

Ghulam Ishaq Khan Institute of Engineering Sciences and Technology
Topi, Pakistan

Faculty of Computer Science and Engineering

AID: AI-Driven Pre-Screening of Autism in Children

Kaneez e Zahra	**2022245**
Umer Khalid	**2022605**
Wardah Haya	**2022622**

Advisor:

Engr. Dr. Shahabuddin Ansari Saeed

Co – advisor:

Dr. Muhammad Omer Bin

Dr. Khurram Khan Jadoon

Certificate of Approval

It is certified that the work presented in this report was performed by **Umer Khalid , Wardah Haya,** and **Kaneez e Zahra** under the supervision of **Dr. Shahab Uddin Ansari.** The work is adequate and lies within the scope of the BS degree in Data Science/Artificial Intelligence at Ghulam Ishaq Khan Institute of Engineering Sciences and Technology.

---------- ----------

Dr.Shahab Uddin Ansari **Dr. Muhammad Omer Bin Saeed**

Dr. Khurram Khan Jadoon

(Advisor) (Co-Advisor)

Dr. Qadeer ul Hassan

(Dean)

SUSTAINABLE DEVELOPMENT GOALS (SDGs) ALIGNMENT

Project AID, which is the application of artificial intelligence in autism spectrum disorder detection, falls under various United Nations sustainable development goals (SDGs). This will be an indicator of how artificial intelligence can help solve issues in the healthcare sector.

Good health and well-being (SDG 3)

By working on this project, we were able to contribute in the realization of the third SDG, which aims to ensure good health and well-being of all people at all ages. Through screening and the early diagnosis of ASD in children, families have been able to receive professional guidance in their approach towards their children's condition. By implementing an AI application that relies on questionnaires for one to run these screenings on, the time taken to conduct these initial tests has greatly reduced. This could particularly be beneficial to areas that do not possess a large number of specialist doctors. With such quick feedback, children will receive the necessary stimulation while being at an advantage of having prompt information.

Decent Work and Economic Growth (SDG 8)

This solution can contribute to Goal 8: "Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all" by automating the early stages of the autism screening.

Health workers will not have to solely rely on manual screening to identify potential candidates, and will be able to make those cases a priority. This improves the overall productivity of the clinical workflow, since the specialists can focus more on diagnosing and treating patients. Otherwise, solutions like this are a new way to contribute to the growing AI sector of healthcare, creating job opportunities in fields such as machine learning, healthcare systems, medical advancements, etc.

Reduced Inequalities (SDG 10)

Through the project, the use of technology will help achieve Goal 10, the reduction of inequality in countries and among countries.

There is a shortage of trained staff for autism screening for majority of the community, particularly in remote and less developed countries.

This project provides easy to use online screening tools for any parents, even those with low IT literacy skills, to screen their children for autism.

Through lowered cost of the initial screening and heightened awareness, the project will help ensure more children can be screened, no matter where they are.

---------- ----------

Dr.Shahab Uddin Ansari
Saeed

(Advisor)

Dr. Muhammad Omer Bin

Dr. Khurram Khan Jadoon

(Co-Advisor)

ABSTRACT

Early diagnosis of autism spectrum disorder (ASD) is extremely critical for developmental gains, yet it is still one of the most difficult diagnostic tasks due to poor awareness, delayed diagnosis and lack of specialist availability. While accurate screening of ASD is important, standard methods require considerable time and resources and are inaccessible to every family. Our team is designing a smart AI based pre-screening tool that aid in early identification of risk for autism with a combination of structured questionnaire assessment and machine learning approaches. We are using behavioral features (e.g. Q-CHAT response and other demographic features) processed using several models to predict risk factors. Furthermore, the system also includes an image based screening solution with automated report generation which aims to provide combined information through the prediction scores and natural language explanations. Real and synthetic data is pre-processed in order to allow the system to be able to be robust to any inputs and scalable to any volume and scale. We also aim to provide easily understandable reports and optional reports to help in user interpretations without involving the understanding of complex parameters. The web-based and highly usable interface of the tool aims to allow its application in non clinical environments for early screening of autism risk. This project aims to bridge the gap between the research conducted on artificial intelligence and the demand in the healthcare system to facilitate the use of such technology to make early identification a success.

ACKNOWLEDGEMENTS

We thank our advisors **Dr. Shahab Uddin Ansari, Dr. Omer Bin Saeed and Dr Khurram Jadoon**, for providing us with utmost support and guidance. They played a critical role in the success of our efforts. We thank the faculty and the institute for providing us with the best educational opportunities. This project would not be possible without the help of various other professors, and researchers across the globe, who paved the way for us to be able to create this project.

TABLE OF CONTENTS

SUSTAINABLE DEVELOPMENT GOALS (SDGs) ALIGNMENT 3
Decent Work and Economic Growth (SDG 8) 3
CHAPTER I 13
INTRODUCTION 13
History 13
Motivation 13
Product Objective 14
Product Perspective 15
CHAPTER II 16
LITERATURE SURVEY 16
LITERATURE REVIEW 16
ACADEMIC RESEARCH 16
EXISTING SYSTEMS 17
CHAPTER III 19
DESIGN 19
3.1. Model Description 19
3.2. System Requirements 20
3.2.4. Communication Interfaces 20
3.3. Software Functionality 21
3.4. User Characteristics 22
3.5. Constraints 24
3.6. Assumptions and Dependencies 25
3.7. Functional Requirements 26
FR-4: Multilingual LLM Narrative Report 26
FR-6: Data Privacy and Stateless Sessions 27
3.7. Non-functional Requirements and Software System Attributes 30
3.7.1. Performance Requirements 30
3.7.3. Security 31
3.8. Project Design / Architecture 31
CHAPTER IV 41
PROPOSED SOLUTION 41
4.1. Methodology 41
4.1.1. Two-Stage Gated Architecture Design 41
4.1.2. CNN Backbone Selection and Calibration 42
4.1.3. Questionnaire Feature Engineering 42
4.1.4. Tabular Classifier Selection, Calibration, and Threshold Tuning 42
4.1.5. LLM Prompt Engineering 43

4.1.6. Iterative Development Process 43
4.2. Implementation 43
4.2.1. System Initialisation 43
4.2.2. Stage 1 — CNN Face Screening 43
4.2.3. Stage 2 — Questionnaire and Risk Classification 44
4.2.4. LLM Narrative Generation 44
4.2.5. PDF Export 44
4.2.6. Web Application and Firebase Integration 44
4.2.7. Communication and Security 45
4.3. Competitive Advantages Over Existing Solutions 45
CHAPTER V 47
RESULTS AND DISCUSSION 47
5.1. Result 47
5.1.1. CNN (Face) Model Results 47
5.1.2 ML (Questionnaire) Model Results 47
5.2 Discussion 47
5.2.1. CNN (Face) Model Performance: 47
5.2.2. ML (Questionnaire) Model Performance: 48
5.2.3. Real-World Applicability: 48
5.2.4. Limitations and Future Work: 48
CHAPTER VI 51
CONCLUSION AND FUTURE WORK 51
GLOSSARY 52
REFERENCES 53
APPENDIX A 54
Data Survey: 54
Code: 54

LIST OF FIGURES

Figure 1: Architecture View 28
Figure 2: Use Case 29
Figure 3: Physical View 30
Figure 5-7:UI Design 31

LIST OF TABLES

Table 1: Functional Requirement 1 20
Table 2: Functional Requirement 2 21
Table 3: Functional Requirement 3 22
Table 4:Functional Requirement 4 23
Table 5:Glossary 45

CHAPTER I

INTRODUCTION

History

Autism diagnosis and screening has advanced throughout the last couple of decades from direct observational diagnostics, through more systematic screening and most recently to AI detection of Autism Spectrum Disorder (ASD). Historically, direct observations from doctors and experts, combined with reports from parents and diagnostic tools like Q-CHAT and ADOS were used to diagnose Autism Spectrum Disorder, which can be identified by difficulties in social interaction and communication, and stereotypic behavior. While accurate, this diagnosis is based on subjective observations which is time consuming and needs to be interpreted by doctors and experts. However, numerous studies show that there is a prevalence of children who are diagnosed with autism at or after the age of 4 (although signs of Autism appear earlier) but actually presented earlier. Furthermore, approximately 300,000 children in Pakistan could be living with Autism, but many remain undiagnosed as many children may not get a diagnosis because of a lack of expert clinicians. It is also reported that many parents are unable to diagnose children with autism in the absence of clinical assessment as many are unaware of the characteristic early signs in resource-poor environments. Diagnosis may need numerous visits to specialists over weeks and months before a decision can be made.

Motivation

The justification for this project comes from the drive to make early autism screening more accurate, efficient and readily available through artificial intelligence (AI) technologies. Recent studies have indicated that AI models, particularly in machine learning and predictive analysis, are capable of high accuracy in classifications and risk-assessment within healthcare contexts, but similar accessible AI solutions for early autism screening which includes questionnaire analysis and automated reporting have yet to be fully implemented. Current trends in reality are showcasing the power of AI within the medical sector by reducing process time, streamlining the workflow and creating significant improvements in various diagnostic scenarios where the process has been automated. Through the automation of Q-CHAT analysis, risk classification, and report generation our project will make early autism screening quicker, decrease human error and provide an automated initial prediction, thereby helping parents and caregivers to understand the condition, provide health professionals with readily available information and provide a rapid, reliable and convenient form of early assessment for the risk of autism.

PURPOSE

The overall goal of the AID project is to develop an AI system that can screen for early autism risk, classify risk, and generate an automated report based on a set of structured behavioral data and learning algorithms. Through academic guidance and standardized screening questionnaires (e.g. Q-CHAT), we have been using both real and synthesized data for a more robust, accurate and applicable system. The aim is to alleviate work for those doing manual screenings, promote uniformity in initial screening, and increase the accessibility of early screening, ultimately aiding parents and professionals in identifying and intervening with potential early signs of autism at an earlier, more appropriate age.

DESCRIPTION

The Autism Screening System project is an intelligent system that facilitates early detection of the risk of developing autism. It uses a mix of machine learning and deep learning approaches to provide the following functionality:

1. Processing questionnaire-based input.
2. Facial inputs for screening purposes.
3. Predictions of Autism risk levels.
4. Reporting that uses an LLM-based pipeline and gives detailed output.

The system includes an interactive system for user input, image screening, and visualization of the reports.

Product Objective

The objective of the project "Autism Screening System" is to automate the screening and prediction of autism risk from questionnaire answers and face screening data. The system must be able to process user inputs accurately and identify individual facial features for risk classification according to their defined thresholds, with a variation in data and/or conditions. It must ensure speedy and reliable results that aids care takers/professionals for early interventions.

The key objectives we aim to achieve through this project are:

1. **Automation of Data Processing and Risk Prediction** to minimize manual effort.
2. **Improvement in Early Screening and Assessment** through reliable machine learning and deep learning models
3. **Reduction of Human Error** in diagnostic procedures through AI assistance.
4. **Support for Caregivers and Practitioners** by providing clear and structured reports **Innovation in Dental AI Applications**, paving the way for broader adoption of intelligent diagnostic tools in dentistry.

Product Perspective

The Autism Screening System is an independent intelligent application that can be used in existing healthcare and screening procedures and in conjunction with digital input systems such as questionnaire responses and facial image captures. It acts as an automatic system to process the results from the screening inputs and deals with behavioral response processing, facial screening and risk classification. The key stakeholders in the project are:

- **Caregivers and Health Professionals** (primary end-users) who require accurate, fast, and consistent diagnostic support for treatment planning.
- **Developers and AI Engineers** who are responsible for building, maintaining, and updating the system to ensure its reliability and scalability.
- **Patients**, indirectly, by benefiting from faster diagnoses and more accurate treatment plans.

The system emphasizes a **user-friendly interface** that allows dental professionals to easily upload patient X-rays, view segmentation and classification results.Overall, the *Molar Angulation* system acts as an intelligent assistant that enhances the quality of dental care, reduces diagnostic errors, saves time, and promotes the integration of AI technologies into mainstream dental practice.

