment.

1.3 Aim and Objectives of the Study

The aim of the research entitled "Anomaly Detection in Child Development (Delayed S peech) Using Machine Learning". In order to realize the above-stated aim, the followin g are the specific objectives of the study:

1. Review literature on how machine learning has been used to predict delayed spee ch development among children.
2. Collect, integrate, and preprocess datasets that have been merged with WHO and CDC speech milestones.
3. Develop a Random Forest classifier using nine critical developmental markers to predict delayed speech development.
4. Design and an approach that helps parents monitor the speech development of t heir children.
5. Implementing a Next.js web app for this approach
6. Perform performance analysis of the models using accuracy, precision, recall, an d F1-score metrics.

1.4 Significance of the Study

This paper has important implications for several stakeholders:

- **Health professionals, paediatricians, and primary care practitioners** - The ChildSpeak model serves as a decision support tool for effectively assessing children at risk of speech delays through primary health workers in underserved areas where such assessments have not previously been possible due to resource constraints.
- **Parents and carers** - Early diagnosis guarantees that children can access speech therapy at the right time, which is extremely important for children aged three to five due to the high neuroplasticity of the brain at this stage.
- **Policymakers and healthcare providers** - This study supplies empirical data to support investment in the development of digital tools and the use of AI in speech delay assessment in Nigeria and similar low- and middle-income countries.
- **Academia** - The study fills a gap in the existing body of literature, which largely consists of research conducted in high-income countries.
- **Children in Nigeria** - This study ensures that no child's communicative development is impaired as a result of speech delay detection coming too late.

1.5 Scope of the Study

The research targets anomaly detection in speech acquisition, in particular, speech delay, among children between the ages of 12 months and 60 months old, when speech milestones are acquired and when early interventions prove to be more effective. The assessment evaluates the expressive speech output of the child with respect to WHO and CDC speech milestones and nine risk factors identified by clinicians. The research will not cover the detection of other anomalies such as motor anomalies, co gnitive anomalies, or social emotional issues. The project implements the research using JavaScript as the main programming language and Next.js for the full-stack web framework. The machine learning prediction engine is a customized Random Forest model trained on a mixed data set of 2,408 data points where the dataset has 1,054 toddlers and 1,054 children's developmental screening data from Kaggle enhanced with 300 WHO/CDC simulated data points. The nine predictive variables used in the machine learning algorithm include: age of the child in months, gender, number of words spoken, pre-mature birth, presence of any speech delay problems in the family, hearing problem, whether the home speaks more than one language, average screen time per day in hours, and daily interaction time between parents and their children in hours.

1.6 Limitation of the Study

Three limitations in the research can be mentioned. The first limitation concerns the representation of the data. For training purposes, the researchers used models based on the ASD datasets from Kaggle, which come from international sources. The labels were used as proxies for developmental delay, while the environmental variables were artificially generated and thus contain some degree of approximation, which would not exist in a dataset developed solely for the purpose of identifying speech delay in Nigeria. The second limitation refers to the absence of the implementation of this research into practice, that is, conducting tests in a real-life clinical setting using actual patients. The third limitation of this project is associated with the period in which it has been conducted.

1.7 Operational Definition of Terms

Delayed Speech Development: This is the main anomaly identified through the use o f this system. It involves a state where a child's word production skill in speech is significantly lower than the milestone standard set by WHO and CDC at their chronological age level but not caused by any underlying factor like total deafness or intellectual impairment.

Speech Development Anomaly: This refers to a deviation in expressive language skills of a child relative to the milestone standards for the same chronological age. This is defined in terms of a combination of fewer word counts than expected and other risk factors flagged as predictive of delayed speech development.

Expressive Language: This is the ability to form spoken words in communication. It is evaluated based on how many unique words the child can speak, in comparison with the WHO and CDC standards for each age group.

WHO Speech Standard: These are the language milestones put forward by the World Health Organization and applied in this study: 5 words at 18 months, 25 words at 24 months, 100 words at 36 months, and 300 words at 60 months.

Machine Learning (ML): This is an area of artificial intelligence that deals with creating algorithms that learn from data without programming specific tasks. ML used in this research study is supervised learning using the custom binary random forest classifier developed in JavaScript.

Random Forest: Supervised machine learning model that constructs multiple decision trees and gives output in the form of mode. In our case, it is implemented from scratch in JavaScript and contains 50 trees with bootstrap sampling and Gini index criteria.

Anomaly Prediction Model: The main machine learning model used in this system is a supervised Random Forest classifier trained to predict anomaly in speech developm ent of children in childhood by classifying children with speech delay versus normally speaking children.

Next.js: An open-source full stack framework based on React and Node.js, which was the main framework used for the speech assessment interface and prediction API im plementation.

NextAuth.js: A freely available authentication module for Next.js that offers session s upport, credential sign-in functionality, and JWT tokens, utilized for creating the functi onality for user sign-up and protected routes.

MongoDB Atlas: A cloud-based fully managed MongoDB offering used for storing us ers' accounts details, their children’s profiles, and the speech assessments done on th em, accessible via the Mongoose object-document mapping framework.

Kaggle: An online site hosting open-access datasets. Two datasets from Kaggle—the Autism Screening for Toddlers dataset (Fabdelja, 2018) and the Autism Screening for Children dataset (Faizunnabi, 2019) — supply the system with 2,108 developmental s creenings records for training purposes.

Hybrid Dataset: 2,408 record dataset used for training made up of 2,108 realistic Kag gle records and 300 synthetic WHO/CDC speech milestone-based records.

Precision: The fraction of children diagnosed by the classifier to be delayed in speec h that indeed have delayed speech. High precision implies a low false positive rate, w hich is medically significant since false positive diagnoses lead to avoidable speciali st consultations.

Recall (Sensitivity): The fraction of children that truly have delayed speech detected by the algorithm. False negatives imply children not flagged, thereby missing crucial early-stage speech therapy.

F1-Score: The harmonic average of precision and recall, a single balanced performance measure accounting for both false positive and false negative.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter, a thorough literature review is provided regarding machine learning-based approaches for the identification of delayed speech development in children. This literature review is compiled from academic journal articles, some of which have been published in prestigious scientific journals, such as PubMed, PubMed Central (PMC), Nature Publishing Group, PLOS ONE, Wiley Online Library, Elsevier ScienceDire ct, Springer Nature Link, JAMA Network Open, BMJ Health Care Informatics, Frontiers in Pediatrics, Health Information Science and Systems, and ISA Transactions. All academic papers used in this chapter have been published in the last eight years (from 2022-2025). The chapter will be divided into five parts, including the theoretical background, literature review, summary table of reviewed papers, and research gap identification.

2.2 Theoretical Background

2.2.1 Delayed Speech Development in Children

Delayed speech development means that a child does not make utterances at the appropriate levels that other children of his/her age should make. Within the first five years, children have definite milestones of producing speech. These milestones include making babble sounds within six months, producing words in their first year, producing five to ten words in eighteen months, saying two-word phrases and producing about twenty-five words in two years, and producing one hundred words or even more within thirty-six months. Failure to meet these milestones in the absence of a known underlying factor means that the problem is one of delayed speech development. It is one of the most common developmental issues seen in paediatrics. Among the well-known risk factors that can lead to delayed speech development are family history of such problems, preterm birth, hearing loss, high exposure to screens, lack of talk-time between the child and the caregiver, and high exposure to more than one language at home. If not detected and referred to speech therapy early enough, consequences of delayed speech development may continue through adulthood, affecting social and academic life and future employability.

2.2.2 Delayed Speech Development in the Nigerian Context

The problem of undiagnosed speech delay among children in sub-Saharan Africa, and Nigeria in particular, has been worsened by the lack of availability of speech-language therapists and developmental pediatricians since they are only available in urban teaching hospitals. Previous research studies have revealed that speech delay makes up 21 percent of communication diagnoses in Nigeria; however, most of the children are diagnosed when their condition has become too serious, causing problems at school. Speech delay in Nigeria usually occurs with other forms of developmental problems like ASD, ADHD, and developmental delay, thus, complicating diagnosis because their behavioral manifestations are highly similar in early life stages. Considering the similarity between the behavioral symptoms of these conditions, the ASD screening databases from Kaggle utilized in this project for training the speech delay model can be regarded as reliable training sources when enriched with speech milestones from WHO and CDC guidelines.

2.2.3 Machine Learning: Core Concepts and Clinical Applications

Machine Learning (ML) is an area of Artificial Intelligence (AI) where computer algorithms learn to detect patterns within the provided data without having been explicitly designed to do so for each case. Supervised Learning refers to ML in which the algorithm trains itself based on a labelled training set, where the results are known. Among the commonly used supervised learning algorithms in pediatric research are Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN). In this project, a Random Forest algorithm written in JavaScript is incorporated into a Next.js application to make predic tions on speech delay. The use of machine learning provides many benefits compare d to the traditional method of clinical evaluation, including identification of complicat ed and nonlinear associations between the features, processing of multiple risk factor s at once, reduction of subjectivity and inter-evaluator variation, and deployment via a web application.

2.2.4 Feature Engineering and Data Sources in Speech Delay Research

The accuracy of a speech delay detection model through machine learning he avily relies on the features selected as input. When using machine learning to detect s peech delay, features used include information gathered from questionnaire tools use d by parents like MacArthur-Bates Communicative Developmental Inventories (MBCD I), Ages and Stages Questionnaire (ASQ), and CDC developmental milestone checklist s, clinical milestone information from electronic health record systems, and demogra phic risk factors like hours of screen time and verbal interaction between parent and c hild. Among some common techniques used during preprocessing are missing value treatment, normalizing the continuous attributes, and performing feature selection usi ng feature importance. In our experiment, Z-score normalization is performed on all ni ne independent attributes, and the training set includes a combined dataset of record s from Kaggle developmental screenings and artificial WHO/CDC speech milestones-based records.

2.2.5 Evaluation Metrics in Clinical Machine Learning for Speech Delay

In clinical machine learning of speech delay diagnosis, one must pay much att ention to the measures other than accuracy, as class imbalance is usually a frequent case. The main measures used in evaluation are: sensitivity/recall, measuring the rat e of positive instances identified by the classifier; precision, measuring the rate of truly positive instances among those predicted to be positive; F1-score, the harmonic mean of precision and recall; and AUC. The benchmark of above 70 percent for sensitivity and specificity has been used in other studies as the cutoff point for an acceptable level of performance (Gasparini et al., 2023; Gasparini et al., 2024). In the current study, however, the trained Random Forest speech delay classifier showed a recall score of 95.74 percent and precision of 98.90 percent, which is well above the benchmark.

2.3 Review of Related Works

In this part, an analysis of sixteen scientific papers published from 2022 to 2025 that deal directly with machine learning as a tool for prediction and identification of speech delay and/or other developmental issues in children will be presented.

