COMMUNITY BASED INTELLIGENCE

Early Health Incident Detection System for Sudan

Project Requirements & Design Document

Version 1.0 - MVP Specification

January 2026

1. Executive Summary

Community Based Intelligence (CBI) is a multi-agent AI system designed to revolutionize health incident reporting in Sudan. The system addresses a critical bottleneck in Sudan's disease surveillance infrastructure: the severe shortage of health officers available to receive, process, and respond to community health reports.

The system leverages WhatsApp as the primary communication channel, enabling community members to report health incidents through natural conversation with an AI agent. Reports are automatically structured, classified, and routed to health officers who can then coordinate rapid response efforts.

Core Value Proposition:

- Enable one health officer to effectively monitor multiple states simultaneously
- Reduce time from incident occurrence to official notification
- Capture structured data from unstructured, stressed conversations
- Provide actionable intelligence for disease outbreak prevention

1.1 MVP Hypothesis

This MVP is designed to validate a specific hypothesis:

> *"Community members in Sudan will proactively report health incidents through WhatsApp when guided by an AI agent, and this data—even when incomplete—can be structured and classified quickly enough to enable faster outbreak response than the current phone-based manual system."*

1.2 Success Criteria

Metric	Description	Target
Adoption	Community members actively using the system to report	Qualitative validation
Data Quality	AI captures usable data from unstructured conversations	70%+ reports actionable
Actionability	Health officers can make decisions based on reports	Officers find reports useful

2. Problem Statement

2.1 Current Situation

Sudan faces a critical challenge in early disease detection and response. The current system relies on community members calling health officers directly to report incidents. This approach has several fundamental limitations:

- Severe shortage of health officers to answer incoming calls
- Officers can only handle one call at a time, creating bottlenecks
- Manual data entry from phone calls is slow and error-prone
- No structured data collection leads to inconsistent reporting
- Delayed notification to authorities hampers rapid response
- Diseases like Cholera and Dengue Fever can spread rapidly before intervention

2.2 Target Diseases for MVP

The MVP focuses on four priority areas based on their prevalence and public health impact in Sudan:

Disease	Transmission	Why Priority
Cholera	Waterborne (contaminated water/food)	Rapid spread, high mortality without treatment, preventable
Dengue Fever	Vector-borne (Aedes mosquitoes)	Increasing incidence, urban spread, seasonal patterns
Malaria	Vector-borne (Anopheles mosquitoes)	Endemic, high burden, requires cluster detection
Clustered Deaths	Various/Unknown	Early warning signal for unknown outbreaks

3. System Architecture

3.1 High-Level Architecture

The system employs a multi-agent architecture built on LangGraph with Agent Swarm orchestration. Three specialized agents work in coordination to collect, classify, and analyze health incident data.

Architecture Flow:

WhatsApp Gateway → Reporter Agent → Surveillance Agent → Health Officer Dashboard

↓

Analyst Agent (on demand)

3.2 Agent Overview

Agent	Primary Role	Key Responsibilities
Reporter Agent	Data Collection	Conversational intake, intent detection, MVS capture, reporter guidance
Surveillance Agent	Classification & Routing	Alert classification, threshold monitoring, case linking, officer notification
Analyst Agent	Analysis & Visualization	Database queries, trend analysis, visualization generation, officer support

3.3 Agent Handoff Rules

Reporter → Surveillance:

- Triggered after reporter confirms the summary of their report
- Includes all captured data fields plus conversation metadata
- Partial reports are flagged but still forwarded for evaluation

Surveillance → Analyst:

- Automatically triggered when alert threshold is exceeded
- On-demand when health officer requests analysis or asks questions
- For situation summary generation to accompany notifications

4. Reporter Agent Specification

4.1 Purpose

The Reporter Agent is the primary interface between community members and the health surveillance system. It handles all incoming WhatsApp conversations, detects health-related reports through natural conversation, and collects essential data while maintaining an empathetic but efficient communication style.