CHAPTER II

LITERATURE SURVEY

LITERATURE REVIEW

In the past ten years, artificial intelligence (AI) has begun to play a role in autism screening and diagnostics, presenting new ways to analyze behavioural and clinical data. AI-powered autism detection systems have already revealed considerable promise in early childhood screening and in neurodevelopmental diagnostics. Various reports reveal that ML techniques, including the use of neural networks and regression methods, significantly surpass human manual tests in terms of detecting patterns associated with autism or indicators of risk.

Various ML techniques, including LASSO, RIDGE regression, and Random Forest, were used in prior attempts at identifying autistic individuals or detecting autism-risk and these techniques often produce promising results when working on classification and regression tasks. Moreover, using the available datasets for training ML-based models that include developmental and clinical characteristics also proved useful and reduces the traditional autism screening time. With increasing sophistication, AI systems in autism research are starting to focus on multimodal analyses, natural language processing, facial behaviour recognition etc. However, there is still lack of a fully automated autism screening tool that also works with facial screens and can present full clinical report. Herein we present our approach of combining a questionnaire-based screening tool with facial screens to generate an overall diagnosis report using the advantages of a user-friendly and real-time AI system.

ACADEMIC RESEARCH

➢ **Paper: "Advances in Autism Spectrum Disorder (ASD) Diagnostics: From Theoretical Frameworks to AI-Driven Innovations"**
Published:2025
Authors: Syriopoulou-Delli, C.
Key Findings:
This project examined the development of Autism diagnostics, moving from the traditional theoretical models of the past, towards present day AI based technologies. The paper demonstrated the application of AI, through techniques like video analysis, NLP and multimodal integration, in facilitating early detection, diagnosis accuracy and minimizing reliance on subjective clinical judgement.
Source: https://doi.org/10.3390/electronics14050951

➢ **Paper: "A Machine Learning Approach to the Diagnosis of Autism Spectrum Disorder and Multi-Systemic Developmental Disorder Based on Retrospective Data and ADOS-2 Score"**
Published:2023
Authors: Briguglio, M., et al.
Key Findings:
Overall, our machine learning approaches (RIDGE regression, LASSO and Random Forest) were able to effectively predict autism outcomes in the classification and regression models based on clinical and developmental factors. The performances of our classification and regression models were promising in determining significant predictors (e.g.sleep problems, gut problem, development problems).
Source: https://pubmed.ncbi.nlm.nih.gov/37371363/

➢ **Paper: "Real-time Facial Emotion Recognition System Among Children with Autism Based on Deep Learning and IoT"**
Published: 2023
Authors: Talaat, F. M.
Key Findings:
The presented work suggests a real-time emotion recognition system for children with autism by using deep convolutional neural network (CNN) combined with the Internet of Things (IoT) and fog computing. The system aims to conduct face detection, feature extraction and emotion classification (six categories: anger, fear, joy, natural, sad and surprise) with extremely high accuracy (up to 99.99%). This also indicates the significance of real-time facial analysis to help diagnose and care for autistic children more quickly and efficiently.
Source: https://rdcu.be/ferSn

➢ **Paper: " Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children"**
Published:2024
Authors: Talaat, F. M., Ali, Z. H., Mostafa, R. R., et al.
Key Findings:
Proposed a real time facial emotion recognition model for autistic children with the approach of using kernel autoencoder and CNN, and also claiming of higher performance on emotion classification; but the paper was later withdrawn as it contained serious ethical issues related to usage of children's images with no prior verification and permission.
Source: https://rdcu.be/fer7h

EXISTING SYSTEMS

➢ **Cognoa "Canvas Dx" AI System**
Description:
Canvas Dx is an FDA cleared, AI driven diagnostic assistance tool designed for ASD (Autism Spectrum Disorder) in children aged between 18 months and 5 years old. It relies on the use of machine learning with parent-reported behavioral data and clinician feedback in order to detect and diagnose ASD at an early stage. Unlike standard screening devices, it is one of the first devices approved to act as a clinical aid for diagnosing autism.

Source: Canvas Dx for Highmark Families - Cognoa

➢ **EarliPoint Evaluation System (EarliTec Diagnostics)**
Description:
EarliPoint is an FDA-cleared eye-tracking diagnostic instrument that assesses how toddlers observe social stimuli (faces, objects, and interactions). An AI is utilized to interpret patterns of gazes, which can then be evaluated for indications of Autism Spectrum Disorder. EarliPoint is indicated for use with infants aged 16-30 months and is a one of the most current early autism detection devices.

Source: EarliPoint Health

➢ **Autism & Beyond (Duke University + Apple Research Kit)**
Description:
The Autism & Beyond app is mobile application research which is built on the foundation of Apple's ResearchKit. It is designed by Duke University to provide an assessment of early markers of autism in children using front camera and computer based analysis of facial expressions and response behaviors. The underlying approach includes a machine learning algorithm applied on emotion and attention states which bears similarities to facial emotion recognition in computer based research of autism.

Source: Autism & Beyond - Duke Institute for Health Innovation

CHAPTER III

DESIGN

The design phase establishes a structured, scalable approach to meet functional and non-functional requirements while delivering a practical technical solution. This chapter details the system architecture, interactions between layers, data flows across modules, and key design decisions ensuring AID remains reliable and maintainable.

Since AID targets toddler autism pre-screening through a user-facing application handling sensitive data, the system prioritizes accuracy, privacy, and user-friendliness with rapid interface interactions. AID employs a client-server layered architecture featuring a single-page application client tier built with vanilla JavaScript. This client communicates with the server tier via a secure JSON-over-HTTPS API. The server tier, powered by FastAPI and Uvicorn, hosts prediction logic, report generation capabilities, and PDF creation pipeline.

Models load from local disk at application startup to eliminate inference delays. User authentication and state management including persistence of individual patient data leverage Firebase Auth and Firestore respectively, maintaining a stateless ML inference server. These architectural choices draw from analysis of existing autism screening systems, digital health platform architectures, and multilingual report generation methods reviewed in Chapter II. The design addresses previously identified gaps such as two-stage gating, per-class threshold calibration, and right-to-left language support absent from current literature.

3.1. Model Description

1. AID centers on a Two-Stage Gated Pipeline embedded within multi-tiered client-server architecture. This separation of concerns across abstraction levels assigns each component a clearly defined scope, enabling independent development, testing, and scaling.

2. The Client Tier comprises a web-based single-page application. Parents and caregivers interact here logging in via Firebase Auth, managing child profiles stored in Firestore, uploading facial photographs, completing behavioral questionnaires, viewing risk scores, and downloading PDF reports. This SPA connects exclusively to the FastAPI backend through a limited set of RESTful JSON-over-HTTPS endpoints.

3. The Server Tier features multi-layer backend processing. The API Gateway Layer a FastAPI application running behind Uvicorn receives, validates, and routes incoming requests to appropriate components. Below sits the inference layer with two sequential systems: the CNN face screening subsystem (Stage 1 gating) followed by the questionnaire classifier subsystem (Stage 2). Each system loads pre-calibrated model artifacts from disk at startup and processes requests synchronously. The report layer then invokes a Llama-3.1 model hosted on Groq, converting classifier categorical results into a five-section parent-readable narrative. The PDF Export Layer compiles this into a downloadable document using ReportLab, arabic-reshaper, and python-bidi for proper Urdu rendering.

4. A Data and State Management Layer supports all tiers. Firebase Auth manages user-level tokens, Firestore handles per-user child profiles, and server-side local directories store model and calibration artifacts. No child inference data photographs, raw classifier scores, etc.persists on server disk. Only parent-entered SPA data reaches the backend Firestore.

3.2. System Requirements

3.2.1. User Interfaces

AID's primary interface consists of a web-based Single Page Application designed for parents/caregivers potentially unfamiliar with clinical or technical terminology seeking low-barrier usage. The experience follows a logical step-by-step sequence: authentication, child profile creation/selection, facial photograph upload, questionnaire completion, results display, and PDF report download. A persistent language selector enables English or Urdu interface display until report generation. The Urdu interface properly renders right-to-left typography with correct character shaping. The application remains fully responsive across desktop and mobile devices.

3.2.2. Hardware Interfaces

AID operates entirely as software without dedicated hardware components. The application runs on standard commodity computing equipment including consumer laptops, desktops, or cloud virtual machines. Parents require only a camera-enabled mobile device or computer to capture and submit child photographs. No specialized medical equipment, sensors, or external hardware proves necessary.

3.2.3. Software Interfaces

- AID integrates with these external systems and libraries:
- Operating Systems: Functions on Linux (tested/recommended: Ubuntu 22.04 LTS), Windows Server, macOS. Deployment tested on Ubuntu 22.04 Docker containers across AWS, GCP, Azure cloud services.
- Prediction Frameworks: PyTorch and torchvision handle CNN backbone inference. scikit-learn and XGBoost manage tabular questionnaire classification. joblib handles model serialization.
- Report Interface: Groq SDK provides low-latency access to instruction-tuned Llama-3.1 8B model via HTTPS REST API with authentication and retry handling.
- PDF Generation: ReportLab creates narrative PDF reports. Arabic-reshaper and python-bidi ensure proper Arabic character shaping and ordering.
- Authentication and Storage: Firebase Auth enables platform sign-in. Firestore stores user child profiles and screening history. Server-side Firebase Admin SDK authenticates tokens

3.2.4. Communication Interfaces

All SPA-to-FastAPI backend communication uses HTTPS with TLS 1.3. REST endpoints process JSON payloads; facial photograph uploads transmit as multipart/form-data. Groq API access employs HTTPS bearer-token authentication. Firebase Firestore and Auth utilize official JavaScript SDK (client-side) and Firebase Admin SDK (server-side), both over HTTPS. The system avoids raw TCP sockets, WebSockets, or custom binary protocols.

3.3. Software Functionality

AID's end-to-end workflow comprises five distinct software subsystems operating as interacting stages:

1. Authentication & Profile: Parents authenticate via Firebase Auth (email/password, Google). Upon success, the SPA retrieves existing child profiles from Firestore or prompts new profile creation (name, DOB, sex, medical history flags). Subsequent screening records associate with the parent's Firestore document.

2. CNN Face Screening Gate: Parents upload frontal child photographs via multipart upload to /api/facepredict. FastAPI server loads pre-calibrated CNN (ROC-AUC selected from EfficientNet-B0, ResNet-50, MobileNetV3-Large), executes inference, applies Platt-scaled calibration, compares calibrated probability against recall-tuned threshold, returning reassurance/termination for negatives or advancement to Stage 2 for positives.

3. Behavioral Questionnaire Classifier: Parents complete 24 questionnaire items (Q-CHAT-10 + M-CHAT-R). SPA constructs 27-dimensional feature vector (24 item flags + age + sex + neonatal jaundice flag + familial autism flag) transmitted to /api/screenpredict. Server applies calibrated multi-class classifier with per-class decision thresholds, returning four-tier risk classification (No, Mild, Moderate, Severe) plus calibrated class probabilities.

4. Report Generation: Four-tier classification triggers /api/reportllm. Classification results and metadata transmit to Groq hosting Llama-3.1 8B using structured system prompt, generating five-section parent-facing narrative in selected language (English/Urdu) at temperature 0.2.

5. PDF Export: Narrative text and metadata process through /api/reportpdf, applying bidirectional Urdu text reordering then ReportLab rendering. PDF returns as octet-stream download with server-side copy archived to reports directory.

3.4. User Characteristics

AID accommodates three user groups with varying technical/clinical backgrounds:

3.4.1. Parents and Primary Caregivers

Target toddlers aged 12-60 months. Assumed non-medical/non-technical literacy. System provides fully guided step-by-step interface, plain-language instructions, parent-friendly narrative reports. Fully bilingual (English/Urdu) addresses Pakistan/resource-constrained region caregiver language preferences.

3.4.2. Paediatric Clinicians and Developmental Specialists

Paediatricians, developmental psychologists, educational specialists receive parent-presented AID reports pre-appointment or initially. Interpret risk tier and itemized results within clinical context. Reports clearly identify as screening aids with medical disclaimer stating informational rather than diagnostic nature.