The first scientific paper is Borovsky, Thal, & Leonard (2021). This research was published in Nature Scientific Reports and has the DOI: 10.1038/s41598-021-85982-0 (PMC8047042). The article provides predictive models of low language outcomes based on parental-report measures of early language abilities. The study integrated two longitudinally collected datasets where the children’s early language development was assessed using the standardized measurements from the MacArthur-Bates Communicative Developmental Inventories and graph theory-based metrics of the structure of lexico-semantic network constructed from toddlers’ expressive vocabularies. The study used Random Forests classifiers to predict which of the children were going to develop poor language skills in the future, resulting in an AUC score of 0.73. The main findings were that there were unique structural predictors for developing language disorders in addition to vocabulary size. The main limitation of the study was its depend ence on parental self-reported data.

The study published by Sharma & Singh (2022) in the journal Computer Methods and Programs in Biomedicine (Elsevier; DOI: 10.1016/j.cmpb.2021.106487) used a 1D-Convolutional Neural Network (1D-CNN) and hybrid deep learning algorithm for automatic classification of children with SLI based on their raw speech data. Instead of relying on handcrafted feature engineering, Sharma & Singh's (2022) methodology relies on training the neural network using raw speech signals. However, the biggest problem with this study is that its sample size was too small, which limits its applicability; also, it needs an audio recording system that is not available in Nigeria.

In Gasparini et al. (2023), which appeared in the Journal of Child Psychology and Psychiatry (Wiley/ACAMH; DOI: 10.1111/jcpp.13733; PMC10952842), machine learning analysis was performed on around 1,990 data obtained from the parent-reported questionnaires of the Early Language in Victoria Study cohort for identifying the minimal set of predictors for low achievement in language at age 11. Using the Random Forest algorithm for feature selection and the SuperLearner ensemble method, they attained 73 percent sensitivity and 77 percent specificity at 24 months, and 75 percent sensitivity and 85 percent specificity at 36 months, using predictor sets containing only seven items each. The main drawback is the relatively fair predictive performance reported, as well as the absence of comorbidities in the cohort participants.

Validation of predictor sets found in the Gasparini et al. study from 2023 was carried out by Gasparini et al. (2024) in the International Journal of Language and Communication Disorders (Wiley; DOI: 10.1111/1460-6984.13086; PubMed ID: 38948964). The six-variable prediction model generated 78 percent sensitivity and 71 percent specificity for predicting poor language outcome at 11 to 12 years old, which provides good proof of generalisability. However, a disadvantage of this study is that, with 71 percent specificity, there was still a considerable number of false positives, and the study did not consider multilingual children or those having concurrent developmental problems.

Georgiou & Theodorou (2024) from the Journal of Technology in Behavioral Science (Springer Nature; DOI: 10.1007/s41347-024-00460-4; arXiv: 2311.15054) created a neural network-based automatic system for diagnosing Developmental Language Disorder in children aged 7-10 years old who are native speakers of Cypriot Greek. Variable importance analysis showed that expressive aspects were more important in predicting this disorder than perceptual ones. The first major flaw is the tiny sample size (n=30), and overfitting is expected as a consequence. Another problem is the dependency on language-specific variables.

Chen et al. (2025) from the journal PLOS ONE (DOI: 10.1371/journal.pone.0324204; PMC12091767) sought to determine whether the frequencies of three commonly recorded therapeutic procedures can forecast developmental delay among children. This research utilized patient data from 2,552 paediatric outpatients in a Taiwanese hospital using the Deep Neural Network, Support Vector Machine, and Decision Tree models. The former had superior results. However, one shortcoming is that using data from a single institution makes generalization difficult, and the study does not distinguish between outcomes of speech delay.

Goodwin Cartwright et al. (2024), in an article in JAMA Pediatrics (DOI: 10.1001/jamapediatrics.2023.5226; PMID: 38048098), undertook a retrospective cohort tim e-series study for assessing first-time diagnosis rates of speech delay in children under five years old, using a nationwide dataset of electronic health records in the United States obtained from Truveta. An increase in first-time diagnoses of speech delays was observed to be statistically significant for the period 2018/2019 and 2021/2022 in all age strata ($p<0.001$), thus indicating a pandemic-related rise. However, EHR-based diagnosis rates can be confounded by healthcare utilization

Rupert, Hughes, and Schoenherr (2023) in American Family Physician (PubMed ID: 37590860; Vol. 108(2):181-188), this work presented an evidence-based clinical review concerning the identification and evaluation of speech and language delays among children in primary care. This study confirmed that parental concern paired with deviation from milestones at 12, 18, and 24 months is indeed a valid reason to take clinical action, while also confirming that family history, hearing loss, premature birth, and poor language input are validated risk factors. This paper may not provide us with a machine learning algorithm, but it does offer us the clinical workflow that the ChildSpeak platform should be integrated within, as well as validation of all risk factor features in use in this project.

Suthar et al. (2022), whose article was published in PLOS Digital Health (https://doi.org/10.1371/journal.pdig.0000041; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9931328), sought to explore the use of Landmark analysis of acoustic signals and new knowledge-based features alongside machine learning classifiers for automatic speech disorders screening among children. Specifically, a new set of features, previously unexplored in any other literature, was developed, and systematic comparison between the performance of linear and non-linear classifiers was made to demonstr ate that the Random Forest classifier with the above-mentioned set of features yielded the best classification performance. However, the principal drawback of the study is that the approach relies heavily on audio recording equipment and transcribers – the resources that are not available in the resource-poor Nigerian environment of primary healthcare.

Shrivastava, Singh & Agrawal (2024) published a paper in Health Information Science and Systems (Springer; doi:10.1007/s13755-024-00277-8, pubmedid: 38464462) the use of multiple ML classifiers including Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF) and Artificial Neural Network (ANN) to analyze Q-CHAT and AQ-10 questionnaires dataset among toddlers, children, adolescents and adults. Missing values were addressed using kNN imputation technique. Random Forest was able to classify all the four data sets with 100 percent accuracy without any sign of overfitting while performing better than the Deep Neural Network and Convolutional Neural Network model. The limitation is that accuracy of 100 percent can be an indicator of the lesser variability of structured questionnaire data set.

Rajagopalan et al. (2024), who reported in JAMA Network Open (doi: 10.1001/jamanetworkopen.2024.29229, pmc: 11333987), ML-based models have been created to forecast ASD from 28 variables in terms of background and health aspects using Simons Foundation SPARK dataset (n = 30,660). Four algorithms were used by the researchers including Logistic Regression, Decision Tree, Random Forest (RF), and Extreme Gradient Boosting (XGBoost), while the latter two algorithms achieved accuracy of nearly 80 percent to predict ASD in individuals below the age of two. The major disadvantage was the fact that the cohort used was from a US research database where there was not much socio-economic variation and also did not have environmental predictors like screen time. The main takeaway from this experiment is the fact that a small set of features based on developmental and background history is sufficient for accurate neurodevelopmental prediction.

Chen et al. (2022) in BMJ Health Care Informatics (DOI: 10.1136/bmjhci-2022-100544; PMC9462117) studied the potential application of big-data based health insurance claims datasets derived from IBM MarketScan (sample size 38,576 children) in developing machine learning models for the prediction of autism spectrum disorders at 18 to 30 months of age. Two approaches - logistic regression with LASSO and Random Forest achieved an area under curve of 0.758 and 0.775 at 24 months, respectively. Age positively correlated with model predictive power. The main constraint of this study was the use of data from American insurance claims that lacked information regarding specific screening and behavioral outcomes for ASD, and speech delay was not assessed independently. This study demonstrates the viability of applying routinely collected clinical markers to identify developmental delay among children, justifying the ChildSpeak strategy of gathering clinical information via questionnaires.

Wan et al. (2025) carried out a retrospective case-control study involving 296 children and analyzed the impact of parental screen time, child screen time, and the quality of parent-child interaction on the likelihood of developing language delay. The participants were classified according to “Bayley Scales of Infant and Toddler Development-III (BSID-III)”. Excessive screen time among children and dysfunctional parent-child interactions were separately linked to increased likelihood of developing language delays. The most important modifiable predictor of the development of language del ay was quality of parent-child interactions. The major drawback is the sole use of a single institute from China, which prevents the extrapolation of results to an African setting. This paper provides empirical evidence on the need to incorporate the number of hours of screen time as well as the number of hours of parent-child interaction as predictors in the ChildSpeak model.

In the paper "PSE Guided Anomaly Detection in Children's Speech," Suthar et al. (2022b), published in ISA Transactions (Elsevier), introduced Process Systems Engineering (PSE) fault detection techniques for detecting speech disorders in children, viewing anomaly detection as analogous to engineering fault detection techniques. Parsimony principle was employed in feature selection and parameters, and landmark acoustic analysis along with SVM, Random Forest, and Logistic Regression was utilized. PSE-inspired feature selection helped enhance generalizability of the models over existing ML methods. The key limitation of the study was dependence on audio signals and a limited corpus of clinical data. This article highlights the theoretical basis for considering childhood speech development abnormalities to be detectable deviations from a normal baseline, an idea adopted by the authors in their research.

The study was undertaken by Moges et al., (2024) and published in the BMC Pediatrics journal (DOI: 10.1186/s12887-024-04862-4; BioMed Central) involving 229 cases and controls, aged between 12 months and 12 years, at Yekatit 12 Hospital in Addis Ababa, Ethiopia. In this case-control study, family history of speech difficulties, difficult birth experiences, minimal verbal engagement by caregivers, and low parental educational status emerged as important independent variables through logistic regression analysis. Single-center recruitment and retroactive bias are the major weakness es. Although this study is based on a sample from an Ethiopian population and not a Nigerian one, it remains the most clinically relevant source of Sub-Saharan African data on risk factors for speech delay, thereby serving as the basis for contextualising the ChildSpeak system in Nigeria.

Boerma et al. (2023) carried out a systematic narrative review summarizing neurobiological evidence of risk factors related to Developmental Language Disorder, considering how genetic predisposition, prenatal problems, hearing loss, bilingualism, and environmental adversity each affect DLD by different neurobiological mechanisms. The review revealed that DLD is characterized by multifactorial aetiology, where each risk factor alone is neither necessary nor sufficient for causing DLD, but the interaction of several moderate risk factors significantly increases the developmental risk. The main problem here is that results were obtained based on the data from Europeans which have little generalizability to the linguistic and natural environment of Sub-Saharan Africa. This research gives us the neurological basis of all the nine predictor variables used in the ChildSpeak model design.

2.4 Summary of Related Literature

Table 2.1 provides a structured summary of the sixteen reviewed studies, presenting methodology, work done, results/findings, limitations, and relevance to the current study.