4.2 Design Principles

Principle	Decision	Rationale
Conversation Style	Empathetic but concise	Avoid frustrating "customer service bot" feel; respect reporter stress/urgency
Question Strategy	One at a time, adaptive	Reduce cognitive load; adapt based on what reporter shares
Incomplete Data Handling	Accept and flag	Any information is better than none; don't block stressed reporters
Language Support	Arabic and English	Primary languages in Sudan; auto-detect from first message
Voice Notes	Not supported (MVP)	Future consideration; text-only for initial validation

4.3 Conversation Tone Examples

Target tone (correct):

> *"I'm sorry to hear this. Can you tell me where this happened?"*
>
> *"Got it. How many people are affected so far?"*
>
> *"Thanks. We've notified the health team. Keep the patient hydrated and isolated if possible. We'll update you soon."*

Too verbose (avoid):

> *"Thank you so much for reaching out to us today. I'm really sorry to hear that you're going through this difficult situation. I want you to know that I'm here to help you..."*

4.4 Conversation Flow

Phase 1: Greeting & Intent Detection

- User sends any message → Agent responds naturally

- Agent operates in "Listening Mode" - conversational but evaluating for health signals
- LLM intelligence (not keyword matching) determines when to activate "Investigation Mode"

Phase 2: Core Data Collection (MVS)

- What: Symptoms, disease description, or event type
- Where: Location (village, district, landmark - accept vague)
- When: Timing ("since yesterday", "few days ago" - accept imprecise)
- Who: Number affected, relationship to reporter
- Symptoms: Observable symptoms (optional but encouraged)

Phase 3: Clarification (only if critical)

- Only ask follow-up if critical information is ambiguous
- Example: "You mentioned Kassala - is this in the city or a village nearby?"

Phase 4: Confirmation & Guidance

- Summarize captured information for reporter confirmation
- Confirm that health officer will be notified
- Provide immediate general guidance (hydration, isolation, etc.)

Phase 5: Follow-up Channel

- Confirm best contact method for follow-up
- Promise to keep reporter updated on response actions

4.5 Intent Detection: Open World Approach

The Reporter Agent uses LLM intelligence rather than keyword matching to detect reportable health events. This enables natural conversation flow where reports can emerge organically.

Operating Modes:

Mode	Behavior	Trigger to Switch
Listening Mode	Conversational, responsive, constantly evaluating for health signals	Health signal detected → Investigation
Investigation Mode	Gently steers conversation to collect MVS data	Summary confirmed → Handoff to Surveillance

Signals That Trigger Investigation Mode:

- Mention of symptoms (vomiting, diarrhea, fever, bleeding, rash)
- Mention of disease names (cholera, malaria, dengue)
- Mention of deaths
- Quantifiers combined with sickness ("three people sick", "many cases")
- Location combined with health concern

Signals That Should NOT Trigger Investigation:

- Educational queries: "What are cholera symptoms?"
- Past events: "I had malaria last year"
- News/general discussion: "I heard there's disease in another country"
- Non-firsthand reports: Rumors without personal connection

4.6 Example Conversation

User: Hi
Agent: Hi, how can I help you today?
User: Things are bad here. No clean water for days.
Agent: That sounds really difficult. Is everyone in your area okay health-wise?
User: Actually my neighbor's children have been sick. Vomiting and diarrhea since yesterday. Three of them.
Agent: [TRIGGERS INVESTIGATION MODE] That's concerning. Are they in the same household?
User: Yes same house
Agent: Which area is this? Village or neighborhood name?
User: Wad Madani, Al Gezira state
Agent: Got it. Are there others nearby with similar symptoms, or just these three?
User: Just them for now

Agent: Thanks. So three children in the same household in Wad Madani with vomiting and diarrhea since yesterday. Is that correct? [Awaits confirmation before handoff]

4.7 Connectivity Handling

Given Sudan's unreliable connectivity, the system handles interrupted conversations gracefully:

- All messages are recorded immediately upon receipt
- Conversation state persists across disconnections
- When reporter reconnects, agent summarizes previous context
- Example: "You mentioned your friend has cholera. When did that happen?"
- Agent reassures accuracy of previous information before continuing

5. Surveillance Agent Specification

5.1 Purpose

The Surveillance Agent performs structured evaluation of incoming reports to determine nature, urgency, and required response. It classifies alerts using established epidemiological frameworks (EBS, CBS, CEBS), monitors thresholds, and generates actionable notifications for health officers.