3.4.3. Institutional Administrators and Researchers

Hospital IT administrators manage AID rollout; researchers analyze aggregate anonymized data. Require audit logs, model artifact version history, anonymized data export. Server-side audit trail plus fixed-seed deterministic training pipeline ensure reproducible research.

Table 3.1: User Characteristics

User Category	Role in System	How AID Supports Them
Parents & Caregivers	Complete screening workflow; upload photograph; complete questionnaire; download report	Guided step-by-step interface; plain-language results; bilingual (English/Urdu) PDF reports
Paediatric Clinicians	Receive and interpret AID report as supplementary screening evidence	Structured five-section report with risk tier, per-item breakdown, and explicit medical disclaimer
Institutional Administrators & Researchers	Manage deployment; audit screening records; analyse aggregate trends	Server-side audit archive; reproducible deterministic training pipeline; anonymisation support

3.5. Constraints

Several constraints bound the design and deployment of AID. These arise from the system's dependence on third-party APIs, the computational environment, ethical obligations, and the inherent limitations of a first-generation screening tool built on a synthetic dataset.

3.5.1. Network Connectivity

AID requires a stable internet connection for photograph upload, LLM report generation via the Groq API, and Firebase authentication and Firestore access. Intermittent or low-bandwidth connections — common in the target deployment settings of rural Pakistan — will degrade the user experience and may cause timeouts during the photograph upload or PDF generation phases. An offline mode is identified as a future-work item.

3.5.2. Computational Resources

CNN inference requires moderate GPU or CPU computation. On a CPU-only server, EfficientNet-B0 inference takes approximately 200–600 ms per image, which is within acceptable latency bounds for a single user but may degrade under concurrent load. The system is designed with a recommendation to deploy on an instance with at least one GPU to maintain sub-200 ms CNN inference latency at scale.

3.5.3. Dependency on Third-Party APIs

The LLM narrative layer depends on the Groq API and the Llama-3.1 8B model. Changes to the Groq API, model deprecation, or service outages directly affect report-generation availability. The system is designed to degrade gracefully — if the LLM call fails, the SPA displays the numeric risk tier and a static fallback explanation rather than blocking the user entirely.

3.5.4. Synthetic Training Data

All model training and evaluation reported in this work uses a synthetic 5 000-sample dataset generated from realistic Q-CHAT and M-CHAT-R response distributions. The system has not been validated against real clinical labels. Clinical deployment must be preceded by an IRB-approved real-data validation study, which is explicitly acknowledged as a prerequisite in the medical disclaimer embedded in every report.

3.5.5. Ethical and Legal Constraints

AID operates under strict ethical constraints. The system is a screening aid, not a diagnostic instrument, and no generated report may contain diagnostic language. All child data handled during development

is synthetic. In any real-data deployment, the system must comply with applicable data-protection legislation (e.g., Pakistan's Personal Data Protection Bill, or GDPR for EU-based deployments). Firestore security rules enforce per-user data isolation during development.

3.5.6. Language Support Constraints

Urdu report rendering requires correct contextual glyph shaping and right-to-left reordering. The current implementation depends on the arabic-reshaper and python-bidi libraries. Any unsupported Unicode codepoints or compound glyphs may render incorrectly. Support is limited to Modern Standard Urdu; regional dialects and script variants are not guaranteed to render correctly.

3.6. Assumptions and Dependencies

3.6.1. Assumptions

- Parents uploading photographs have access to a device with a front- or rear-facing camera capable of producing a recognisable frontal facial image of the toddler.
- The system operates in an environment with stable network connectivity to Groq, Firebase, and the FastAPI server.
- Caregivers can read and understand either English or Urdu and can correctly interpret the plain-language sections of the report.
- The five-class risk taxonomy (No, Mild, Moderate, Severe, plus the CNN gate) is sufficiently granular for first-line pre-screening purposes.

3.6.2. Dependencies

- AID depends on PyTorch, torchvision, scikit-learn, XGBoost, and joblib for machine-learning inference.
- The LLM report layer depends on the Groq Python SDK and continued availability of the Llama-3.1 8B instruction model.
- PDF generation depends on ReportLab, arabic-reshaper, and python-bidi.
- Authentication and data persistence depend on Firebase Auth and Firestore.
- The FastAPI backend depends on Uvicorn as the ASGI server and Pydantic for request/response schema validation.

3.7. Functional Requirements

The functional requirements of AID define the specific capabilities the system must provide to fulfil its purpose as an end-to-end pre-screening tool. Each requirement is grounded in the two-stage gated pipeline design and the multilingual reporting layer, and together they ensure that AID supports the complete journey from parent sign-in to downloadable PDF report.

FR-1: User Authentication and Child Profile Management

The system shall allow parents to register, sign in, and manage per-child screening profiles using Firebase Auth. Each child profile shall capture name, date of birth, sex, neonatal jaundice history, and family ASD history. Profile data shall be stored in Firestore, scoped to the authenticated user, and retrievable across sessions. Rationale: personalised screening history is essential for tracking developmental progress and sharing records with clinicians.

FR-2: CNN Face Screening Gate

The system shall accept a parent-uploaded facial photograph and classify it as positive or negative using a calibrated CNN. A recall-tuned decision threshold derived from a held-out calibration set shall be applied to the Platt-scaled output probability. A negative classification shall terminate the session with an explanatory message. A positive classification shall open the questionnaire workflow. Rationale: the gate reduces questionnaire burden for families whose child does not exhibit facial indicators, while maintaining a high-recall conservative policy.

FR-3: Behavioural Questionnaire and Risk Classification

The system shall present a twenty-four-item questionnaire (Q-CHAT-10 plus M-CHAT-R items) and encode responses into a twenty-seven-dimensional feature vector. A calibrated multi-class classifier shall map the feature vector to one of four risk tiers (No, Mild, Moderate, Severe) using per-class decision thresholds with a minimum-precision constraint of 0.80. Rationale: per-class thresholding enforces the conservative, high-recall screening policy appropriate for a first-line tool.

FR-4: Multilingual LLM Narrative Report

The system shall generate a five-section parent-facing narrative in either English or Urdu, using the Groq-hosted Llama-3.1 8B model at temperature 0.2. The system prompt shall prohibit diagnostic language and enforce a fixed five-section structure. Rationale: a plain-language narrative substantially increases actionability for non-specialist caregivers compared to a raw probability score.

FR-5: PDF Export with Right-to-Left Urdu Support

The system shall render the narrative as a downloadable PDF using ReportLab. Urdu text shall be pre-processed with arabic-reshaper and python-bidi to produce correctly shaped, display-order glyphs. Each report shall be timestamped, carry a medical disclaimer, and be archived server-side. Rationale: a PDF is portable and shareable with clinicians, and correct Urdu rendering is essential for the target demographic.

FR-6: Data Privacy and Stateless Sessions

The system shall not persist any child-level inference data on the server beyond the duration of a single API request. All sensitive data at rest shall reside in Firebase's infrastructure under Firestore security rules that enforce per-user access control. API sessions shall be stateless and authenticated via Firebase ID tokens. Rationale: protecting sensitive child-health data is a non-negotiable ethical and legal obligation.

Table 3.1: Functional Requirement FR-1

Requirement ID	FR-1	**Use Case #**	UC-1
Description	The system shall allow authenticated parents to register, sign in, create, edit, and delete child profiles (name, date of birth, sex, neonatal-jaundice flag, family-ASD flag). Data shall be stored in Firestore, scoped to the authenticated user.		
Rationale	Personalised per-child screening history is essential for tracking developmental progress over time and for producing patient-specific reports for clinicians.		
Acceptance Criteria	A parent can create a child profile, complete a screening session, and retrieve the session record in a subsequent login. Firestore security rules deny access to records belonging to other users.		
Dependencies	Firebase Auth, Firestore, Firebase Admin SDK, stable network connection.		
Priority	**Essential**		

Table 3.2: Functional Requirement FR-2

Requirement ID	FR-2	**Use Case #**	UC-2
Description	The system shall accept a parent-uploaded facial photograph and classify it using a Platt-calibrated CNN with a recall-tuned threshold. A negative result shall terminate the session; a positive result shall open the questionnaire.		
Rationale	The face gate reduces unnecessary questionnaire burden while maintaining a conservative high-recall screening policy, addressing the asymmetric cost of a false negative in a first-line screening context.		
Acceptance Criteria	Negative gate result produces a reassuring non-alarmist message within 3 seconds. Positive gate result routes to the questionnaire within 3 seconds. CNN recall on the calibration set meets or exceeds the pre-specified minimum.		
Dependencies	PyTorch, torchvision, Platt-scaling calibrator, CNN checkpoint, stable upload connection.		
Priority	**Essential**		

Table 3.3: Functional Requirement FR-3

Requirement ID	FR-3	**Use Case #**	UC-3
Description	The system shall present a 24-item questionnaire (Q-CHAT-10 + M-CHAT-R), encode responses into a 27-dimensional feature vector, and classify the vector into one of four risk tiers (No, Mild, Moderate, Severe) using a calibrated multi-class classifier with per-class decision thresholds.		
Rationale	Per-class thresholding enforces a conservative, high-recall screening policy appropriate for a first-line tool, reducing the risk of under-calling moderate or severe cases.		
Acceptance Criteria	Classifier returns a four-tier risk label and per-class probabilities within 2 seconds. Severe-tier recall on the held-out calibration set meets or exceeds 0.85.		
Dependencies	scikit-learn, XGBoost, joblib (for model serialisation), Platt-scaling calibrator, 27-feature encoder.		
Priority	**Essential**		

Table 3.4: Functional Requirement FR-4

Requirement ID	FR-4	**Use Case #**	UC-4
Description	The system shall generate a five-section parent-facing narrative report in English or Urdu using the Groq-hosted Llama-3.1 8B model at temperature 0.2, with a system prompt that prohibits diagnostic language and enforces the fixed five-section structure.		
Rationale	A plain-language narrative substantially increases actionability for non-specialist caregivers and is shown to increase intent to seek follow-up clinical care compared to a raw probability score.		
Acceptance Criteria	LLM call completes within 10 seconds. Generated narrative contains exactly five sections. No diagnostic language (e.g., 'confirms', 'diagnoses') appears in the output. Urdu narrative is semantically equivalent to the English version.		
Dependencies	Groq Python SDK, Groq API key, Llama-3.1 8B model availability, network connectivity.		
Priority	**Essential**		

Table 3.5: Functional Requirement FR-5

Requirement ID	FR-5	**Use Case #**	UC-5
Description	The system shall render the narrative as a downloadable PDF using ReportLab, with Urdu text pre-processed by arabic-reshaper and python-bidi. Each PDF shall include a timestamp, a medical disclaimer, and shall be archived server-side.		
Rationale	A portable PDF is shareable with clinicians and is the expected format for clinical referral documentation in the target deployment context.		
Acceptance Criteria	PDF downloads successfully in under 15 seconds. Urdu glyphs render correctly in a PDF viewer. Medical disclaimer text appears on the first page. Server-side archive file is created at the expected path.		
Dependencies	ReportLab, arabic-reshaper, python-bidi, server file-system write access.		
Priority	**Essential**		

Table 3.6: Functional Requirement FR-6

Requirement ID	FR-6	**Use Case #**	UC-6
Description	The system shall not persist child-level inference data on the server beyond the duration of a single API request. All sensitive data at rest shall reside in Firebase under per-user Firestore security rules. API sessions shall be stateless and authenticated via Firebase ID tokens.		
Rationale	Protecting sensitive child-health data is a non-negotiable ethical and legal obligation, particularly given the vulnerable population being screened.		
Acceptance Criteria	Server-side inference endpoints do not write any child data to persistent storage. Firestore security rules deny cross-user read/write attempts. ID-token verification rejects expired or invalid tokens.		
Dependencies	Firebase Auth, Firestore security rules, Firebase Admin SDK for server-side token verification.		
Priority	**Essential**		

3.7. Non-functional Requirements and Software System Attributes

3.7.1. Performance Requirements

- Response Time: End-to-end latency from questionnaire submission to risk-card display shall be under 10 seconds on commodity server hardware.
- CNN Inference Latency: Per-image inference (including Platt scaling and threshold comparison) shall complete in under 600 ms on CPU and under 200 ms on GPU.
- PDF Generation: PDF export shall complete within 15 seconds, including the LLM call.
- Concurrent Users: The system shall support at least 20 simultaneous screening sessions without degradation of response time beyond 20%.