Table 2.1: Summary of Related Literature on Machine Learning for Speech Delay and Developmental Anomaly Detection (16 Studies)

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
1	Borovsky et al. (2021)	Scientific Reports (Nature)	Random Forest; network science; MBCDI longitudinal datasets	Constructed predictive models of poor language performance through parental report vocabulary and linguistic network characteristics.	Language disorder prediction ROC AUC 0.73. Network structural properties have independent prediction capacity regardless of word count.	Parental self-reporting bias. Small sample size reduces generalizability	Validates Random Forest and expressive vocabulary as the primary speech delay predictor for this study
2	Sharma & Singh (2022)	Computer Methods (Elsevier)	1D-CNN, hybrid CNN-LSTM; raw speech waveform input; SLI corpus	Specific Language Impaired children classified using raw speech data without manual feature extraction	1D-CNN proved superior to traditional hand-engineered acoustic features. Deep learning is capable of modeling the nuances in speech	Small size of data leads to overfitting issues. Audio capture setup not possible in limited-resource settings	Motivates the questionnaire-based approach adopted in this study for Nigerian primary healthcare settings
3	Gasparini et al. (202	J. Child Psychol	Random Fores	Predictive sets were ide	Sensitivity = 73%, Specificity = 77% at	Accuracy was only moder	Directly informs feature s

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
	3)	ogy (Wiley/ACAMH)	t, Super Learner ensemble; ELVS cohort; ~1,990 parent-reported items	ntified for poor language outcomes by age 11 using parent-reported items at ages 24 months and 36 months	age 24 months. Sensitivity = 75%, Specificity = 85% at age 36 months	ate. Australian sample alone. Participants with comorbid disorders excluded	election and validates the 12 to 60 month assessment age range adopted in this study
4	Gasparini et al. (2024)	Int. J. Lang. Comm. (Wiley)	Random Forest, Super Learner; LSAC birth cohort (n=5,107); age 11-12 language outcome	External validation of the 6-predictor set from the 2023 study in a large independent population cohort	78% sensitivity, 71% specificity confirmed across 5,107-child cohort. Evidence of predictor generalisability established	71% specificity implies notable false positive rate. Multilingual children excluded	Confirms generalisability of questionnaire-based speech delay prediction approach
5	Georgiou & Theodorou (2024)	J. Tech. Behavioral Sci. (Springer)	Neural network; k-fold cross-validation; perce	Developed automated neural network identifying Developmental Lan	High accuracy, precision, recall, F1, AUC on unseen data. Production features outperformed perceptual features	Very small sample (n=30) raises significant overfitting concerns. Language-spe	Confirms expressive speech production features are stronger diagnostic marke

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			ptual and language production data	guage Disorder in Cypriot Greek-speaking children aged 7-10		cific — not transferable to other populations	rs than perceptual assessment features
6	Chen et al. (2025)	PLOS ONE	DNN, SVM, Decision Tree; routinely collected therapy-frequency data; 2,552 paediatric outpatients Taiwan	Investigated whether therapy-service frequencies could predict developmental delay in paediatric outpatients	DNN achieved highest performance. Routinely recorded low-cost administrative data generates clinically useful predictions	Single-institution dataset restricts generalisability. No separate performance metric for speech delay as distinct outcome	Validates low-cost routinely collectible data approach and confirms DNN performance on developmental prediction tasks
7	Goodwin Cartwright et al. (2024)	JAMA Pediatrics (AMA)	Time-series STL analysis; retrospective cohort; Truveta national	Examined temporal trends in first-time paediatric speech delay diagnoses before and after th	Statistically significant increases in speech delay diagnosis rates 2018-2022 across all age strata (p<0.001). Pandemic surge documented	Descriptive analysis only — no predictive ML model developed. EHR rates may reflect changes in healthc	Provides critical epidemiological justification for the urgency of scalable automated speech delay scree

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			EHR database; USA 2018-2023	e COVID-19 pandemic		are-seeking behaviour	ning tools
8	Rupert et al. (2023)	American Family Physician	Systematic evidence-based clinical review; primary care diagnostic guidelines	Reviewed identification, evaluation, and management of childhood speech and language delays in the primary care setting	Parental concern and milestone deviation at 12/18/24 months confirmed as reliable referral trigger. Risk factors confirmed	Clinical review only — no ML performance metrics reported. US-focused context may not fully reflect Nigerian realities	Defines clinical workflow and confirms all 9 risk factor features used in this study as clinically validated speech delay predictors
9	Suthar et al. (2022)	PLOS Digital Health	SVM, Random Forest, Logistic Regression; acoustic Landmark analysis; GFCC features; child speech	Developed novel knowledge-based acoustic features and ML classifiers to automate speech disorder screening in children, reducing dependence on s	Random Forest with proposed features achieved highest classification accuracy. Novel features outperformed existing LM-based features in distinguishing speech-disordered from typical children	Requires audio recording infrastructure and specialist transcription; not deployable in low-resource settings without clinical recording equipment	It serves as a demonstration of the ability of knowledge-driven acoustic measures and machine learning methods in the automated screening process of speech disorders

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			corpus	ubjective auditory perceptual analysis			in children.
10	Shrivastava et al. (2024)	Health Inf. Sci. Syst. (Springer)	RF, SVM, KNN, ANN; kNN Imputer; MinMaxScaler; Q-CHAT and AQ-10 questionnaire datasets across all age groups	Classified ASD and typical development across toddler, child, adolescent, and adult ASD questionnaire datasets using multiple ML classifiers with missing value imputation and normalization	Random Forest achieved 100% accuracy on all four age-group datasets with no overfitting, outperforming DNN and CNN-based approaches in generalization for real-time ASD detection	Single-dataset evaluation per age group; 100% accuracy may reflect dataset characteristics rather than true generalizability; clinical validation not performed	These results show the potential of applying artificial intelligence technology in developmental disorders screening and offer some useful information to develop such decision-making systems.
11	Rajagopalan et al. (2024)	JAMA Network Open (AMA)	Logistic Regression, Decision Tree, RF, XGBoost; SPARK	Developed and validated ML models predicting ASD from minimal background a	RF and XGBoost achieved ~80% accuracy for children under age two. Developmental milestones and eating behaviour were top pred	Cohort is predominantly Western and drawn from a research registry; clinical deployment i	This paper offers a good template for creating intelligent decision-making systems that cou

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			database (n=30,660); 28-feature background and medical history predictors	nd medical information, including developmental milestones, using a large US registry database	ictors. Good generalizability confirmed on independent cohort	n routine paediatric settings not validated; limited coverage of environmental risk factors	ld facilitate the early diagnosis of developmental problems among children.
12	Chen et al. (2022)	BMJ Health Care Informatics (BMJ)	Logistic Regression with LASSO; Random Forest; IBM MarketScan health claims database (n=38,576; 2005–2016); ASD prediction at 18–30	Examined feasibility of large-scale real-world health claims data to build ASD prediction models for young children using medical diagnoses and healthcare procedure codes as predictor variables	LR and RF achieved AUROC of 0.758 and 0.775 respectively for ASD prediction at 24 months. Prediction accuracy increased with age. Perinatal complications and developmental milestone deviations were top predictors	Claims data may lack completeness; excluded ASD-specific screening results; single-country US context; does not isolate speech delay as a distinct outcome	This proves that it is possible to build predictive models to forecast the risk of developing ASD in toddlers by analyzing real-world health claim data on a large scale.

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			months				
13	Wan et al. (2025)	Frontiers in Pediatrics	Retrospective case-control study; Bayley Scales BSID-III; NCATS parent-child interaction scale; logistic regression and ROC analysis; n=296 children (2020–2024)	Investigated the combined effect of parental and child screen time, parent-child interaction quality, and emotion regulation on language development delay in young children aged under five years	Excessive child screen exposure, dysfunctional parent-child interaction, and parental screen habits were all independently associated with elevated risk of language development delay; interaction quality was the strongest modifiable predictor	Retrospective case-control design limits causal inference; single hospital in China; parent-child interaction measured by structured scale may not reflect naturalistic daily interaction	Uses information from medical diagnoses and health care procedures as predictive measures in identifying early onset of ASD.
14	Suthar et al. (2022b)	ISA Transactions (Elsevier)	Process Systems Engineering (PSE) fault-det	Applied PSE fault-detection principles to speech disorder detection, u	PSE-guided feature reduction improved model generalizability. Demonstrated that disease/disorder detection can be	Audio-based approach; requires speech recording infrastructure not available	Shows the success achieved in using the parsimony principle in childhood sp

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			ection framework; Landmark acoustic analysis; parsimony principle for feature reduction; SVM, RF, Logistic Regression	sing the parsimony principle to minimise feature and parameter spaces for more parsimonious and generalisable classification of childhood speech disorders	framed as an engineering fault-detection problem with improved clinical parsimony	in community healthcare settings; small clinical corpus limits generalizability	eech disorders classification in an automated way.
15	Moges et al. (2024)	BMC Pediatrics (BioMed Central)	Case-control study; structured clinical interview; logistic regression; determinants analysis; children	Identified clinical and environmental determinants of speech and language delay in an African primary care hospital setting, examining family history,	Family history of communication disorders, birth complications, limited caregiver verbal interaction, and low parental education were significant independent predictors of speech and language delay in an African context	Single-institution design in Ethiopia; does not develop an ML model; retrospective case-control design introduces recall bias; findings not validated in Nigerian or West African	It shows which clinical and environmental causes play an important role in speech and language disorder development, such as family background, birth history, social

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			aged 12 months to 12 years; Addis Ababa, Ethiopia	birth history, socioeconomic factors, and caregiver interaction as predictors		populations	status, etc.
16	Boerma et al. (2023)	Neuroscience & Biobehavioral Reviews (Elsevier)	Systematic narrative review; neurobiological risk factor synthesis; DLD cohort studies across Europe; genetic, environmental, and neurodevelopmental risk factor meta-	Reviewed the neurobiological mechanisms underlying Developmental Language Disorder risk factors, synthesising evidence on how genetic susceptibility, perinatal complications, hearing impairment, and environmental exposure combine to produ	Confirmed multilayered aetiology of DLD: genetic factors, hearing impairment, prematurity, limited language input, and adverse environmental exposures all independently elevate DLD risk through distinct neurobiological pathways	Biological mechanism review does not provide ML performance metrics; findings drawn from European cohorts with limited applicability to Sub-Saharan African linguistic and environmental contexts	It is a good review about neurological processes that cause Developmental Language Disorder.

S/N	Author(s)	Journal	Methodology	Work Done	Results / Findings	Limitations	Relevance
			analysis	ce DLD			

2.5 Research Gap

Critical examination of the sixteen articles identified five important gaps that make a strong case for the introduction of the ChildSpeak solution. Firstly, the current body of knowledge is overwhelmingly concentrated within wealthier countries and those speaking English, including Australia, the USA, Taiwan, Cyprus, Sweden, China, and Ethiopia. With regards to Sub-Saharan Africa, the only contribution made by any paper in this context was by Moges et al. (2024). However, no machine learning-based research for speech delay detection has been carried out specifically within the Nigerian setting yet. Secondly, although Wan et al. (2025) and Boerma et al. (2023) emphasized the importance of screen time and parent-child interaction from a clinical point of view, the risk factors such as screen time, parent-child interaction hours, and bilingual family environment, which comprise the environmental risk factors, have not been incorporated into any of the machine learning algorithms reviewed, except for ChildSpeak. Third, the methods that rely on audio analysis for prediction in the works by Sharma and Singh (2022), Suthar et al. (2022), Suthar et al. (2022b), and Georgiou and Theodorou (2024) depend on infrastructure unavailable in Nigeria's primary care facilities; those methods based on questionnaires developed by Shrivastava et al. (2024), Rajagopalan et al. (2024), and Chen et al. (2022) continue to be the only practical solution. Fourth, all reviewed predictive models failed to create an accessible mobile app for caregivers without specialized equipment; our method creates a completely functional web application using React framework, which can be accessed anywhere and anytime. The fifth factor is that the rise in speech delay among children as a result of the pandemic was statistically significantly recorded by Goodwin Cartwright et al. (2024), and screen time research by Wan et al. (2025) proves the persistence of pandemic-induced behavioral effects which pose risks of development delays on a global scale – especially in Nigeria. The ChildSpeak tool addresses all five aspects mentioned.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.0 Introduction

This chapter is dedicated to the system analysis and design of ChildSpeak anomaly prediction system for detecting delayed speech development in 12-60 month-old children referred to as "Anomaly Prediction Model for Child Development (Delayed Speech) Using Machine Learning". In this chapter, you will find information on the following topics: analysis of the current system, motivation of the limitations which require solving of the problem with the proposed new system, objectives of the proposed system, system requirements specification, system design approach, UML diagrams, system design (input, output, database, and architectural).