5.2 Core Responsibilities

- Classify incoming reports by disease type, urgency, and alert category
- Monitor case counts against Ministry of Health defined thresholds
- Link related cases from same geographic area (intelligent deduplication)
- Generate notifications for health officers with appropriate urgency level
- Produce investigation instructions tailored to suspected disease
- Generate follow-up questions for community members when data is missing
- Request analytical support from Analyst Agent when thresholds are exceeded

5.3 Report Classification

Data Completeness Tiers:

Tier	Data Present	System Behavior
Actionable	Disease/symptoms + Location + Recency confirmed	Route to officer, full notification
Partial	2 of 3 above, or one is vague	Log, attempt one clarification, route with "incomplete" flag
Insufficient	Only 1 element or not current/local	Log for records, don't alert officer, thank reporter

Alert Type Classification:

Alert Type	Definition	Example
Suspected Outbreak	Multiple linked cases exceeding threshold	5 cholera cases in same village within 7 days
Cluster	Related cases below outbreak threshold	3 cases of fever in neighboring households
Single Case	Isolated case of notifiable disease	1 suspected cholera case (still reportable)

Rumor	Unverified report requiring investigation	"I heard people are sick in the next village"

5.4 Urgency Classification

Level	Criteria	Expected Response	Notification
Critical	Any cholera, hemorrhagic fever, or cluster of deaths	< 6 hours	Dashboard + WhatsApp/Email alert
High	Threshold exceeded for notifiable disease	< 24 hours	Dashboard + alert
Medium	Single cases, unclear reports needing verification	< 48 hours	Dashboard only

5.5 Disease Thresholds (MoH Defined)

Note: Actual threshold values to be provided by Ministry of Health. Structure shown below:

Disease	Alert Threshold	Outbreak Threshold	Notes
Cholera	1 suspected case	3+ cases same area within 7 days	Single case is always reportable
Dengue Fever	5 cases/week in one locality	20+ cases/week	Adjust for endemic baseline
Malaria	Cluster detection (above seasonal baseline)	Significant deviation from expected	Requires historical comparison
Clustered Deaths	2+ unexplained deaths same area	5+ deaths same area	Unknown cause triggers investigation

5.6 Case Linking Logic

The Surveillance Agent intelligently links related reports to identify potential outbreaks:

Linking Criteria:

- Geographic proximity: Same village, district, or within defined radius
- Temporal proximity: Reports within 7-day window
- Symptom similarity: Matching or related symptom profiles
- Disease match: Same suspected disease or compatible symptoms

Example:

Day 1: Reporter A reports 3 children with vomiting in Kassala

Day 2: Reporter B reports 2 adults with diarrhea in Kassala

→ System links these as potentially related (5 cases, same area, similar symptoms)

→ Threshold exceeded → Critical alert generated

5.7 Health Officer Notification Content

When thresholds are met, the health officer receives a structured notification containing:

Notification Package:

- Urgency Level: Critical / High / Medium with color coding
- Suspected Disease: Based on symptom analysis and reporter information
- Location: Normalized location with map reference if available
- Case Summary: Number of cases, deaths, and affected population
- Timeline: When symptoms started, when reported
- Related Cases: Links to other reports in same area if applicable

5.8 Investigation Instructions

For each notification, the Surveillance Agent generates disease-specific investigation guidance:

- Questions for health officer to ask during field investigation
- Medical tests required and recommended sample quantities
- Personal protective equipment (PPE) requirements
- Additional data to collect from community members
- Follow-up questions to send to original reporter(s)

6. Analyst Agent Specification

6.1 Purpose

The Analyst Agent provides data analysis, visualization, and natural language database access capabilities. It serves both the Surveillance Agent (automated analysis) and health officers (interactive queries).