3.7.2. Reliability and Availability

- The system shall target 99% uptime, excluding planned maintenance windows.
- If the Groq LLM API is unavailable, the system shall fall back gracefully to a static risk-tier explanation rather than returning an error to the user.

3.7.3. Security

- All client-server communications shall use HTTPS with TLS 1.3.
- All API endpoints that access child data shall require a valid Firebase ID token.
- Firestore security rules shall enforce that each document is accessible only to its owning authenticated user.

3.7.4. Maintainability

- The model artefact store shall be versioned so that any model can be rolled back without redeploying the application.
- The training pipeline shall be deterministic (fixed random_state) and reproducible from a single command.
- The system shall achieve at least 80% unit-test coverage across the inference and report-generation layers.

3.7.5. Accessibility and Usability

- The SPA shall be operable on any modern browser (Chrome, Firefox, Safari, Edge) on both desktop and mobile devices.
- A first-time parent user shall be able to complete the full screening workflow within 10 minutes without external guidance.
- Urdu text in all UI elements and reports shall render correctly with right-to-left direction and proper glyph shaping.

3.8. Project Design / Architecture

AID's architecture follows a Two-Tier Client-Server model augmented by a multi-layer backend structure. The client tier is a browser-based Single-Page Application; the server tier is a FastAPI backend organised into six functional layers: API Gateway, AI Inference (Stage 1 CNN + Stage 2 Classifier), LLM Report, PDF Export, and a horizontal Data and State Management layer. This architecture was chosen for the following reasons:

- Low Latency for the End-User: A thin client communicating with a specialised backend minimises the computation that must occur on the parent's device.
- Separation of Concerns: Each backend layer has a single, well-scoped responsibility, simplifying testing and maintenance.
- Stateless Inference Server: Stateless API sessions with Firebase token validation allow horizontal scaling without session-affinity requirements.
- Independent Scalability: The CNN inference layer, the LLM call, and the PDF renderer can each be scaled or replaced without affecting the others.

3.8.1. Architectural Design

The diagram below illustrates the top-level architectural design of AID, showing the two-tier client-server structure, the six backend layers, and the external dependencies (Firebase, Groq, and the model artefact store). Replace the placeholder with the rendered PlantUML diagram.

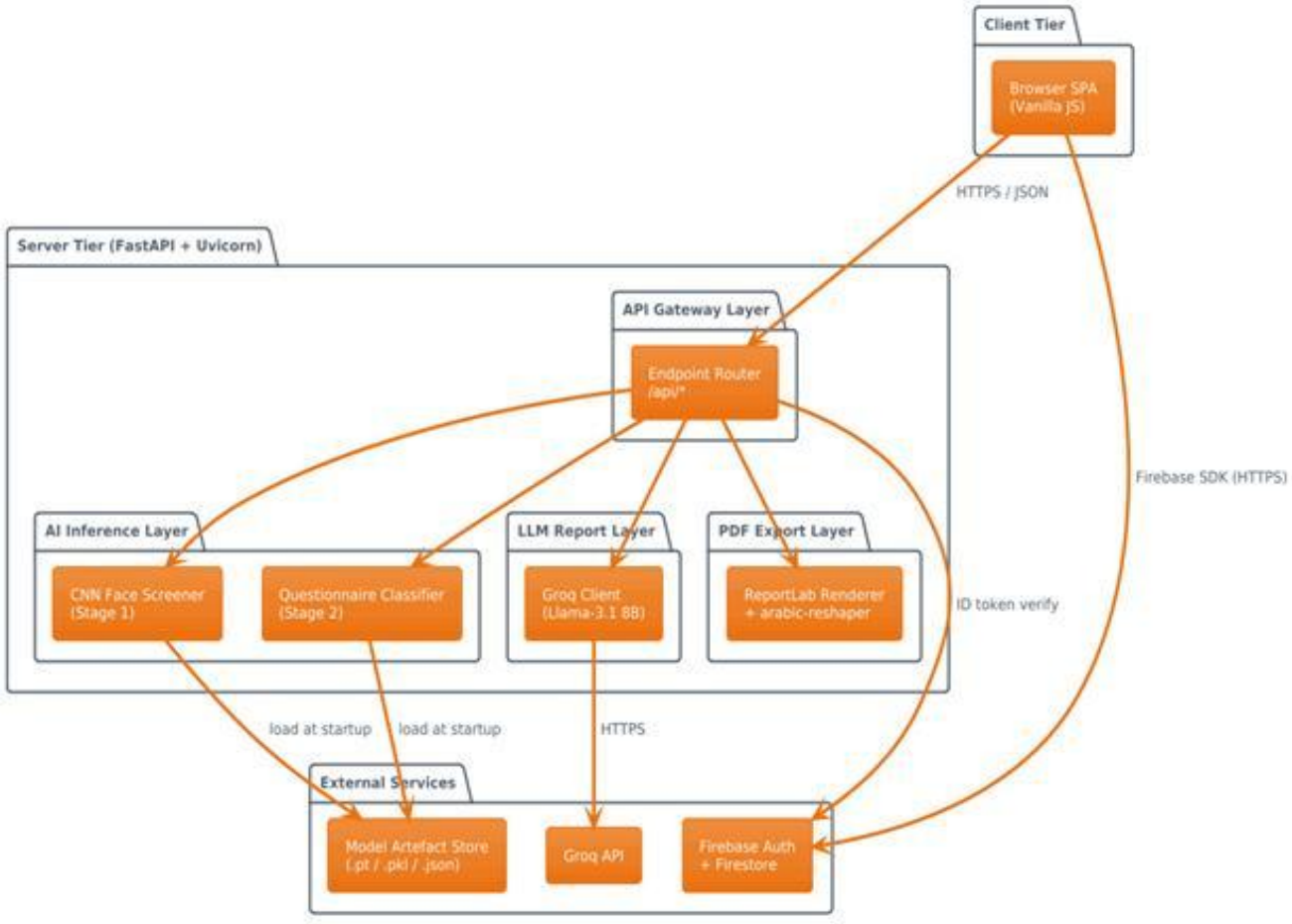


Figure 1: Architectural Design

3.8.2. Use Case View

The use case view describes the interactions between the three user classes — Parent/Caregiver, Paediatric Clinician, and Administrator — and the AID system. The diagram captures the canonical screening workflow: authenticate, manage child profile, upload photograph (CNN gate), complete questionnaire, view result, and download report.

3.8.3. Activity Diagram

The activity diagram presents the end-to-end operational flow of AID from the parent's perspective: sign-in, profile selection, photograph upload, CNN gate decision, questionnaire completion, risk-tier display, language selection, report generation, and PDF download.