3.1 System Analysis

3.1.1 Analysis of the Existing System

The current approach for assessing delayed development in speech among children in Nigeria entails largely a specialist-based assessment process. The conventional approach involves the initial identification of a problem in the development of the child's language by a parent through consultation at a primary health care facility. This is achieved through conducting informal developmental tests such as the Denver Developmental Screening Test, the Ages and Stages Questionnaire (ASQ) and the CDC Speech Milestones Checklist. For speech disorders, the child is referred to the Speech Language Pathologist (SLP) where an assessment is conducted. The process of assessment involves carrying out a diagnosis using standardized methods, audiometry among other measures, a process that normally takes weeks or even months. In rural and semi-urban environments, this cannot be done. This is because all data are recorded manually.

3.1.2 Limitations of the Existing System

The following limitations of the current practice were revealed based on the problem statement described in Chapter One and from the critical evaluation of literature revie w done in Chapter Two:

- Lack of speech pathologists; specialist speech services are available only in a few selected tertiary urban hospitals thus being unattainable for most Nigerian children.
- Late diagnosis and missed therapy time frame: The lengthy procedure of refer rals causes delays that make a child miss the crucial neuroplastic time frame of three initial years of life.
- Subjective speech testing process; the existing diagnostic method is subjectiv e and depends on the experience of an observer and memories of parents, hen ce resulting in many misdiagnosed cases.
- Lack of consideration for integration of environmental risk factors: Current scr eening tools ignore important predictors such as daily screen time, parent-child interaction time, and bilingual home environment.
- Lack of digital speech delay screening tool availability: Currently, there are no a ccessible digital screening tools based on data that can give estimates of spee ch delay risk in the primary healthcare sector of Nigeria.
- Use of paper records: Paper-based record keeping prevents tracking of develop ment milestones in speech and surveillance of speech delay prevalence in the general population.

3.1.3 Objectives of the Proposed System

The suggested ChildSpeak solution attempts to resolve all of the six limitations listed above. The ChildSpeak system would have the following main goals:

1. The ability to eliminate the reliance on expert evaluation for the initial diagnosis of speech delay through data-based risk assessments using the specially developed nine-item checklist available to any adult on any device capable of connecting to the Internet.
2. Dramatically shorten the amount of time needed for the diagnosis through instant delivery of speech delay risks assessment within three seconds after the assessment submission.
3. Mitigate subjectivity by employing a standardized process of Random Forest classification algorithm to deliver the same consistent objective outcome irrespective of clinicians/caregivers for similar input data.
4. Include all nine validated predictors for speech delay identified so far, including those based on environmental risks, which have been neglected by all other validated screening tools.
5. Develop a complete web-based application available to be used from any internet-connected device without any special equipment or installation requirement.
6. Achieve continual tracking of individual children's progress by incorporating authentication, profiling, and progress charting of each child using MongoDB Atlas.

3.2 System Requirements Specification (SRS)

3.2.1 Functional Requirements

1. The system shall present a conversation-style interface for evaluating speech through the presentation of developmentally appropriate questions one-by-one along with calming prompts based on context.
2. The system shall take answers to fifteen questions designed around identifying the nine key attributes related to speech.
3. The system shall send the results of the test to the pre-trained Random Forest model running on a Next.js API route where it will be processed and returned with a result in under three seconds.
4. The system shall provide a results dashboard displaying information on the severity of the problem, probability score, number of words counted versus WHO milestones, causes, action plan, and recommendations.
5. There should be user registration and login functionalities with password hashing using bcrypt and sessions using JWT with NextAuth.js.
6. There should be storage for completed assessments in the MongoDB Atlas database for users who authenticate their accounts.
7. There should be functionalities for viewing child profiles, viewing assessment history with graphs, creating a PDF document, and WhatsApp & email sharing.

3.2.2 Non-Functional Requirements

- Accessibility: Available through any modern web browser without the need for extra software installation on any computer, laptop, tablet, or mobile phone device.

- Performance: Predictive API will provide results within three seconds of completing the form.
- Security: User password must be stored using bcrypt hashing algorithm; session token needs signing using a secret key; child information shall not be shared with any external entity.
- Reliability: The speech delay detection tool should give consistent results for the same input during multiple sessions.
- Scalability: The architecture involving the use of Next.js and MongoDB Atlas should be able to handle growing user numbers without undergoing any major restructuring.
- Usability: The speech delay evaluation interface should not need any previous technical know-how from its users.

3.3 System Design Methodology

The design of ChildSpeak follows a mixed approach that integrates the strengths of both of the following frameworks in one system – Object-Oriented Analysis & Design (OOAD) and the Cross-Industry Standard Process for Data Mining (CRISP-DM). This was not an accidental choice because ChildSpeak is, after all, a system that consists of two different subsystems that require two different methods of designing. These two subsystems are: 1) The user interface component (web-based), which involves user interactions, assessment, authentication, and results delivery, and 2) the ML prediction component. OOAD is the well-known methodology for modeling and defining the behavior of software systems, whereas CRISP-DM is the worldwide recognized methodology for designing machine learning models using data. These two methodolo gies make up a unified hybrid methodology design where both play the role of manag ing subsystems they excel in, but none of them is complete on its own.

3.3.1 Object-Oriented Analysis and Design (OOAD)

The OOAD design paradigm involves grouping various parts of the system into objects that have well-defined properties and behaviors, then describing how these ob jects interact with each other in order to model system structure and behavior. When l ooking at the ChildSpeak system, three key reasons influenced the decision to choose OOAD as the design approach for its application layer.

One, the object-oriented approach to architecture (OOAD) matches well with th e modularity of Next.js and React, wherein all pages, components, and API routes are self-contained modules with clear input/output interfaces. The advantage here is that it avoids adding unnecessary layers of abstraction that may not match the architectu re and, instead, matches the design with actual implementation. Two, OOAD provides a way to improve individual modules such as the assessment form, prediction API ro ute, results dashboard, and authentication without affecting the architectural design, considering the interface needs of the software developed for caregivers. Thirdly, and more significantly, OOAD enables the creation of UML diagrams as a standardized la nguage independent approach for modelling systems prior to their implementation. T he four UML diagrams created in the research – use case diagram, activity diagram, c lass diagram, and sequence diagram – are the OOAD artifacts generated in this study and will be described in Section 3.4.

3.3.2 Cross-Industry Standard Process for Data Mining (CRISP-DM)

The CRISP-DM approach has been the most used and cited methodology in ac ademia in the field of machine learning and data mining processes. CRISP-DM is an agnostic model that uses six iterative phases in defining the complete life cycle of the machine learning process. These phases include: (1) Business Understanding, (2) Dat a Understanding, (3) Data Preparation, (4) Modelling, (5) Evaluation, and (6) Deploym ent. The CRISP-DM approach has been chosen for the ChildSpeak prediction engine f or the following three main reasons. First, the CRISP-DM approach encompasses the entire process of ML from deployment. The SEMMA approach, on the other hand, onl y reaches the assessment stage without any deployment, while the KDD process focu ses more on knowledge discovery in databases than on prediction. Deployment is ine vitable with respect to this task, where the ChildSpeak classifier must be serialised an d incorporated into a Next.js API route in less than three seconds. This is the output th at needs to be developed as the key outcome. Second, CRISP-DM is the only ML proc ess framework to make Business Understanding as its initial phase. This method was responsible for influencing the design of the ChildSpeak model itself: the recall lower bound and precision goal of greater than 90 percent were set prior to the choice of the dataset, which guaranteed that the model would be assessed based on clinically sign ificant performance metrics rather than arbitrarily high levels of accuracy. Third, CRIS P-DM is an iterative process, allowing for revisits to previous phases based on finding s from subsequent stages. In the case of the ChildSpeak methodology, this became n ecessary after discovering five missing features during Phase 2 (Data Understandin g).

3.3.3 Justification for the Hybrid Combination

Hybrid OOAD and CRISP-DM methodology has been chosen because, althoug h each of these approaches can be used separately, neither of them on its own can adequately cover both subsystem designs for the application of ChildSpeak.

This methodology would not be possible using the OOAD approach alone, since OOAD does not have the constructs for dealing with training data, feature distribution, evaluation metrics, and benchmarks. The OOAD of the ML portion will result in class diagrams related to the Random Forest classifier; however, it does not give any guidelines on what kind of data should be used, how data needs to be pre-processed, which algorithm needs to be applied, and how the clinical adequacy of the model can be ensured. On the other hand, CRISP-DM does not enable the creation of the ChildSpeak application portion as CRISP-DM is completely unaware of the concept of user interface, component hierarchy, user authentication process, and database schema.

The integration of the two methodologies in such a manner is addressed by allocating each methodology to the particular subsystem they were originally intended for: OOAD takes care of what the system looks like, how its elements are arranged, and how they communicate; CRISP-DM takes care of the learning capabilities of the system, what performance level should be achieved, and how the model is going to be deployed after training. The boundary between the two methodologies lies at the level of the prediction API service: it is an element of OOAD on the side of the application, and a deployment artifact of CRISP-DM Phase 6 on the side of machine learning..

3.3.4 CRISP-DM Design Phases Applied to ChildSpeak

The six phases of CRISP-DM are used in relation to the ChildSpeak project design and specification in this chapter to lay down the plan that will be implemented in Chapter Four.

Phase 1 - Business Understanding: In terms of business understanding, the objective is the correct classification of whether a child between the age of 12 and 60 months is delayed or normal in speech development, based on nine parameters obtained from a non-expert caretaker using a questionnaire. The data mining objective has been formulated as: Create a model that has at least 80 percent recall rate and greater than 90 percent precision for the testing dataset, thus significantly improving upon the 70 to 78 percent benchmarks for sensitivity set out in Chapter Two. Target variable: Binary category (1 = delayed; 0 = normal). Deployment criterion: A predictor that provides an answer in less than three seconds using a Next.js API route from any Internet-capable device.

Phase 2 - Data Understanding: Two publicly available Kaggle datasets were selected as sources of raw data: the Autism Screening for Toddlers dataset (Fabdelja, 2018; 1,054 observations; age group 12 to 36 months) and the Autism Screening for Children dataset (Faizunnabi, 2019; 1,054 observations; age group 4 to 11 years). Both datasets were considered in terms of class balance, feature set completeness, covered age groups, and consistency with nine required predictor variables. Q-CHAT-10 form and ASD classification used in both datasets were taken as an analog of developmental delays due to a high correlation of ASD and speech delay noted by Rupert et al. (2023) and Boerma et al. (2023). Five essential characteristics of the ChildSpeak framework that include words_spoken, screen_time_hours, parent_interact_hours, hearing_problems, and preterm_birth were discovered not to be part of either source, thus outlining the augmentation needs covered in Phase 3.