6.2 Core Capabilities

For Surveillance Agent:

- Generate situation summaries when thresholds are exceeded
- Provide contextual data for notifications (historical comparison, trends)
- Estimate potential spread based on disease characteristics (using LLM general knowledge)

For Health Officers:

- Natural language database queries ("Show me all cholera cases in Kassala last month")
- Trend analysis and visualization generation
- Heat map creation showing geographic distribution
- Report and dashboard building on demand

6.3 MVP Scope

Feature	MVP Status	Notes
Situation summaries	Included	Text-based summaries for notifications
Database queries (natural language)	Included	Agent writes and executes queries
Trend analysis	Included	Cases this week vs. last week, etc.
Visualization generation	Included	Agent writes code to generate charts
Heat maps	Included	Geographic distribution visualization
ML Predictive modeling	Deferred	Requires historical data; use LLM estimates for MVP
Advanced dashboards	Deferred	Basic visualization sufficient for validation

6.4 Spread Estimation (MVP Approach)

For MVP, the Analyst Agent uses LLM general knowledge to provide spread estimates based on disease characteristics:

- Cholera: High spread potential via contaminated water; can affect entire communities rapidly
- Dengue: Limited to mosquito vector range; urban areas more vulnerable
- Malaria: Endemic patterns; seasonal variations; localized spread

These estimates are clearly marked as preliminary and subject to officer verification with real-world data. They help officers prioritize but do not replace clinical judgment.

7. Data Model

7.1 Report Entity

Each health incident report is stored with the following structure:

Field	Type	Required	Description
id	UUID	Yes	Unique identifier
created_at	Timestamp	Yes	When report was created
updated_at	Timestamp	Yes	Last modification time
status	Enum	Yes	open, investigating, resolved, false_alarm
reporter_phone	String	Yes	Contact number (for follow-up)
reporter_relationship	Enum	No	witness, family, health_worker, community_leader, unknown
symptoms	String[]	No	List of observed symptoms
suspected_disease	Enum	No	cholera, dengue, malaria, unknown
location_text	String	Yes	Raw location as provided
location_normalized	String	No	Standardized location
location_coords	Object	No	Latitude/longitude if available
onset_date	Date	No	When symptoms started
onset_text	String	No	Raw timing as provided
cases_count	Integer	No	Number of cases reported
deaths_count	Integer	No	Number of deaths reported
affected_description	String	No	Description of affected population
urgency	Enum	No	critical, high, medium, low
alert_type	Enum	No	suspected_outbreak, cluster, single_case, rumor
conversation_id	String	Yes	Link to full chat history
data_completeness	Float	Yes	0-1 score of data quality
flags	String[]	No	incomplete_location, needs_verification, etc.
linked_reports	UUID[]	No	Related report IDs

officer_assigned	String	No	Assigned health officer
investigation_notes	String	No	Officer notes
outcome	String	No	Final resolution

7.2 Privacy Considerations

- Reporter identity limited to phone number for follow-up capability
- No additional personal information collected
- Data retention policy to be defined based on MoH requirements
- Conversation history retained for audit trail and ML training

8. User Flows

8.1 Community Reporter Flow

1. Community member sends message via WhatsApp
2. Reporter Agent engages in natural conversation
3. Health signal detected → Investigation Mode activated
4. Agent collects MVS data through adaptive questioning
5. Agent presents summary for confirmation
6. Reporter confirms → Report handed to Surveillance Agent
7. Reporter receives confirmation + immediate guidance
8. Reporter kept updated on response actions

8.2 Health Officer Flow

9. Officer receives notification on web dashboard (+ alert for critical cases)
10. Reviews case details: urgency, location, suspected disease, case count
11. Views investigation instructions and required actions
12. Can query Analyst Agent for additional context/analysis
13. Assigns case and initiates field investigation
14. Updates case status based on real-world findings
15. Closes case with outcome documentation

8.3 Feedback Loop to Reporter

The system maintains communication with reporters to build trust and encourage future reporting:

- Immediate: Confirmation that report was received
- Immediate: General health guidance appropriate to suspected condition
- On Investigation: Notification that team has been dispatched (if applicable)
- On Resolution: Summary of actions taken
- Follow-up: Request for additional information if needed

9. Health Guidance Protocol

9.1 Guidance Scope

The Reporter Agent provides general health guidance that does not require medical expertise. This helps reporters take immediate protective action while awaiting official response.

Permitted Guidance:

- Isolation: Keep sick person in separate room if possible
- Hydration: Encourage fluid intake, especially for diarrhea/vomiting
- Hygiene: Wash hands frequently, especially after contact
- Sanitation: Use separate bathroom facilities if available
- Monitoring: Watch for worsening symptoms
- Seeking care: Recommend visiting health facility if symptoms severe

Not Permitted:

- Medication recommendations
- Diagnosis statements
- Treatment protocols
- Any guidance requiring medical expertise

9.2 Medical Review Requirement

All health guidance templates must be reviewed and approved by a medical professional before deployment to ensure safety and appropriateness.