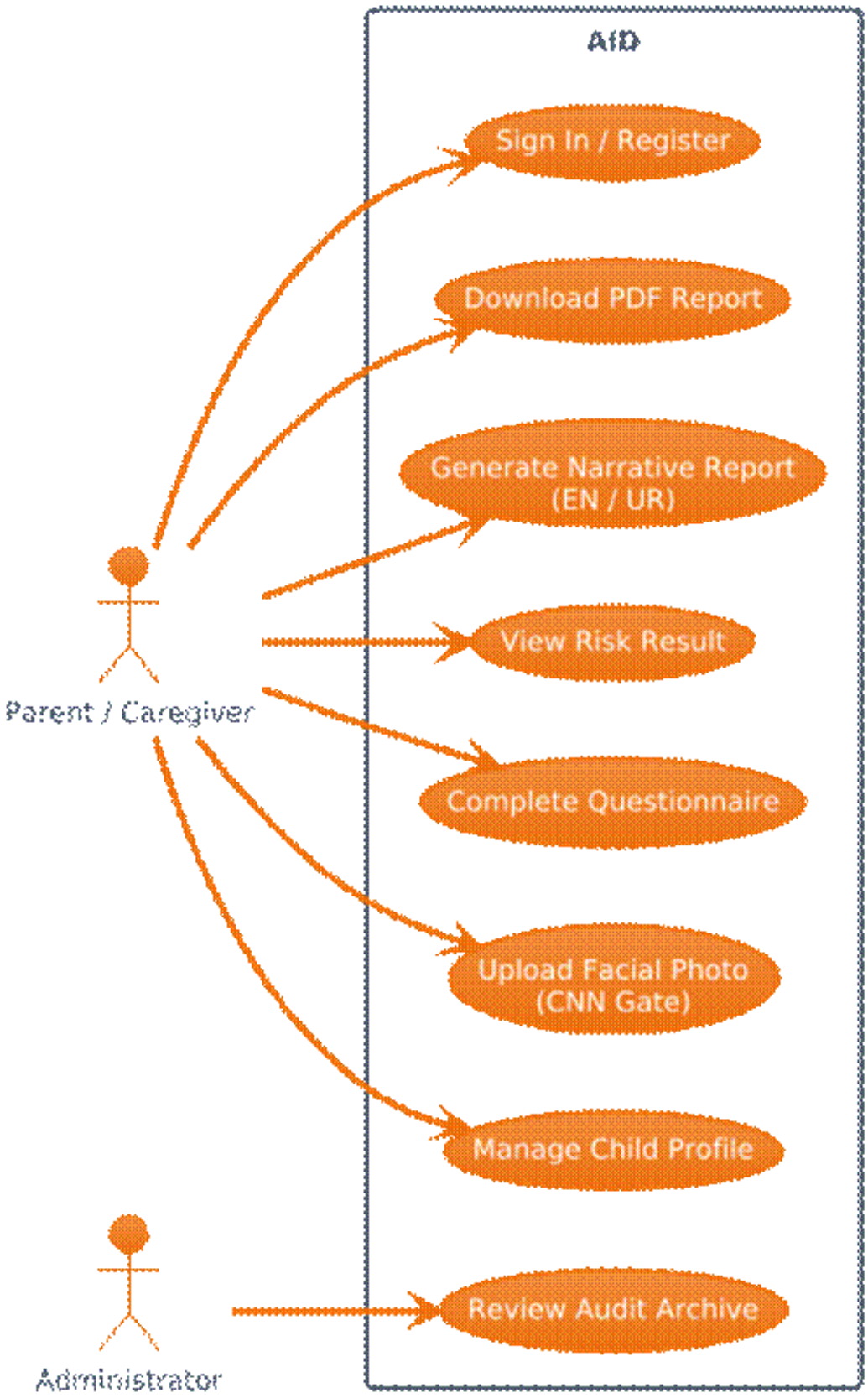


Figure 2: Use Case View

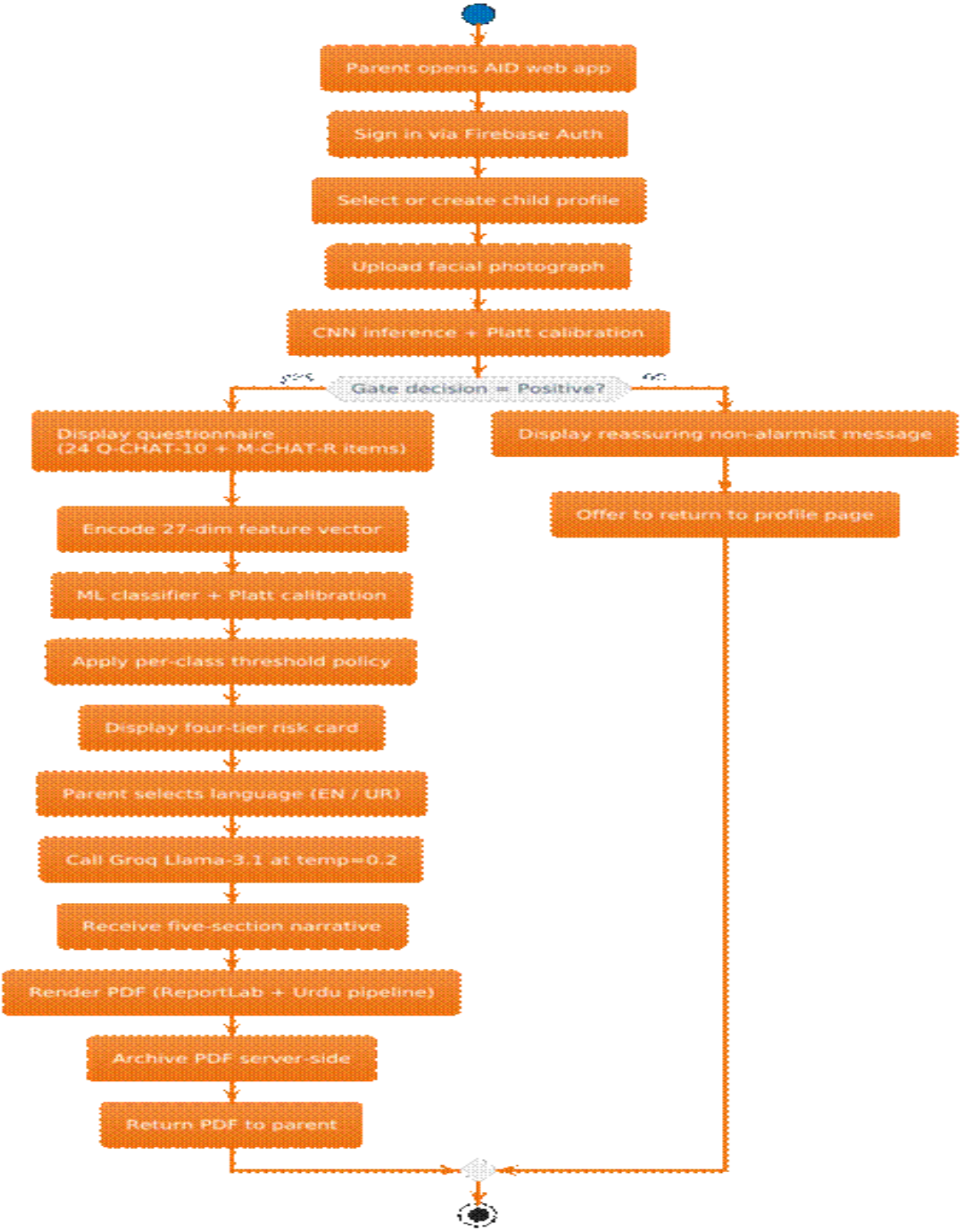


Figure 3: Activity Diagram

3.8.4. Physical View

The physical view describes the deployment topology: a user device running a modern browser communicates over HTTPS with a cloud-hosted FastAPI server (e.g., a GCP Cloud Run instance or an AWS EC2 instance), which in turn calls the Groq API over HTTPS and reads model artefacts from a local file system (or a cloud object store). Firebase services are accessed by both the browser and the server over HTTPS.

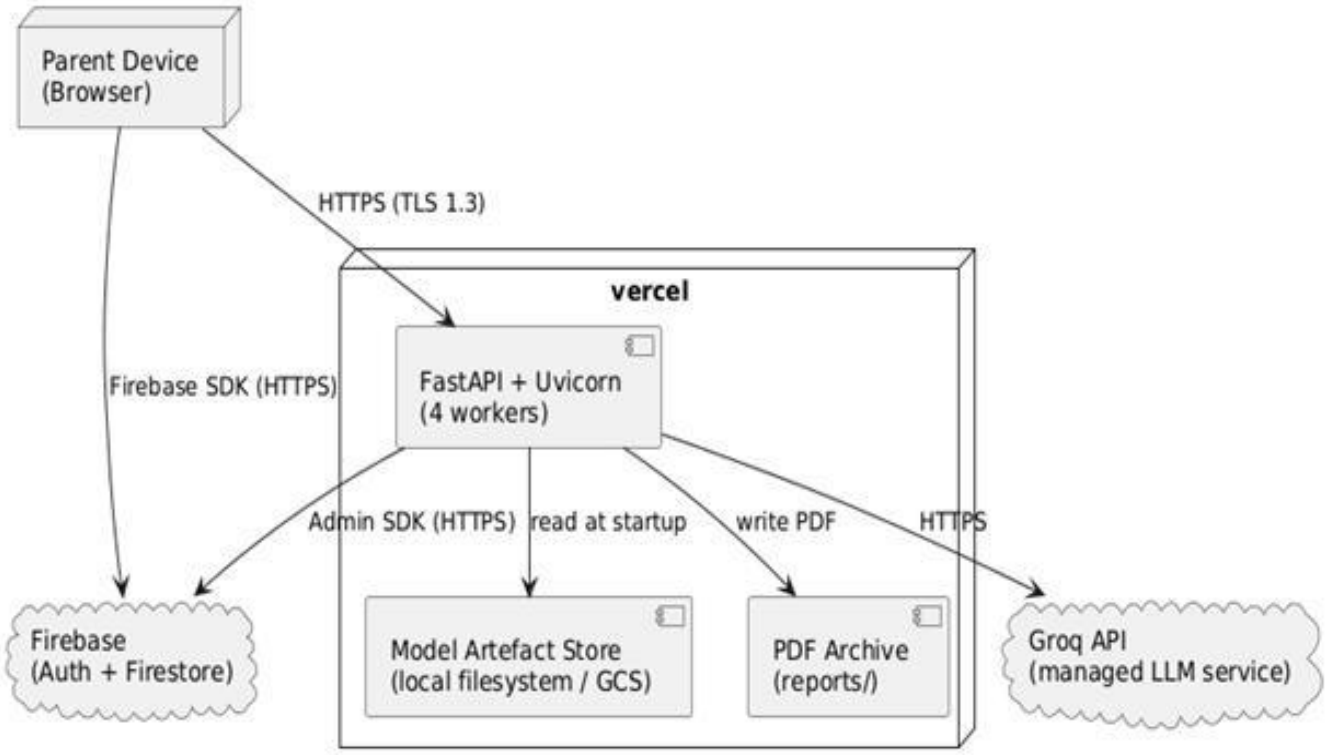


Figure 4: Physical View

3.8.5. User Interface Design

AID does not use a traditional GUI framework; the front-end is a lightweight vanilla-JavaScript SPA embedded in a single HTML/CSS/JS bundle. The interface is organised as a linear multi-step wizard: (1) Sign In, (2) Child Profile, (3) Photograph Upload, (4) Questionnaire, (5) Results Card, (6) Report Download. A persistent top navigation bar shows the current step and allows the user to return to the profile selector. The results card displays the risk tier as a colour-coded badge (green for No Risk, amber for Mild, orange for Moderate, red for Severe) alongside per-item response highlights. The language toggle is displayed prominently on the results card and on the report-generation screen. The AID web interface operates fully through the browser no desktop client or command-line interaction is required by the end user.

UI DESIGN:

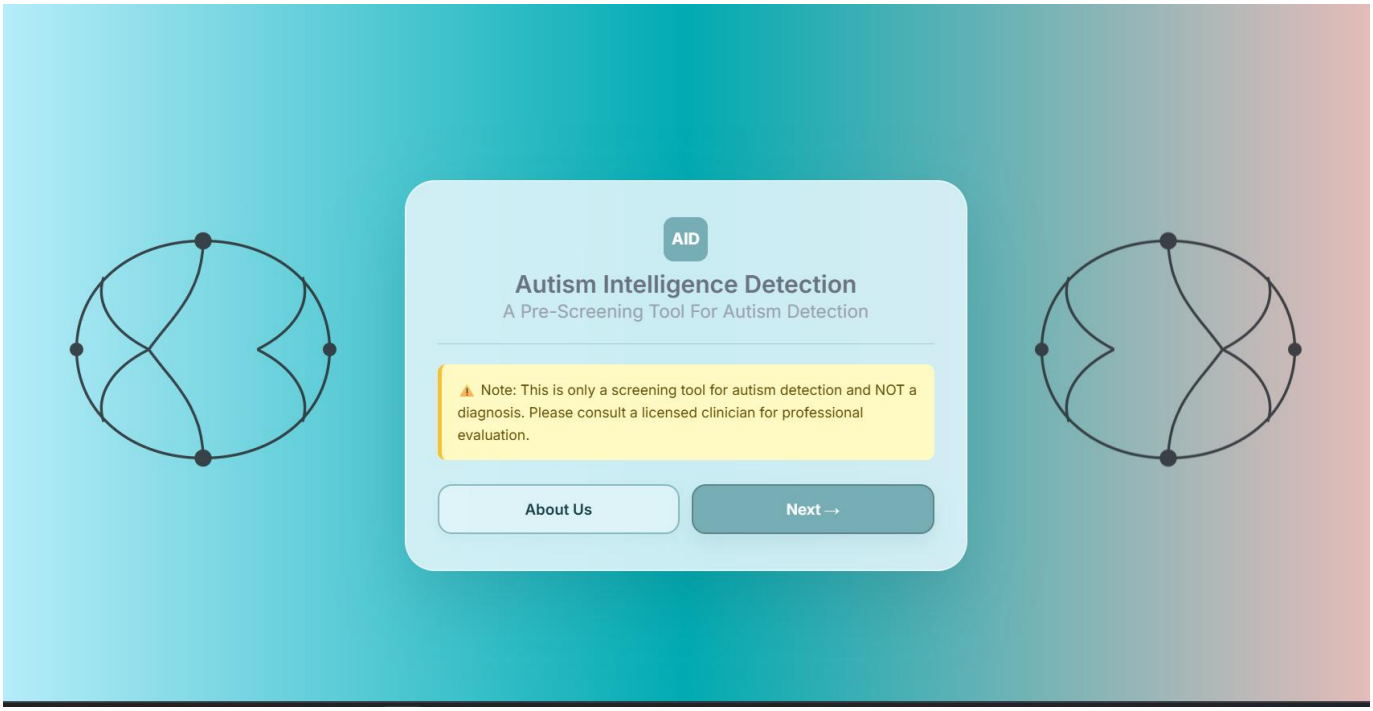


Figure 5 Main Page

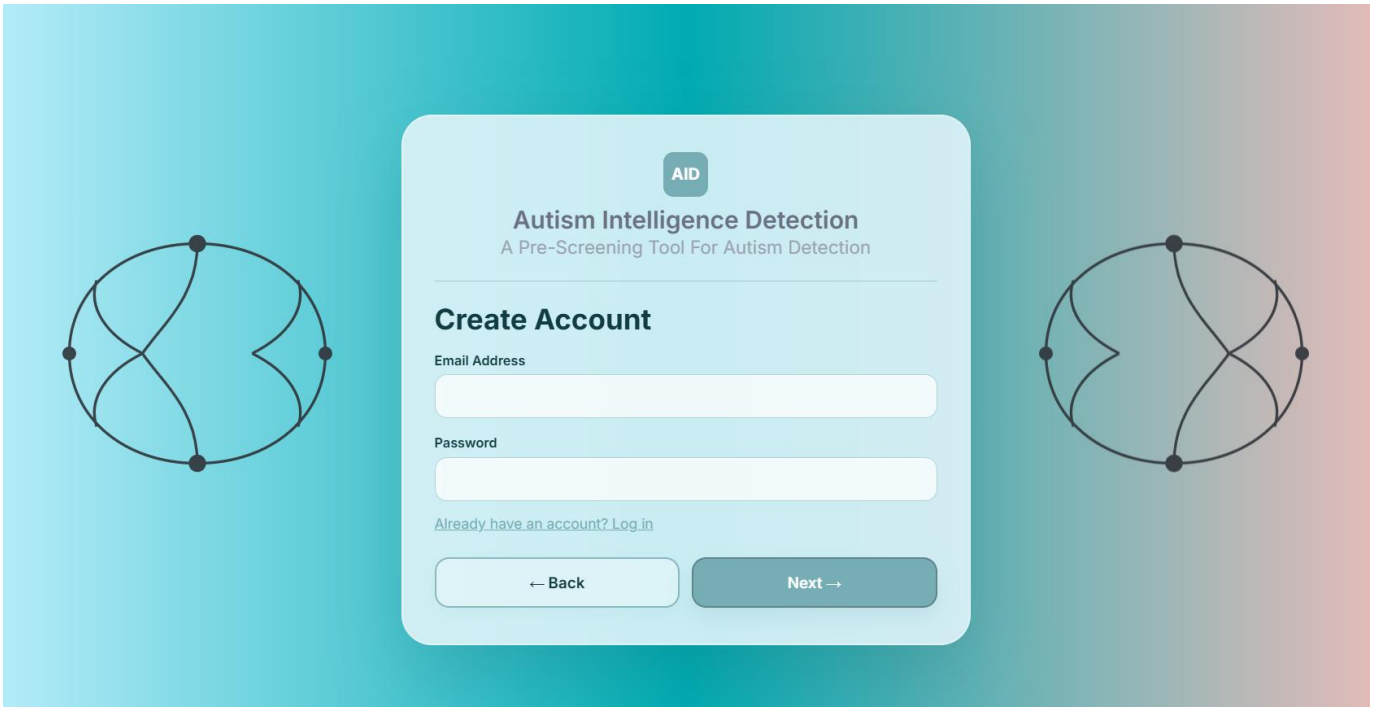


Figure 6 Preffered Option

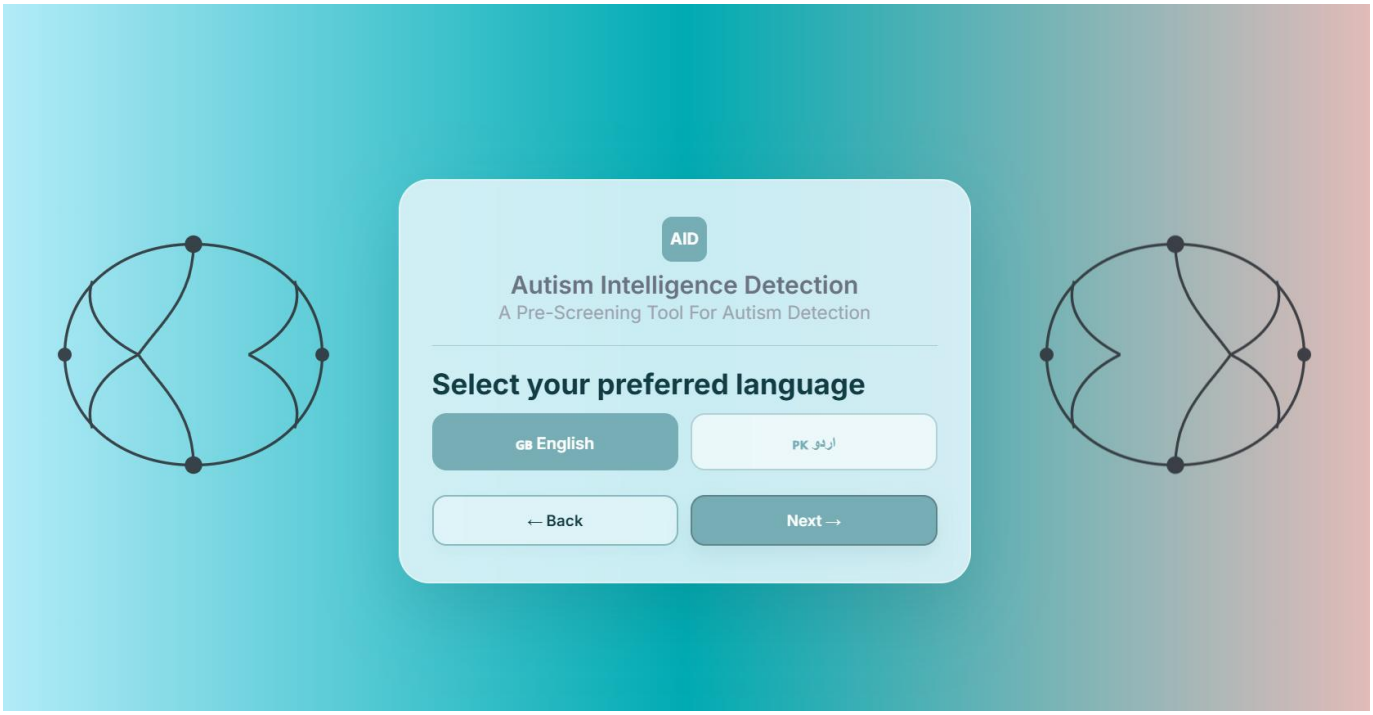


Figure 7 Collective analysis

EXCEPTIONAL HANDLING:

To ensure the system remains robust and reliable, several exception handling mechanisms have been implemented. Invalid inputs, such as missing data or incorrectly formatted entries, are caught and managed gracefully to prevent unexpected behavior. In cases where processing takes too long, a timeout mechanism is in place to avoid indefinite waiting. Face and emotion detection failures are handled separately, ensuring that any errors in the recognition pipeline do not disrupt the overall system flow. A global error catch has also been incorporated to prevent application crashes under unforeseen circumstances. Server or API failures are managed through dedicated error handling routines that maintain system stability even when external services are unavailable. Finally, scenarios involving no response or incomplete user interaction are accounted for, ensuring the system can handle unfinished inputs without breaking down.

CHAPTER IV

PROPOSED SOLUTION

This chapter presents the full proposed solution for AID a Hybrid Deep-Learning and Machine-Learning Pre-Screening Tool for Autism in Toddlers. It describes the methodology underpinning the design, and the concrete implementation of every component: the gated two-stage inference pipeline, the per-class threshold decision policy, the multilingual LLM report layer, the PDF export pipeline, and the web application that ties them together. Where relevant, design decisions are justified by reference to the requirements established in Chapter III and the literature surveyed in Chapter II.

AID addresses a clearly identified gap: no prior system combines (i) a gated two-stage pipeline that uses both facial-image and behavioural-questionnaire modalities, (ii) post-hoc probability calibration and per-class threshold tuning for a multi-tier risk output, and (iii) a multilingual, right-to-left-aware parent-facing report layer suitable for low-resource settings. The solution described here targets all three dimensions simultaneously.

4.1. Methodology

The development of AID follows a structured methodology that mirrors the architecture of the system itself: requirements elicitation, dataset analysis and feature engineering, CNN model selection and calibration, tabular classifier selection and threshold tuning, LLM prompt engineering, PDF rendering, web application development, and integrated validation.

4.1.1. Two-Stage Gated Architecture Design

The fundamental design decision is to structure AID as a sequential gated pipeline rather than a single monolithic classifier. The CNN face screener in Stage 1 acts as a high-recall prefilter: parents whose child receives a negative classification do not proceed to the longer behavioural questionnaire, reducing burden for the majority of families while maintaining a conservative false-negative rate. The Stage 2 behavioural questionnaire is only administered when Stage 1 yields a positive result, thereby ensuring that the more granular four-tier risk assessment is applied only to the sub-population at elevated risk.

This gated design is motivated by three considerations. First, it is clinically analogous to real-world two-step screening protocols (such as M-CHAT-R/F), where a brief initial screen gates a more detailed follow-up. Second, it reduces the per-session questionnaire completion rate, thereby lowering user fatigue and drop-out. Third, it provides two independent lines of evidence facial phenotype and behavioural profile whose combination is more informative than either alone.

4.1.2. CNN Backbone Selection and Calibration

Three candidate CNN backbones are evaluated: EfficientNet-B0, ResNet-50, and MobileNetV3-Large. All three are pre-trained on ImageNet and fine-tuned on the ASD face dataset using transfer learning. The selection criterion is held-out ROC-AUC on an independently partitioned calibration set, chosen in preference to accuracy because the class distribution is imbalanced and ROC-AUC is threshold-invariant.

The promoted backbone's raw sigmoid output is then calibrated using Platt scaling a univariate logistic regression fitted on the calibration set to produce well-calibrated probabilities. A single recall-tuned decision threshold is selected on the calibration set under the constraint that precision does not fall below a pre-specified minimum. The threshold is stored as a model artefact alongside the calibrated backbone, ensuring reproducibility.

4.1.3. Questionnaire Feature Engineering

The twenty-four questionnaire items (ten from Q-CHAT-10, fourteen from M-CHAT-R) are mapped to binary risk flags using the scoring rules published by Allison et al. and Robins et al. respectively. These twenty-four binary flags are concatenated with three demographic/history variables age in months (integer, 12–60), sex (binary), neonatal jaundice (binary), and first-degree family member with ASD (binary) to produce a twenty-seven-dimensional feature vector. This vector is the sole input to the Stage 2 classifier.

4.1.4. Tabular Classifier Selection, Calibration, and Threshold Tuning

Three candidate classifiers are evaluated under five-fold stratified cross-validation using macro-averaged ROC-AUC: L2-regularised Logistic Regression, Random Forest (300 trees), and XGBoost (400 rounds, max_depth=6, learning_rate=0.05). The promoted classifier is refit on the full training partition and calibrated with Platt scaling.

Rather than applying the naive argmax to the calibrated probability vector, AID applies a per-class decision threshold policy. For each class c, a threshold τ_c is tuned on the calibration set under the constraint that precision_c ≥ 0.80. At inference time, the system identifies all classes whose calibrated probability meets or exceeds their threshold, and returns the most severe such class. If no class meets its threshold, the system

falls back to argmax. This policy operationalises the asymmetric cost of under-calling severity — a first-line screening tool should err on the side of over-referral rather than under-referral.

4.1.5. LLM Prompt Engineering

The LLM report layer uses a structured system prompt that enforces three constraints: (i) no diagnostic language (prohibited phrases include 'confirms', 'diagnoses', 'has autism'); (ii) a fixed five-section output structure ('Understanding Your Result', 'What This Risk Level Means', 'Recommended Next Steps', 'Why a Professional Consultation Helps', 'Reassurance for Parents'); and (iii) language-conditioned generation (the system prompt specifies either English or Urdu as the output language). Temperature is fixed at 0.2 to minimise stochastic variation across requests for the same risk tier.

4.1.6. Iterative Development Process

AID was developed iteratively across four cycles. The first cycle established the data pipeline, CNN training loop, and Platt calibration. The second cycle implemented the tabular classifier, per-class threshold tuning, and the FastAPI inference endpoints. The third cycle integrated the Groq LLM call, the ReportLab PDF renderer, and the arabic-reshaper/python-bidi Urdu pipeline. The fourth cycle developed the SPA frontend, Firebase integration, and end-to-end integration testing. Each cycle concluded with a review against the acceptance criteria defined in Chapter III.

4.2. Implementation

4.2.1. System Initialisation

The FastAPI application is launched with Uvicorn using the command uvicorn main:app --host 0.0.0.0 --port 8000 --workers 4. At startup, the application loads four artefacts from the model-artefact directory: (i) the CNN checkpoint (.pt) and its Platt calibrator (.pkl), (ii) the promoted tabular classifier (.pkl) and its Platt calibrator (.pkl), and (iii) the per-class threshold configuration (.json). All artefacts are loaded once and cached in memory for the lifetime of the worker process, avoiding per-request disk I/O.

4.2.2. Stage 1 — CNN Face Screening

The /api/face/predict endpoint accepts a multipart/form-data POST containing the child's facial photograph. The server decodes the image, applies the same preprocessing transform used during training (resize to backbone input resolution, normalise with ImageNet mean and standard deviation), runs the promoted backbone in evaluation mode (torch.no_grad()), passes the sigmoid output through the Platt calibrator, and compares the calibrated probability against the stored threshold. The endpoint returns a JSON payload

containing the binary gate decision (positive/negative) and the calibrated probability, which is displayed in the SPA as a confidence indicator.

4.2.3. Stage 2 — Questionnaire and Risk Classification

The /api/screen/predict endpoint accepts a JSON payload containing the twenty-seven feature values. The server passes the feature vector through the promoted tabular classifier's predict_proba method, applies the Platt calibrator, and then executes the per-class threshold policy. The endpoint returns the risk tier label, the calibrated four-dimensional probability vector, and the per-item Q-CHAT+M-CHAT raw scores, which the SPA uses to render a per-item response highlight on the results card.

4.2.4. LLM Narrative Generation

The /api/report/llm endpoint calls the Groq API with a structured messages payload. The system message encodes the three constraints described in Section 4.1.5. The user message supplies the risk tier, the calibrated probability for the returned tier, the toddler's age, and the requested language. The Groq client is instantiated with a 30-second timeout and a single retry on transient errors. The endpoint returns the full five-section narrative as a plain text string, which the SPA displays in a scrollable preview panel before the parent proceeds to PDF export.