Phase 3 - Data Preparation Design: The approach used for the preparation of d ata was created with the intention to get a clean and complete dataset of 2,408 data points. The approach consists of four steps which include source specific feature mapping of Q-CHAT-10 attributes to the nine attribute format; augmentation of the five missing features through generation of synthetic values based on clinically derived probability distribution from Rupert et al. (2023) and World Health Organisation and Centre for Disease Control criteria for each class type; creation of 300 records through risk scoring algorithm with seed =42 based on World Health Organisation speech milestones; and data preprocessing involving Z-score normalisation, shuffle, and an 80/20 split generating 1,927 training and 481 test records.

Phase 4 - Modelling Design: Random Forest algorithm was chosen due to convergent support from Chapter Two studies: Gasparini et al., (2023, 2024), Shrivastava et al., (2024) and Rajagopalan et al., (2024) independently corroborated that Random Forest is the best algorithm for the structured/ tabulated developmental screening application. The classifier will be implemented customarily through JavaScript code - DecisionTree and RandomForest classes - utilizing Bootstrap Aggregating and Gini Split Criterion method without any external Machine Learning library. Model parameters include: trees = 50; max_depth = 8; min_samples = 5; random_state = 42. Majority vote is used as classification criterion; probability score as a proportion of trees supporting delay class prediction.

Evaluation Phase 5: Four evaluation measures are stated for evaluation of the results on the 481-record test dataset: accuracy, precision, recall, and F1-score. Confusion matrix with TN, FP, FN, and TP is used primarily to evaluate the system performance. False negatives, i.e., late detection of the condition by the algorithm, represent th e most clinically damaging misclassification errors. Minimum acceptable level of recall is 80%, which is higher than the sensitivity level of 78% reported by Gasparini et al. (2024).

Phase 6 - Deployment Design:

The trained classifier is serialized into src/ml_model/model.json containing all the forest architecture along with Z-score normalization coefficients. The deployment interface is the Next.js API endpoint located at src/app/api/predict/route.js. It receives the input variables, runs Z-score normalization on-the-fly from saved data, evaluates all 50 trees and outputs a complete prediction in the form of a JSON file within 3 seconds, including classification, probabilistic measure, severity score, WHO milestone assessment, causes and action plan.

3.3.5 Method of Data Collection

The dataset that is used for the experiment in this paper was built by combining two real-world datasets that are publicly available on Kaggle with some artificial data added. Firstly, there is the Autism Screening for Toddlers data set (Fabdelja, 2018), accessible at https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers, which includes 1,054 entries about 12-36 month-old children containing ten Q-CHAT-10 questions related to their behavior along with information about their age (in months), sex, jaundice, family background, and ASD diagnosis. Secondly, there is the Autism Screening for Children data set (Faizunnabi, 2019), accessible at https://www.kaggle.com/datasets/faizunnabi/autism-screening, which includes 1,054 entries about 4-11-year. The ASD categories were used as substitutes to indicate the probability of a developmental delay because of the clear clinical similarities between autism spect rum disorder and delays in speaking ability. Environmental factors not considered in t he original data, such as words uttered, screen time, parental engagement, hearing pr oblems, and bilingualism at home, were artificially included based on distribution par ameters determined clinically. A total of 300 new data points were produced utilizing an algorithm that assigned weighted risk scores, based on WHO and CDC criteria for speech milestones.

3.4 System Modelling Using UML

In this section, the Unified Modeling Language (UML) is utilized for developing visual models of the architecture of the suggested system. This will be achieved by pr esenting four different UML diagrams, which are explained below.

3.4.1 Use Case Diagram

The Use case diagram shows the following main actors: the User (caregivers, parents, or primary health care professionals) and the System (Next.js application and speech delay prediction engine). The use cases performed by the User are: Register/Login; Pe rform Speech Assessment; Submit for Analysis; See Results Dashboard; See Possible Causes and Action Plan; Download Report as a PDF file; Share on WhatsApp or Emai l; Manage Child Profiles; and See History of Assessments. On the other hand, the resp onsibilities of the System are: Validate Inputs; Perform ML Prediction; Save Results in MongoDB Atlas; and Provide Recommendations. The association among the followin g use cases is demonstrated: the submit use case and validate inputs/ML prediction use cases.

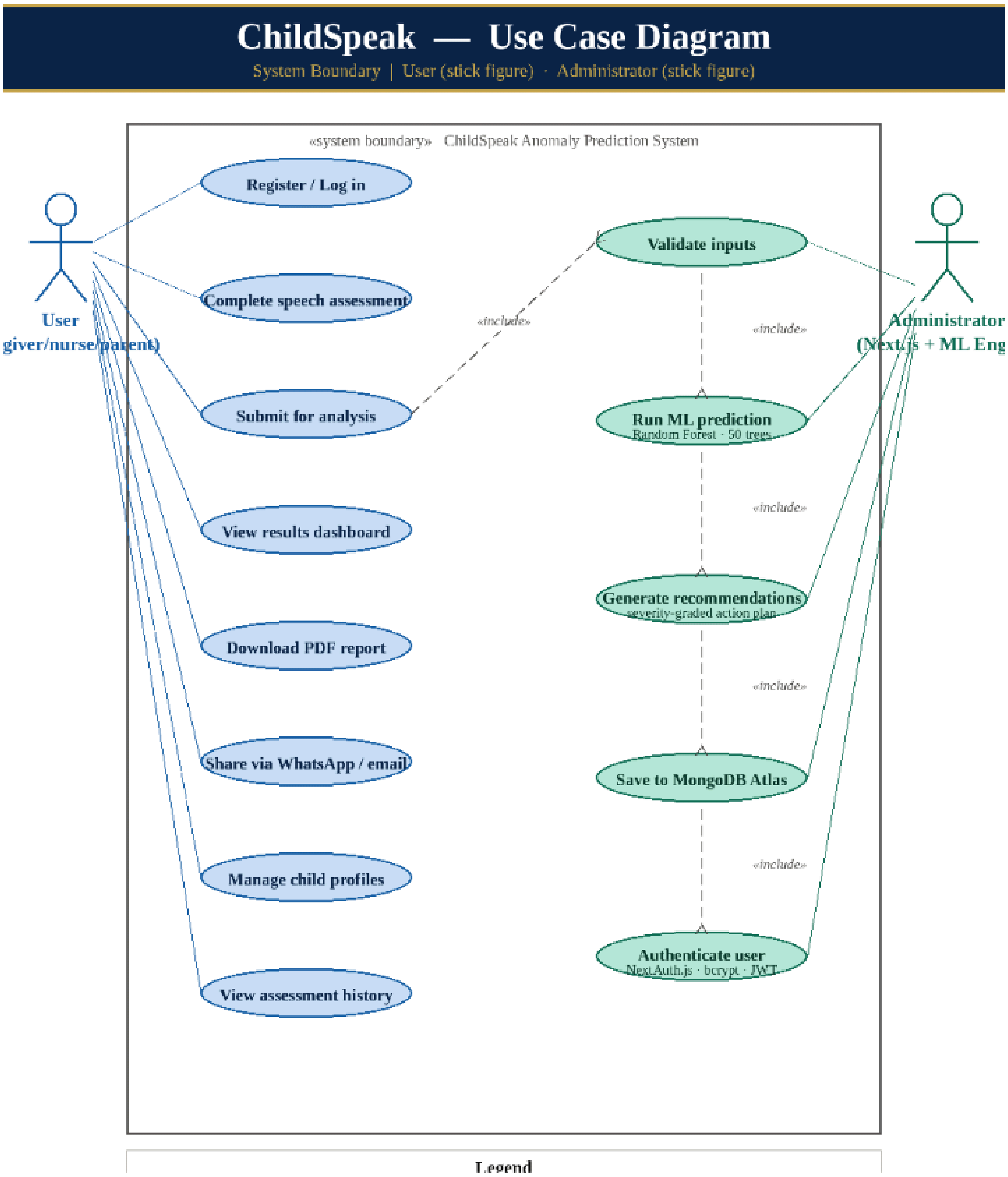

Figure 3.1: Use case diagram of the proposed system

3.4.2 Activity Diagram

The Activity Diagram illustrates the process of actions performed by the system in tw o swim lanes which are User and System. Actions performed by the user are: go to th e home page, sign-in/sign-up, answer 15 questions and press submit button. Decision diamonds illustrate the authentication and validation of the input, which have loops i n case of unsuccessful authentication and alternate path in case of invalid input. The process performed by the system includes: Input validation, loading the ML model (in

JSON format), normalising features using Z-score method, Decision Tree (50 decision trees with scores), classification with severity levels and recommendations, storing th e results into MongoDB atlas and return JSON response.