10. Technical Decisions

10.1 Technology Stack

Component	Technology	Rationale
Agent Framework	LangGraph	Supports complex multi-agent workflows with state management
Orchestration	Agent Swarm	Enables dynamic agent coordination and handoffs
Messaging Gateway	WhatsApp Business API	Most accessible channel in Sudan
LLM Provider	TBD	Requires Arabic language support evaluation
Database	TBD	Needs to support geospatial queries
Dashboard	Web-based	Accessible from any device with browser

10.2 Key Technical Decisions

Decision	Choice	Alternative Considered	Reason
Intent Detection	LLM Intelligence	Keyword matching	Handles natural conversation; adapts to context
Data Collection	Flexible/Adaptive	Structured forms	Accommodates stressed reporters; partial data acceptable
Language Detection	Auto-detect	Ask user	Reduces friction; seamless experience
Voice Support	Deferred to post-MVP	Include in MVP	Adds complexity; text sufficient for validation
Predictive ML	Deferred to post-MVP	Include in MVP	Requires training data; LLM estimates sufficient
MoH Integration	Standalone (MVP)	Direct integration	Reduces dependencies; faster validation

11. MVP Scope Summary

11.1 Included in MVP

- Reporter Agent: Full conversation flow with intent detection, MVS collection, confirmation
- Surveillance Agent: Classification, threshold monitoring, case linking, notifications
- Analyst Agent: Database queries, basic visualization, situation summaries
- Health Officer Dashboard: Notifications, case details, investigation instructions
- Reporter Feedback: Confirmation, guidance, status updates
- Languages: Arabic and English
- Diseases: Cholera, Dengue, Malaria, Clustered Deaths
- Alerts: Dashboard + WhatsApp/Email for critical cases

11.2 Deferred to Post-MVP

- Voice note support
- ML-based predictive modeling
- MoH system integration (DHIS2)
- Advanced escalation protocols
- Multi-officer assignment and case routing
- Additional disease categories
- Mobile application (dashboard)

11.3 Out of Scope

- Medical diagnosis or treatment recommendations
- Direct patient care coordination
- Supply chain management for medical supplies
- Full epidemiological modeling

12. Future Considerations

12.1 Post-MVP Features

Pending successful validation, the following features should be prioritized:

- Voice note transcription for reporters with limited literacy
- ML models trained on collected data for improved predictions
- Integration with national health information systems
- Escalation workflows for unresponsive officers
- Multi-language support (regional dialects)
- Offline-capable mobile dashboard for field officers

12.2 Scalability Considerations

- Geographic expansion to additional states
- Additional disease categories based on MoH priorities
- Multiple communication channels (SMS, Telegram, local apps)
- Integration with laboratory systems for test result tracking

12.3 Success Metrics for Scaling Decision

Metric	MVP Target	Scale Threshold
Reporter adoption	Qualitative validation	>50 reports in pilot period
Data quality (actionable)	70%+ actionable	>80% actionable reports
Officer satisfaction	Positive feedback	Confirmed time savings
Response time improvement	Faster than phone	>50% reduction in time-to-notification
False positive rate	Acceptable	<20% false alarms

Appendix A: Glossary

Term	Definition
CBI	Community Based Intelligence - this project
MVS	Minimum Viable Signal - essential data points for actionable report
EBS	Event-Based Surveillance - detection of unusual health events
CBS	Community-Based Surveillance - community reporting of health events
CEBS	Community Event-Based Surveillance - combined approach
MoH	Ministry of Health
DHIS2	District Health Information System - common health data platform
PPE	Personal Protective Equipment

Appendix B: Open Questions

The following questions require resolution before or during implementation:

16. Exact threshold values from MoH for each disease
17. Geographic unit for reporting (village vs. district resolution)
18. Data retention policy and privacy compliance requirements
19. Medical review of health guidance templates
20. LLM provider selection (Arabic language capability)
21. Pilot location and target reporter population
22. Health officer onboarding and training approach
23. Data validation strategy for incoming reports

— End of Document —