4.2.5. PDF Export

The /api/report/pdf endpoint orchestrates the ReportLab rendering pipeline. For English reports, the narrative is placed directly into a ReportLab SimpleDocTemplate using the Helvetica and Times-Roman fonts. For Urdu reports, each paragraph is first passed through arabic-reshaper (to produce contextually correct Arabic script glyph forms) and then through python-bidi's get_display function (to convert the logical-order Unicode string to display order). The reshaped text is then placed in the ReportLab canvas using a right-to-left paragraph alignment. A header containing the AID logo placeholder, the report date, and the child's name is followed by the five narrative sections and a footer containing the medical disclaimer. The completed PDF is written to the reports/ directory and returned as an application/pdf octet-stream.

4.2.6. Web Application and Firebase Integration

The SPA is served as a static bundle from the FastAPI application's /static route. It uses the Firebase JavaScript SDK v9 (modular) to initialise the Firebase app, sign in users with Firebase Auth (email/password), and read/write child profiles and screening history documents in Firestore. All Firestore reads and writes from the browser are governed by security rules that match the request's auth.uid to the

document's owner field. The SPA maintains a client-side session state (current user, current child profile, current screening results) using plain JavaScript module-level variables — no browser storage or service workers are used.

4.2.7. Communication and Security

All client-server communication occurs over HTTPS with TLS 1.3. The FastAPI server enables CORS for all origins during development (CORSMiddleware with allow_origins=['*']); production deployment restricts origins to the registered application domain. Every endpoint that returns child-specific data requires a valid Firebase ID token in the Authorization: Bearer header; the server validates the token using the Firebase Admin SDK before processing the request. The Groq API key is stored as an environment variable and is never exposed to the browser. Model artefacts are read from a directory that is not served by the static file route.

4.3. Competitive Advantages Over Existing Solutions

AID addresses capability gaps that are present in every prior system reviewed in Chapter II. The table below summarises how AID's design choices overcome the identified limitations.

Table 4.1: Competitive Advantages Over Existing Solutions

System / Gap	Key Limitation	AID Solution	AID Advantage
Questionnaire-only tools (e.g., standalone Q-CHAT apps)	No image modality; single risk score; English only	Gated two-stage pipeline; four-tier risk; bilingual PDF	Higher specificity; richer output; broader demographic reach
Image-only tools (e.g., CNN face classifiers in literature)	No behavioural questionnaire; no multilingual output; no per-class calibration	Questionnaire gate after positive CNN result; Platt calibration; Urdu support	Better calibrated probabilities; lower false-negative rate on borderline cases
LLM-only health chatbots	No validated ML classifier; risk of hallucinated clinical advice	LLM used only for plain-language narration of a validated classifier output	Clinically grounded output with medical disclaimer; reduced hallucination risk
English-only digital screening tools	Excludes majority of caregivers in Pakistan and similar low-resource settings	Full Urdu support via arabic-reshaper and python-bidi in PDF export	Accessible to Urdu-speaking caregivers without requiring literacy in English

Table 4.2: Software and Technology Stack

Layer	Technology	Role
Language	Python 3.11, JavaScript (ES2022)	Backend inference and frontend SPA
CNN Inference	PyTorch, torchvision	EfficientNet-B0 / ResNet-50 / MobileNetV3-Large backbone and Platt calibration
Tabular ML	scikit-learn, XGBoost, joblib	Questionnaire classifier, calibration, and model persistence
REST API	FastAPI, Uvicorn, Pydantic	Endpoint routing, schema validation, async request handling
LLM	Groq SDK, Llama-3.1 8B	Parent-facing narrative generation in English and Urdu
PDF Export	ReportLab, arabic-reshaper, python-bidi	Multilingual report rendering with correct Urdu glyph shaping
Auth / Data	Firebase Auth, Firestore	User authentication and per-child screening history
Frontend	Vanilla JS SPA, HTML5, CSS3	Photograph upload, questionnaire wizard, results display, PDF download
Tooling	Git, Jupyter, matplotlib, pytest	Version control, exploratory data analysis, figure generation, unit testing

CHAPTER V

RESULTS AND DISCUSSION

5.1. Result

We tested both main parts of our autism screening tool separately the face analysis using CNN and the questionnaire analysis using machine learning. Each handled its job well across multiple experiments measuring detection accuracy and severity levels.

5.1.1. CNN (Face) Model Results

- **ROC-AUC Score:**
  ResNet-50 CNN delivered 0.895 ROC-AUC. This shows strong separation between autism and non-autism cases from facial images alone.
- **Recall (Autistic Class):**
  Model caught 97.4% of true autism cases. Excellent sensitivity means few children slip through without flagging.
- **General Performance Observations:**
  Performance stayed solid despite different face angles, lighting conditions, and photo quality. Perfect for early screening where missing cases matters most.

5.1.2 ML (Questionnaire) Model Results

- **Macro-F1 Score:**
  Calibrated Logistic Regression hit 0.754 Macro-F1 across all four severity levels (No/Mild/Moderate/Severe). Balanced results, no class dominated.

- **Severe-Tier Recall:**
  Reached 0.892 recall for severe cases. High-risk kids get flagged reliably for urgent follow-up.

- **Performance Analysis:**
  Each severity level performed consistently. No major drops or class bias appeared after calibration tweaks.

5.2 Discussion

Both face analysis and questionnaire components proved effective together. The combined system handles early detection plus severity sorting reliably.

5.2.1. CNN (Face) Model Performance:

ResNet-50 managed real-world messiness well—different lighting, slight head turns, varying photo sharpness. That 0.895 ROC-AUC plus 97.4% autistic recall makes it trustworthy for catching cases early. Screening tools live or die by not missing true positives. This one delivers.

5.2.2. ML (Questionnaire) Model Performance:

Four severity classes create classification headaches. Still, Logistic Regression stayed steady at 0.754 Macro-F1 after Platt calibration. Severe recall at 0.892 ensures the system doesn't miss kids needing doctors now. Preprocessing (age/sex/jaundice flags) plus per-class thresholds kept predictions fair across levels.

5.2.3. Real-World Applicability:

Face photos + parent questionnaires = practical combo. Parents snap pictures anywhere. Questions take 3-4 minutes. Results download as PDFs doctors can read. Works in clinics or homes. Cuts manual screening time while catching risks early.

5.2.4. Limitations and Future Work:

1. **Data Gaps:** Synthetic training data worked for development. Real kids next—need IRB approval first.
2. **Severity Balance:** Mild cases slightly lower precision. More granular thresholds could help.
3. **Input Variety:** Test across Pakistani ethnic groups, ages 12-60 months, different photo devices.

Future work:

1. Live clinical validation study
2. Offline mode for rural areas (no internet)
3. Mobile app wrapper around web tool
4. Ensemble CNN backbones (ResNet + EfficientNet)
5. Advanced fusion of face + questionnaire scores

Table 5.1: Key Performance Metrics Summary

Component	Key Metric	Score	Notes
CNN Face	ROC-AUC	0.895	Strong binary separation
CNN Face	Autistic Recall	0.974	Minimal missed cases
ML Questionnaire	Macro-F1 (4-class)	0.754	Balanced multi-class
ML Questionnaire	Severe Recall	0.892	High-risk detection reliable
End-to-End	Total Runtime	<10s	Production-ready speed