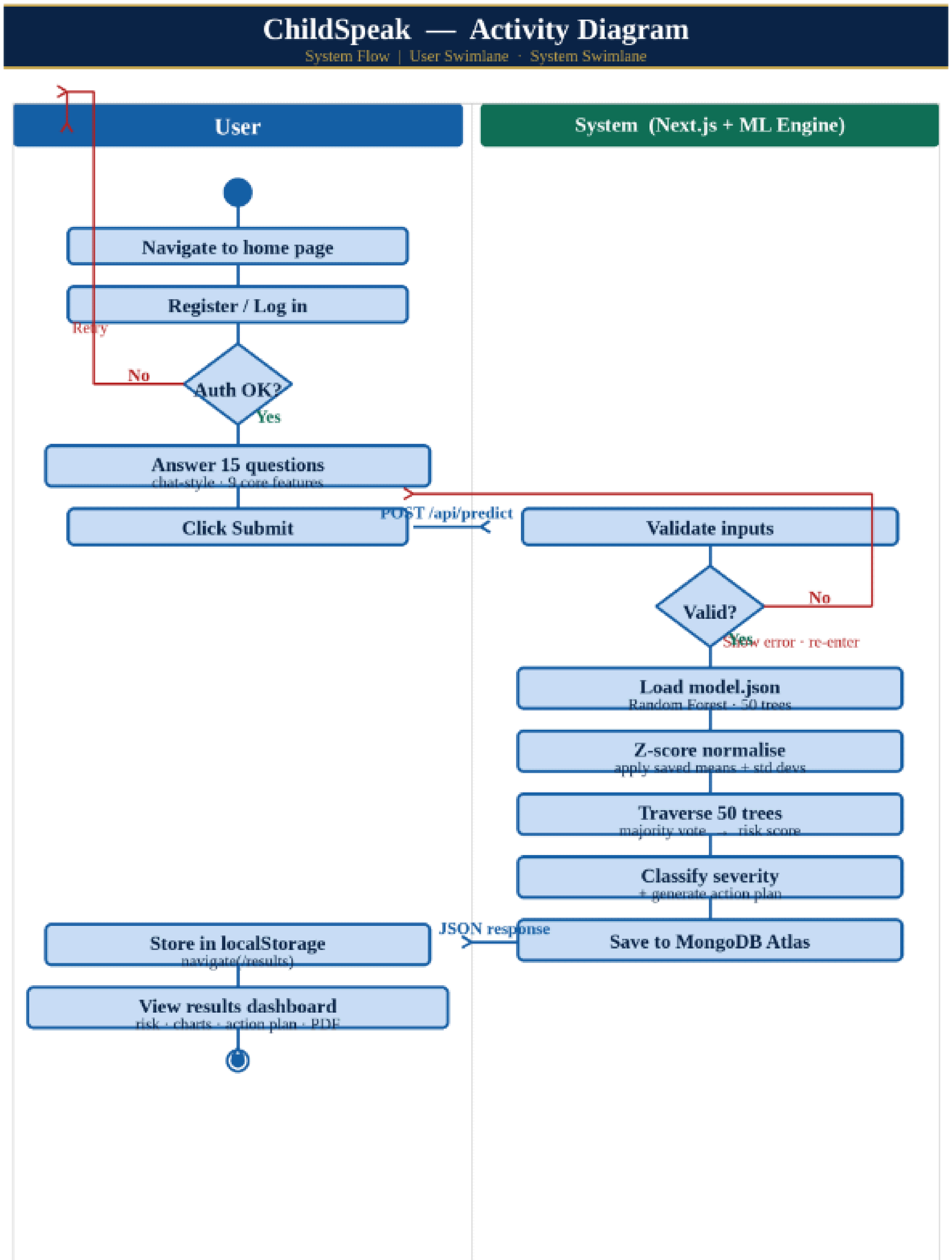

Figure 3.2: Activity diagram of the proposed system

3.4.3 Class Diagram

The class diagram highlights the major classes, their attributes, functions, and associations. The application classes consist of AssessmentForm, PredictionEngine, ResultsDashboard, PDFReport, and ShareResults classes. Mongoose schema classes include User, Assessment, and Child classes linked with ObjectId data type. The asso ciations indicate that AssessmentForm uses PredictionEngine using an HTTP request method, while PredictionEngine stores information on the Assessment schema using MongoDB Atlas. The association between User, Assessment, and Child indicates that User has one-to-many relationships with both Assessment and Child classes.

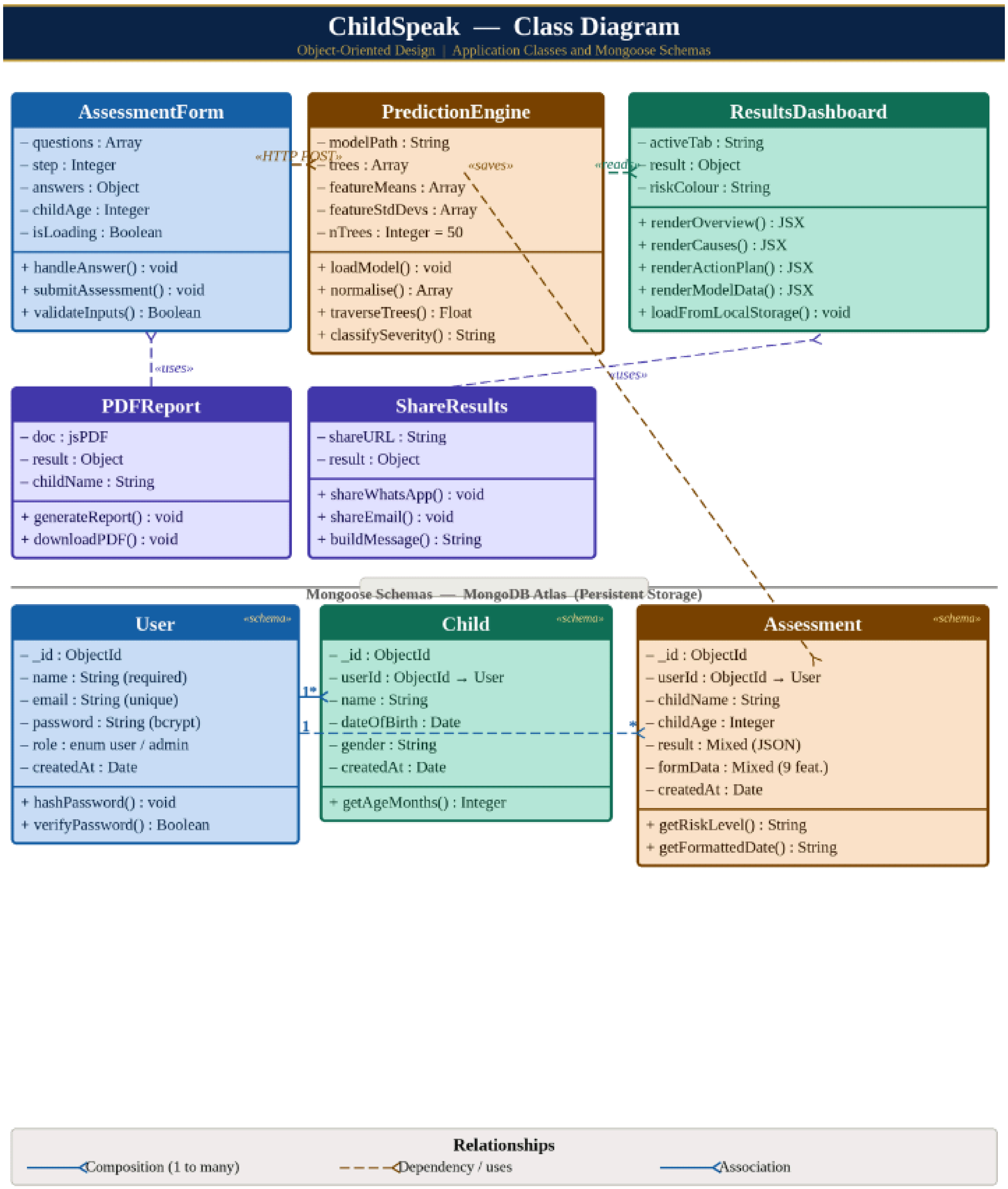

Figure 3.3: Class diagram of the proposed system

3.4.4 Sequence Diagram

The sequence diagram captures the flow of messages chronologically between six lifelines: User, Assessment Form, API Route (/api/predict), Speech Delay Predictor, MongoDB Atlas, and Results Dashboard in the span of 15 numbered messages starting from form submission until the rendering of results on the dashboard.

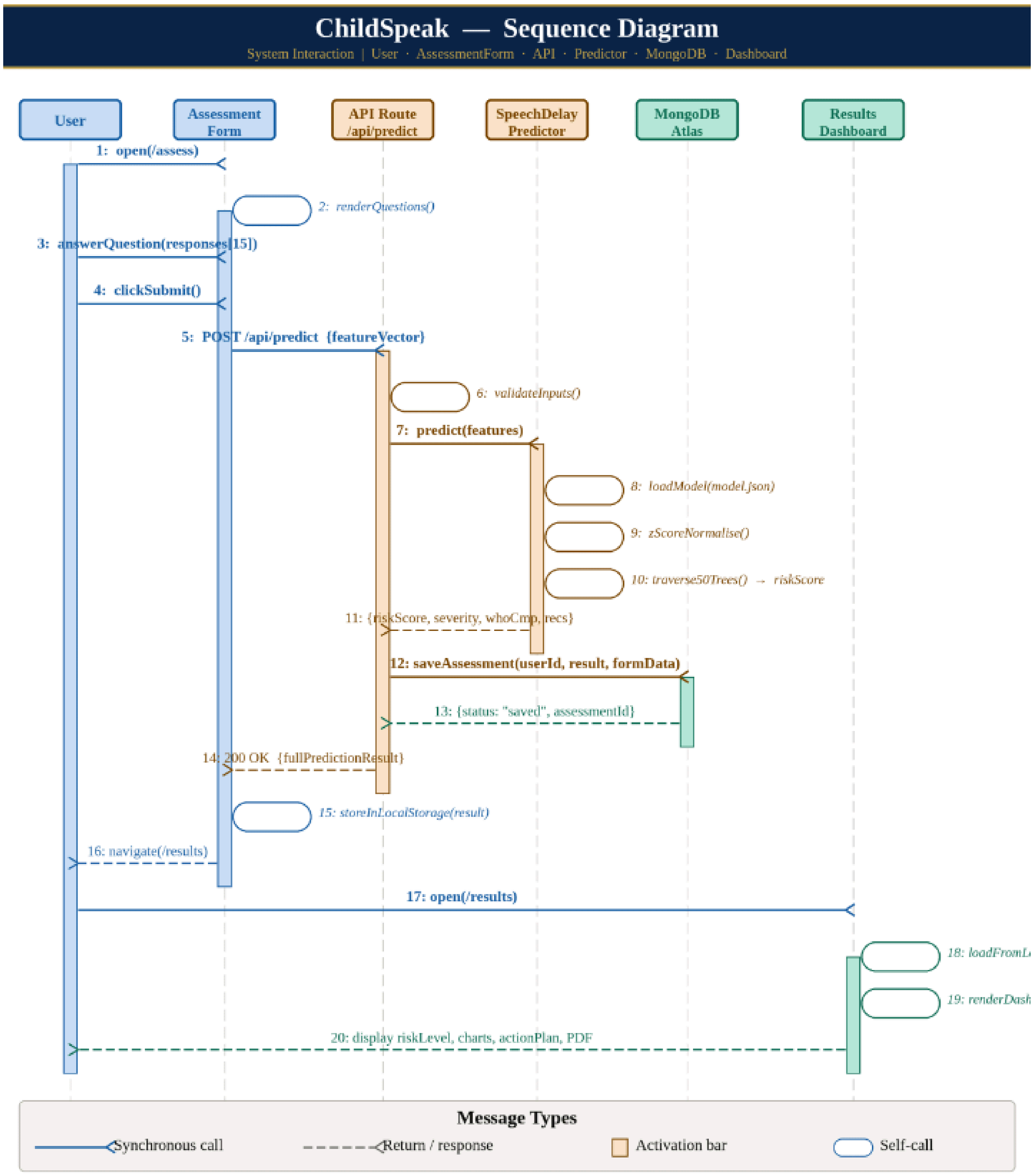

Figure 3.4: Sequence diagram of the proposed system

3.5 System Design

3.5.1 Input Design

The input interface consists of a conversational or chat-like speech assessment form which can be accessed from the /assess endpoint. The interface contains fifteen questions presented one by one in a messenger style interface with the doctor and user interactions appearing in chat bubbles. The question comes along with a soothing message about the clinical importance of that particular question. The output q uestion for the speech test is primarily the number of words the child says, and this is measured against the expected number of words according to the WHO speech milestones based on the age of the baby in months.

3.5.2 Output Design

This output interface consists of a results dashboard at the /results route that is made up of four tabs. In the Overview tab, one will find the risk banner (which has color coding according to different risk levels such as Normal - green, Low Risk - yellow, Moderate Risk – orange, and High Risk – red), probability score, Recharts Donut Chart, Recharts Bar Graph comparing the number of words said against the expected word count by the WHO milestone, the assessment summary, and personalized recommendations. Action plan provides a tab that contains the immediate actions categorized into different levels of severity and proven daily practices. The data model tab shows the four evaluation measures presented through a Recharts bar chart, a confusion matrix, and a clinical disclaimer. There are download and share buttons in the result tab header.