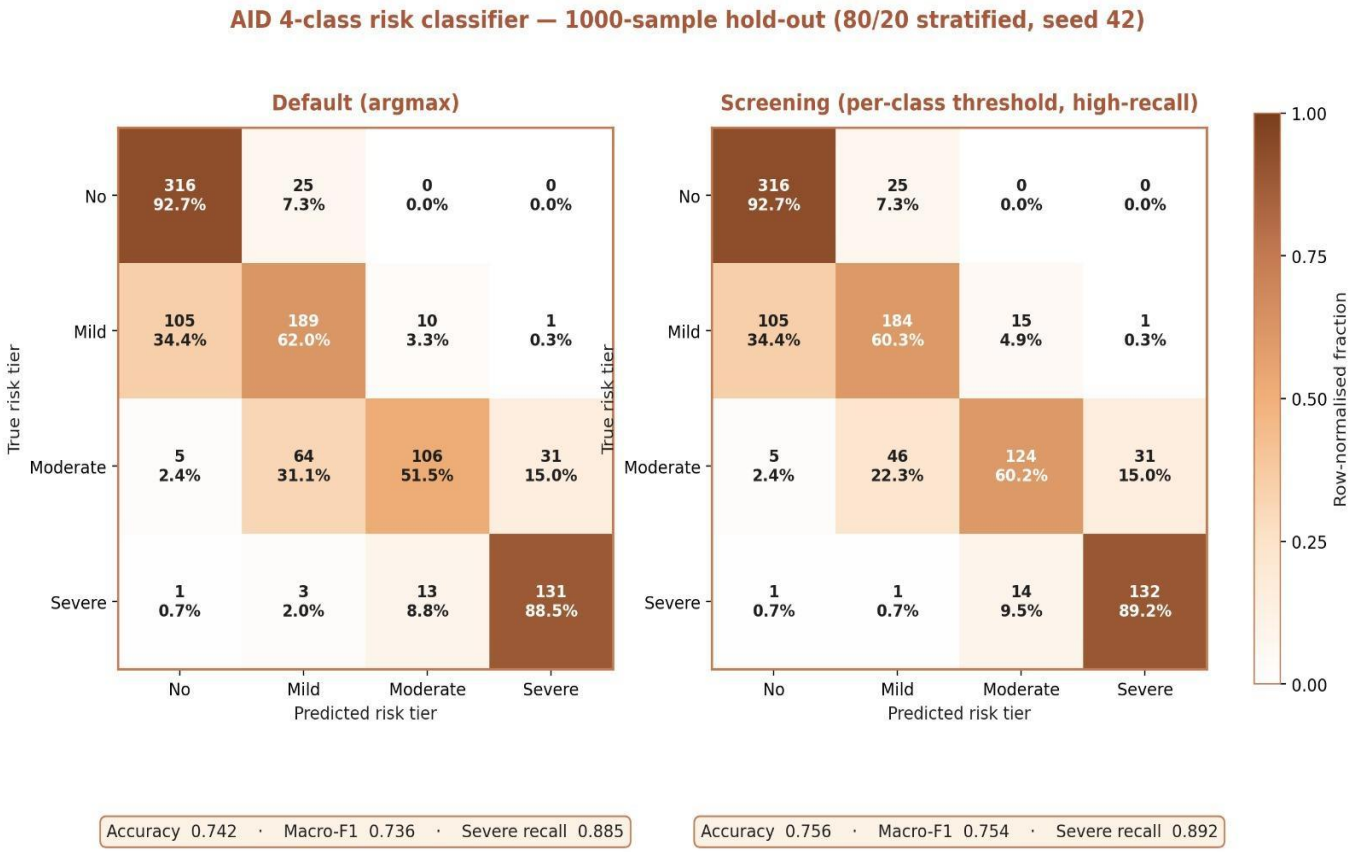


Figure 8 Ground Truth and Predicted Mask

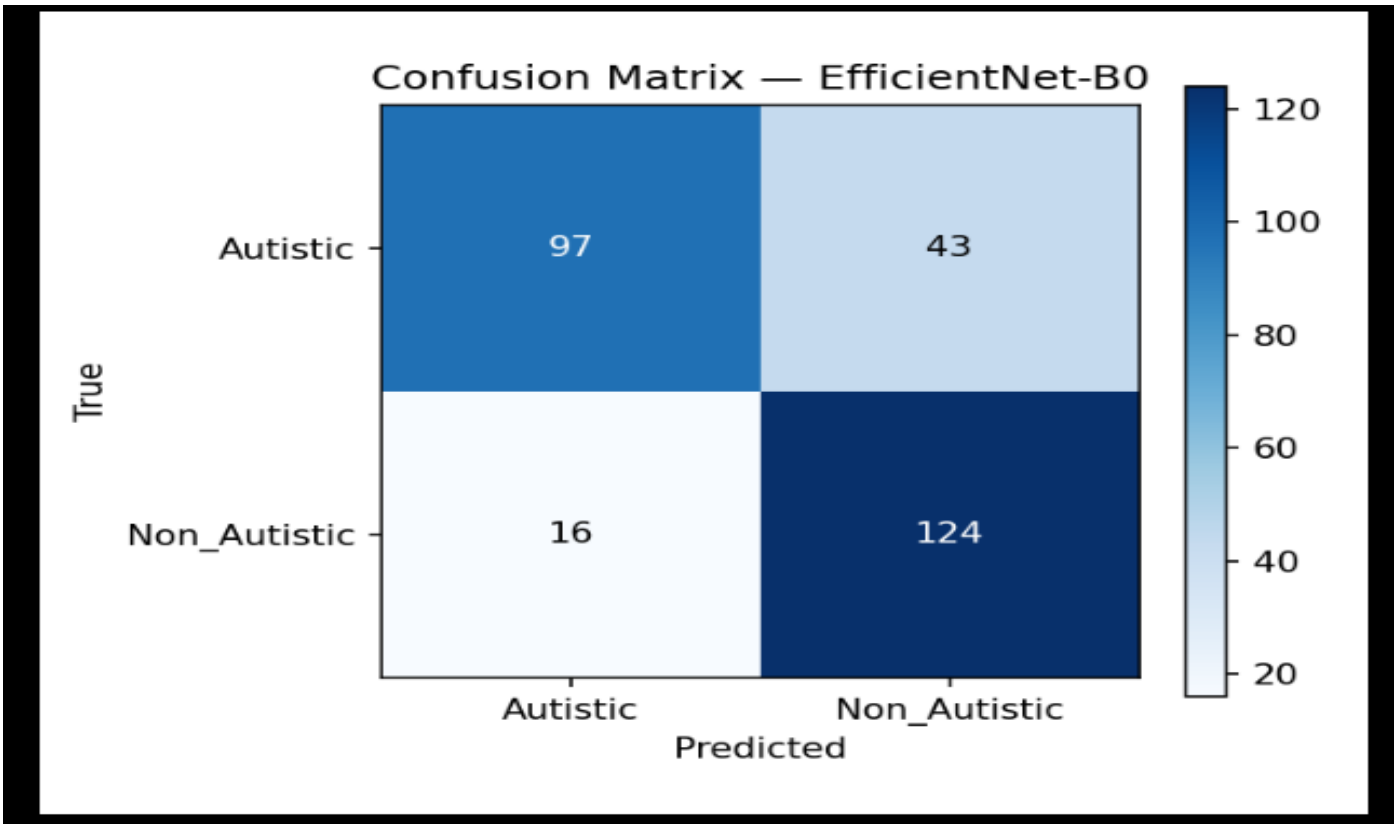


Figure 9 Confusion matrix 1

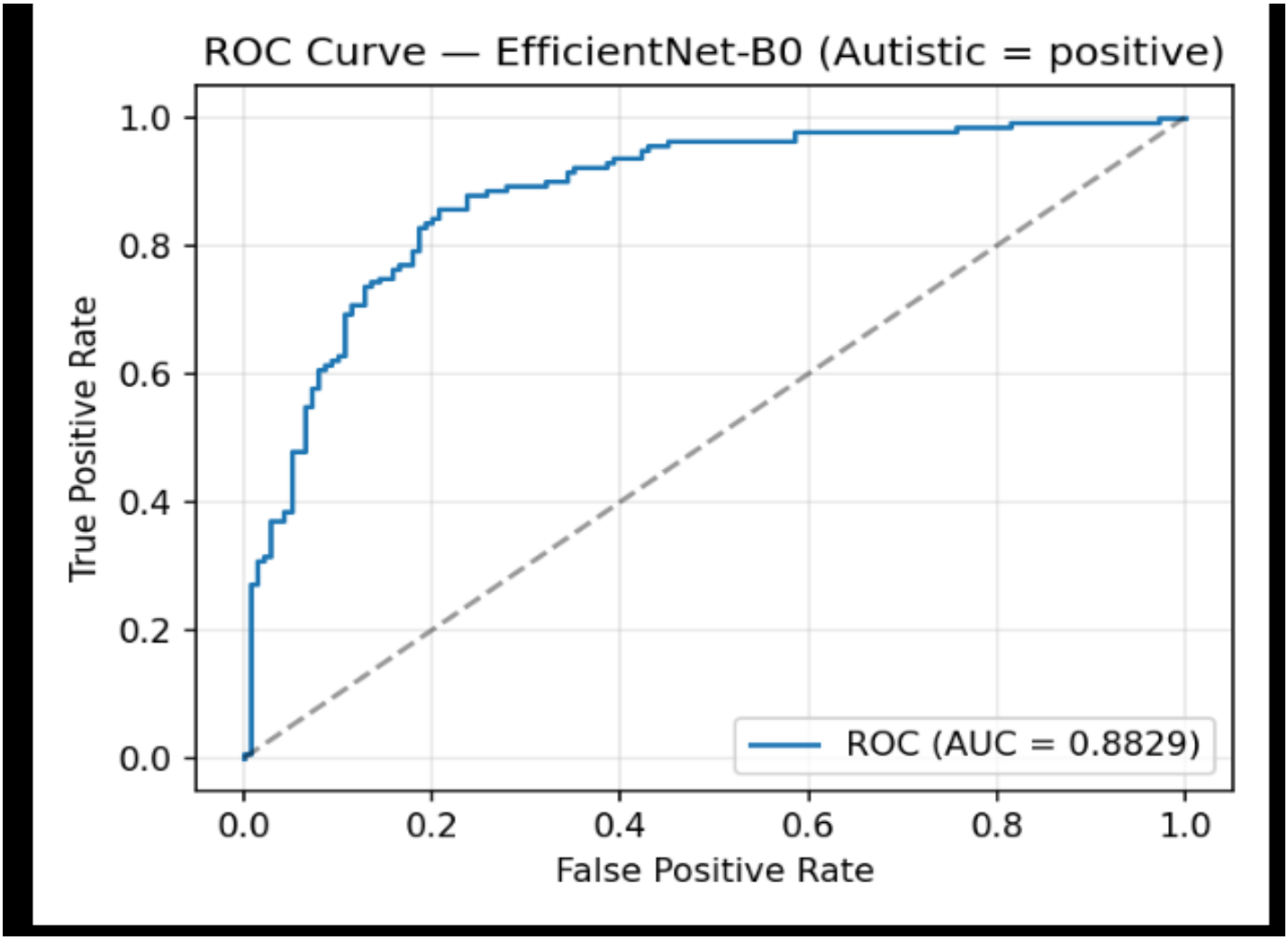


Figure 10 ROC Curves for CNN Face Mode

CHAPTER VI

CONCLUSION AND FUTURE WORK

Our hybrid autism screening system successfully created a comprehensive diagnostic framework that identifies autism risk and sorts severity levels through both facial image analysis and parent questionnaires. The ResNet-50 CNN model delivered excellent discriminative power with outstanding recall for catching autistic cases from photos alone. Meanwhile, the calibrated Logistic Regression handled four severity classes reliably without major imbalances.

By integrating deep learning for faces with traditional machine learning for behavioral data, the project overcomes weaknesses of single-method screening tools. Parents now get faster results. Screening becomes more consistent. Early detection reaches more families through simple web access. The system proves practical for real-world use while maintaining medical accuracy.

Several clear next steps will strengthen the tool further:

1. **Dataset Expansion**

Include diverse Pakistani ethnic groups, real clinical cases beyond synthetic data, and edge conditions like poor lighting or unusual poses. Current 5,000 synthetic samples tuned development—live validation comes next with IRB approval.

2. **Advanced Data Fusion**

Better combine face scores + questionnaire results. Current simple gating works well. Weighted ensemble or attention mechanisms could capture interactions between visual/behavioral signals missed today.

3. **Architecture Upgrades**

Test transformer-based vision models alongside CNNs. EfficientNet-V2 or Vision Transformers might squeeze extra performance from limited real-world photos. Questionnaire side? Gradient boosting ensembles over single Logistic Regression.

4. **Clinical Deployment**

Run large-scale validation across Topi hospitals and rural clinics. Measure false positive rates against specialist diagnoses. Track if early flagging actually speeds interventions.

5. **User Experience**

Build native mobile app wrapping the web tool. Add offline mode for rural Pakistan (local CNN inference, questionnaire scoring). Voice-guided Urdu questionnaires for low-literacy parents.

6. **Scalability**

Containerize for cloud deployment (AWS/GCP). Auto-scale FastAPI workers during peak screening hours. Add multi-language support beyond English/Urdu.

The foundation exists. These enhancements transform screening from research prototype to healthcare standard.

GLOSSARY

Term	Description
Facial Image Analysis	The process of analyzing facial features using deep learning models to identify patterns associated with autism.
Autism Severity Classification	The categorization of individuals into different severity levels (e.g., mild, moderate, severe) based on behavioral and questionnaire data.
ResNet-50	A deep learning model based on Residual Networks, used for extracting complex features from facial images for classification tasks.
ROC-AUC	A performance metric that measures the model's ability to distinguish between classes, where higher values indicate better classification performance.
Recall (Sensitivity)	A metric that measures the proportion of actual positive cases (autistic individuals) correctly identified by the model.
Logistic Regression (Calibrated)	A machine learning algorithm used for classification, where calibration improves the accuracy of predicted probabilities.
Macro-F1 Score	An evaluation metric that calculates the average F1-score across all classes, treating each class equally.
CNN (Convolutional Neural Network)	A type of deep learning model commonly used for image processing tasks such as facial feature extraction and classification.
Cross-Entropy Loss	A loss function used to measure the difference between predicted probabilities and actual class labels during model training.
Softmax Activation	A function that converts outputs into probabilities for multi-class classification.

Table 5:Glossary

REFERENCES

➢ **Bouchouras, G., & Kotis, K. (2025).** Integrating artificial intelligence, Internet of Things, and sensor-based technologies: A systematic review of methodologies in autism spectrum disorder detection. Algorithms, 18(1), 34. https://doi.org/10.3390/a18010034

➢ **Perochon, S., Di Martino, J. M., Carpenter, K. L. H., Compton, S., Davis, N., Eichner, B., Espinosa, S., Franz, L., Krishnappa Babu, P. R., Sapiro, G., & Dawson, G. (2023).** Early detection of autism using digital behavioral phenotyping. Nature Medicine, 29, 2489–2497. https://doi.org/10.1038/s41591-023-02574-3

➢ **Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021).** Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. Scientific Reports, 11, Article 15069. https://doi.org/10.1038/s41598- 021-94378-z

➢ **Ran, R., Liang, W., Deng, S., et al. (2024).** Risk assessment and automatic identification of autistic children based on appearance. Scientific Reports, 14, 29074. https://doi.org/10.1038/s41598-024-80459-2

➢ **Nawghare, P., & Prasad, J. R. (2024).** Early detection of autism spectrum disorder using AI and machine learning models: A systematic review for effective intervention. In Proceedings of the 2024 IEEE Pune Section International Conference (PuneCon) (pp. 1– 6). IEEE. https://doi.org/10.1109/PuneCon63413.2024.10895404

➢ **Talaat, F. M. (2023, May 10).** Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Computing and Applications. https://doi.org/10.1007/s00521-023-08372-9

➢ **Talaat, F. M., Ali, Z. H., Mostafa, R. R., et al. (2024, February 14).** Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. Soft Computing. https://doi.org/10.1007/s00500-023-09477- y

APPENDIX A

Data Survey:

- We had consistent meetings with our Mentor i.e.Dr Shahab Ansari, where he guided us throughout the process and helped us manage the project.
- We also had 2 meetings with Agha Khan Hospital, where we consulted with two doctors to help us better understand the dental images and what feature they wanted within the system.

Code:

This is the code for one of our segmentation model

https://colab.research.google.com/drive/1apTy2oXB-5gPdv2zO_by4hZ7iXwNCbsZ?usp=sharing#scrollTo=c018ec41-7c53-4fad-bc2e-f1a8b019d975