3.5.3 Database Design

MongoDB Atlas is being used with access via Mongoose. There are three schemas specified. The User schema has name(String, required), email(String, required, unique), password(String, hashed with bcrypt), role(String, enum[user/admin], default:user), and createdAt(Date). The Child schema has userId(ObjectId, referencing User), name, dateOfBirth, gender, and createdAt. The Assessment schema has userId(ObjectId, referencing User), childName, childAge, result(Mixed, full prediction response consisting of probability, severity, word count comparison, metrics, and recommendations), fo rmData(Mixed, consists of nine features), and createdAt. The browser localStorage is used as a temporary storage on the client side to pass the prediction result from the a ssessment form to the results dashboard.

3.5.4 System Architecture Design

The system architecture follows the three-tier web application design. Presenta tion layer is built with Next.js using React and Tailwind CSS; seven pages include Ho me, Assessment, Results, Children, History, Admin, Login/Signup. Application logic is made of Next.js API routes, which are: the main API to predict speech delay at src/ap p/api/predict/route.js, authentication APIs with NextAuth.js, sign up API, children & as sessments' CRUD APIs, and finally admin stats API. The third tier (machine learning model) is made of the serialized Random Forest classifier stored as src/ml_model/m odel.json and the training script stored at src/ml_model/train_model.js; both use two Kaggle CSV files together with the synthetic generation of WHO/CDC to form the hybr id training dataset of 2,408 entries. The data preprocessing pipeline incorporates feat ure augmentation, Z-score normalization using mean and standard deviation for each feature, shuffling, and 80-20 train-test split that produced 1,927 training samples and 481 test samples. The trained Random Forest classifier resulted in 98.96% accuracy, 98.90% precision, 95.74% recall, and 97.29% F1 score on the 481 samples in the test set, based on a confusion matrix of TN=386, FP=1, FN=4, and TP=90.

CHAPTER FOUR

SYSTEM IMPLEMENTATION

4.0 Introduction

This chapter offers an in-depth description of how the anomaly prediction mod el called ChildSpeak was implemented in order to determine whether a child who is be tween 12 to 60 months old experiences a speech delay. The implementation process uses the hybrid approach set forth in Chapter Three: the object-oriented analysis and design (OOAD) methodology applies to the application layer; on the other hand, the m achine learning development cycle operates under the CRISP-DM methodology, whos e importance and rationale in the context of the hybrid approach are stated in Section s 3.3.2 and 3.3.3 of Chapter Three, respectively. This chapter represents a complete i mplementation of the CRISP-DM architecture designed in Chapter Three - each of the phases of implementation is directly equivalent to its corresponding phase in CRISP-DM design as outlined in Section 3.3.4 of Chapter Three. This chapter covers informa tion related to the development environment, programming language, and developme nt tools utilized; CRISP-DM-based machine learning training pipeline design and devel opment of the random forest classifier; creation of the Next.js web app together with i ts six modules; testing processes carried out to ensure accuracy of the system; and evaluation results including interface screens and sample outputs. In the chapter, it has been illustrated how each one of the five specific objectives outlined in Chapter One was handled during the implementation stage, based on the system analysis, data set analysis, modeling approach, and UML models discussed in Chapter Three.

4.1 Development Environment and Tools

4.1.1 Programming Language and Libraries

All parts of the ChildSpeak system – namely, the machine learning model, the web app itself, the server-side API routes, and the database interactions – were coded in JavaScript, utilizing the Node.js runtime environment (v20). JavaScript was chosen as the sole programming language in which to implement the solution because JavaScript allows one to write all the code, including the machine learning logic, without relying on an extra programming language for that part. The Random Forest custom classifier can therefore be instantiated and used right in the server-side API route code during predictions.

Table 4.1: Programming Languages, Frameworks, and Libraries

Library / Framework	Version	Role in the System
JavaScript (Node.js)	20	Primary implementation language for all system components — ML model, API routes, and web application
Next.js	16.2.3	Full-stack framework providing React frontend, server-side rendering, App Router API routing, and Turbopack development server
React	18	Client-side component library for all user interface pages including assessment form, results dashboard, history, children, and admin
Tailwind CSS	3	Utility-first CSS framework for all component styling and responsive layout across desktop, tablet, and mobile

Library / Framework	Version	Role in the System
Recharts	2	React data visualisation library for word count charts, metrics bar charts, risk donut chart, and progress line charts on the history page
NextAuth.js	Latest	Authentication — email and password credential login, JWT session management, and protected route middleware at middleware.js
Mongoose	Latest	Object-document mapping — defines User, Child, and Assessment schemas and manages all MongoDB Atlas database operations
bcryptjs	Latest	Password hashing before storage in MongoDB Atlas; password verification during login authentication
jsPDF	Latest	PDF generation implementing the downloadable speech assessment report feature on the results dashboard page
Axios	Latest	HTTP client dispatching POST requests from the assessment form component to the /api/predict API route
Node.js fs / path	Built-in	File system access — loads the trained model.json file from disk within the API route at prediction time

4.1.2 Development Platform and Hardware Requirements

The model was created on a PC running Windows OS, using VS Code as the integrated development environment. VS Code was chosen due to its complete JavaScript and Next.js integration, built-in terminal for executing scripts to train the model as well as the development server, Git repository integration, and availability of extension s that provide syntax highlighting, code completion, and code linting capabilities. The development server was run locally using the command npm run dev and accessed o n http://localhost:3000. The database service utilized is MongoDB Atlas, thus avoidin g the need to install a database server within the local computer. Neither a GPU nor a ny other dedicated hardware is necessary for performing the model training because t he custom implementation of the Random Forest algorithm uses CPU only. The traini ng process for the entire data set takes between three and five minutes.

4.2 Dataset Description and Preprocessing (CRISP-DM Phase 3: Data Preparation)

As outlined in Chapter Three, the training set is a mixed dataset that comprises 2,408 labeled records from three different sources, which include: the Kaggle dataset of Autism Screening for Toddlers with 1,054 records (data/toddler_autism.csv), the K aggle dataset of Autism Screening for Children with 1,054 records (data/children_auti sm.csv), and 300 synthetic records produced by the training script (src/ml_model/trai n_model.js) using a seed random number generator (seed=42) based on WHO and C DC speech milestones for 12 to 60-month-old children. Class balance results in 553 d elayed (22.9%) and 1,855 normal (77.1%) cases. Each record has the nine predictor v ariables and one target variable that takes the value 0 for normal and 1 for delayed. T he entire preprocessing process involved four steps that were carried out sequentially using JavaScript inside the training script. First, both Kaggle CSV data sets were pars ed and processed using source specific functions that converted the original columns into the standard nine-feature form and added five features which were missing usin g distribution values from clinical studies. The second step involved merging all three processed arrays using the spread operator into a single array containing 2,408 data sets. In the third step, z-normalization was done on all nine columns of features and n ine calculated means and standard deviations of features were stored inside the mod el.json file to be used during prediction.

4.3 Model Architecture and Training (CRISP-DM Phase 4: Modelling)

The algorithm for predicting the onset of delayed speech was developed as a c ustom implementation of a Random Forest model written in JavaScript and consistin g of two classes – DecisionTree and RandomForest, located in src/ml_model/train_m odel.js. The class DecisionTree is responsible for creating a decision tree that uses Gi ni impurity as an impurity metric to split nodes. For finding the best split, the bestSplit function searches through all available splits over all nine features and calculates the weighted Gini impurity. The algorithm has three stop criteria: maximum depth of eigh t levels, minimum samples of five records, and pure nodes. The RandomForest object builds 50 trees, using as input bootstrap samples of 1,927 instances each. When pred icting, each of the 50 trees makes its own vote; if there is more than one vote in favor of class 1, then the outcome is 1, whereas the ratio of trees voting for class 1 gives us the probability score shown on the dashboard.

Table 4.2: Random Forest Training Parameters

Parameter	Setting	Justification
Number of trees (nTree	50	Provides a stable ensemble on the 1,927-re

Parameter	Setting	Justification
s)		cord training set without excessive computation time
Maximum tree depth (maxDepth)	8 levels	Limits individual tree complexity to reduce overfitting to the training data
Minimum samples per leaf (minSamples)	5 records	Prevents leaf nodes from being defined by too few training examples
Split criterion	Gini impurity	Standard criterion for binary classification — measures class impurity at each node
Bootstrap sampling	With replacement	Introduces diversity among the 50 trees, reducing ensemble variance
Feature normalisation	Z-score (mean=0, std=1)	Ensures all 9 features contribute on a comparable numerical scale
Train/test split	80% / 20%	1,927 training records; 481 held-out test records
Random seed	42	Ensures full reproducibility of synthetic data generation and shuffle
Implementation	JavaScript (Node.js 20)	Enables direct integration into Next.js API route without external ML serving infrastructure

The training script is run using the command node src/ml_model/train_model.js. This outputs status messages confirming that the CSV files from Kaggle have been loaded, a total of 2,408 data items, out of which 553 have been delayed while 1,855 were normally delivered, and 1,927 training instances and 481 test instances. During execution, it generates output for each tree generated, in intervals of ten trees. After ru nning to completion, it outputs a model.json file containing all 50 decision tree structures along with the means and standard deviation of nine features, and the names of the keys.

4.4 System Implementation and Modules (CRISP-DM Phase 6: Deployment)

The development of the ChildSpeak system was in terms of a Next.js web application that had six main modules associated with the functional requirements discussed in Chapter Three Section 3.2.1, which correspond with CRISP-DM Phase 6 (Deployment) as outlined in Chapter Three Section 3.3.4, and described above in Section 4.1. Each of these modules is discussed below, in relation to the class diagram (Figure 3.3) and sequence diagram (Figure 3.4) in Chapter Three.

4.4.1 Prediction API Route (src/app/api/predict/route.js)

The prediction API route is the key integration component, mapped to the PredictionEngine class of Figure 3.3 and the API Route lifeline in Figure 3.4. Upon receiving the HTTP POST request from the assessment form to /api/predict, the prediction API route extracts the nine feature values from the HTTP request payload; checks for presence of all required parameters, returning HTTP 4the nine input values utilizing the saved feature means and standard deviations; iterates through all 50 decision trees to get the binary prediction and probability; calculates the severity of the speech delay risk depending on probability ranges (less than 0.25 - green/Normal, between 0.25 - 0.49 - yellow/Low Risk, between 0.50 - 0.74 - orange/Moderate Risk, greater than 0.75 - red/High Risk); calculates the number of expected words based on the WHO milestones at the specific age of the child in months; creates customized clinical advice; saves the results of the assessment into MongoDB Atlas if user is authorized via the Assessm ent Mongoose schema; and responds with the result via JSON HTTP 200 response.

4.4.2 Assessment Form (src/app/assess/page.jsx)

The assessment form is related to the AssessmentForm class shown in Figure 3.3. The assessment form is a client side React component which renders the chat-lik e conversation interface of speech test. This component maintains the following stat es: questions (the 15 question objects including questions text, calming text, input typ e, and possible answers); step (the integer representing the currently displayed questi on from 0 to 14); answers (an object storing all entered answers under feature name k ey); and isLoading. Numerical questions contain an input box for numbers along with a Next button, while choice questions display radio buttons that have labels. There is a progress bar at the top showing the level of progress. After clicking the Submit butt on, the submitAssessment() method maps the answers received into the feature para meters, sends a POST request through Axios to /api/predict, stores the response in lo calStorage as JSON under the key childspeak_result, and redirects to the results page using the Next.js router.

4.4.3 Results Dashboard (src/app/results/page.jsx)

The result dashboard is related to the ResultsDashboard component shown in Figure 3.3. When mounted, it loads the result from localStorage and creates the four-t abbed result dashboard according to the guidelines given in Section 3.5.2. In addition, the overview tab displays the result in the form of a risk banner containing a colour-c oded severity and probability percentage. It further includes the donut graph showing the probability score using the Recharts graphing library. This is followed by the bar g raph showing the comparison between the number of words said with WHO mileston es word count. Under the Action Plan tab, severity-based immediate actions and daily habits are displayed based on the clinical research found in chapter two. The Model Data tab displays the evaluation criteria as a Recharts bar chart, a color-coded confu sion matrix table, and the disclaimer message. The page also displays the PDFReport component that uses jsPDF to generate report documents as well as ShareResults co mponent for sharing results via WhatsApp, e-mail, and clipboard.

4.4.4 Authentication Module

The authentication process will be done using the NextAuth.js configuration d one at the src/lib/auth.js file. The Credentials provider takes the email and password fields, makes an inquiry on MongoDB Atlas for a corresponding user record using Mo ngoose, and compares the entered password using bcryptjs. Upon a successful valid ation process, the JWT session token is saved within an HTTP only cookie containing the user ID, username, email, and the user's role. The /assess, /results, /children, /hist ory, /profile, and /admin pages are protected with middleware.js and any unauthorize d user will be redirected to the /login page. On the other hand, the route.js file of the sr c/app/api/auth/signup/ folder takes input values from the frontend, verifies if there i s already another account with the same email address, encrypts the password using bcryptjs with salt round 12, and saves a new document in MongoDB Atlas database.

4.4.5 Child Profiles and Assessment History

Authenticated users can generate many children profiles through the child prof iles page located at src/app/children/page.jsx where they can fill their child's name, d ate of birth, and gender details. The children profiles are saved in the database as Chi ld documents in MongoDB Atlas connected with the authenticated user's ObjectId. It also shows the profile cards that include child's name, gender, and age in months. All the previously undertaken assessments for the currently signed-in user are fetched on the Assessment History Page available at src/app/history/page.jsx, which displays th e assessments in chronological order by rendering the cards that contain risk level ba dge, risk probability score, age of the child at time of assessment, words spoken by th e child, and date of assessment. When two or more assessments have been made on a child, the Recharts LineChart shows the trend of risk probability over time.

4.4.6 Administrative Panel (src/app/admin/page.jsx)

The Admin Panel is only accessible to users with the User document in the Mo ngoDB Atlas where role=admin. At the load time, it makes an API call to the /api/adm in/stats end-point which checks for admin role and performs calculations using Mon goose aggregate pipelines such as the total number of users, total assessments mad e, number of risky assessments made, top 10 latest user creation and assessments, n umber of assessments made each day of last 7 days, and risk levels of all 4 types. Th e component creates three tabs that include Overview (four summary statistic cards, Recharts bar graph of daily assessment numbers, Recharts pie graph of risk level dist ribution, and model info panel showing dataset summary and evaluation metrics); Us ers (table showing the ten most recent users along with their name, email address, rol e badge, and registration date); and Assessments (table showing the ten most recent assessments along with the name, child’s name, age, risk level badge, probability, and date).

4.5 Testing and Evaluation (CRISP-DM Phase 5: Evaluation)

4.5.1 Evaluation Metrics

The results obtained from the trained random forest model predicting speech delay were subjected to four different performance metrics for binary classification, which include accuracy (fraction of total records correctly classified by the model), precision (fraction of records predicted to be speech delayed that are actually speech delayed), sensitivity (fraction of actual speech delayed records correctly classified by the model), and F1 score (harmonic average of precision and recall). These performance metrics measure the effectiveness of the model not only in terms of its false-positive rate but also its false-negative rate, which is crucial in the clinical sense because false negatives would lead to the unnecessary referral of children without speech delay.

4.5.2 System Testing

The testing process for End-to-End was performed via running the entire application locally, where each scenario was tested fully. The first scenario was testing a new user signup, where valid and invalid inputs were used; the application successfully validated all inputs, hash the password and saved the User document into MongoDB Atlas before redirecting the user to the Login Page. The second scenario involved testing authentication; correct credentials gave a valid NextAuth.js session, while incorrect credentials did not create a session. Thirdly, the process of assessment from start to end was completed by responding to all 15 questions in order, the progress bar was working fine, soothing messages were shown for each question type, upon submitting the POST request was fired, and after that, the results page loaded in three seconds. Fourthly, the results page: all components of the risk banner color, probability score, wo rds graph, causes tab, action plan tab, and model data tab were displaying properly with various combinations of inputs. Fifthly, PDF report: was created and was checked to have all correct information regarding the assessment results, risk level, action plan and model metrics. Lastly, responsive design: testing performed in four different resolutions desktop (1920x1080), laptop (1366x768), tablet, and mobile (390x844).

4.5.3 User Interface Testing and Walkthrough

User interface testing has been conducted via navigation through all pages with the aim of ensuring that all features work correctly. Home page renders with a description of the system's goal, an action button that leads the user to the assessment page, and statistics on model accuracy. Signup and login pages display with appropriate form validation and error handling. The assessment page renders with the conversation interface with the doctor character, chat bubbles, soothing messages, and progress bar. The results page displays all four tabs along with the risk banner, Recharts graphs, possible causes, actions plan, and the model data table. The children page displays the child profile cards along with the add-child form. The history page displays assessment cards along with the Recharts progress line graph. The admin page displays the three tabs along with statistics and data tables. All components were found to render as expected and work as defined by Section 3.2 of the requirements.

4.6 Results Presentation and Analysis

4.6.1 Output Samples and Interface Screens

ChildSpeak model was evaluated from end to end with a sample scenario involving a 24-month-old boy who uttered 8 words out of an estimated 100 words according to WHO standards, with four and half hours screen time on average, 1.2 hours par ent contact every day, no hearing problems, no early deliveries, family history of delayed speech development, and a non-bilingual household. The result given by the model was 87% probability and high risk with a red banner severity indicator. In the WHO Comparison Chart, only 8 words were listed compared to the anticipated 100 words, indicating that there was a shortfall of 92 words. From the Possible Causes tab, the four factors found responsible for this condition included little parent interaction, too much exposure to screens, less number of words than milestone words, and family history. The Action Plan tab suggested the following quick actions:

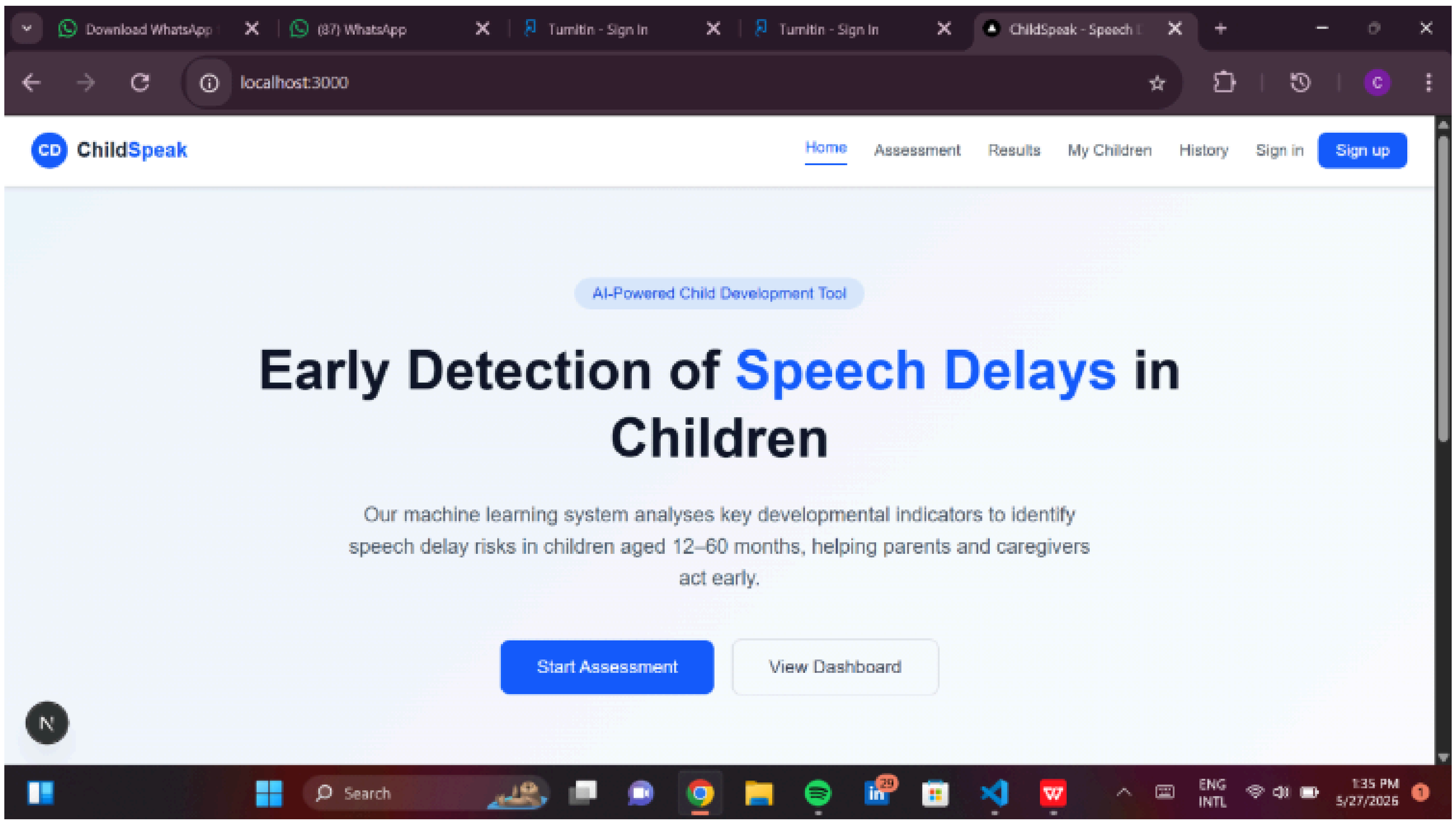

Figure 4.1: Home page of the ChildSpeak system

The homepage of ChildSpeak system is shown in figure 4.1, which can be accessed from root (/). The page provides information on the purpose of the system along with important statistics about accuracy of the models (98.96%) and features a prominent call to action button for navigating to speech assessment section. Navigation bar contains login and sign up options for un-authenticated users.