Leveraging Local Somali Radio Stations for Real-Time Food Security Insights: An AI-Powered Approach

Preliminary Phase

1. Introduction

1.1 Background

Somalia continues to face severe food insecurity challenges due to recurring droughts, political instability, and armed conflict that have disrupted agricultural production across the country. Traditional food security monitoring systems often suffer from data gaps, delayed reporting, and limited coverage in conflict zones, creating critical information barriers for humanitarian response. The lack of timely and accurate data prevents early warning systems from functioning effectively, especially in remote and insecure regions where vulnerability to food crises is highest. Humanitarian organizations struggle to allocate resources efficiently without reliable, up-to-date information on local conditions, market prices, and population movements. These information gaps can lead to delayed interventions, allowing food security situations to deteriorate to emergency levels before adequate assistance reaches affected communities.

1.2 Research Question

"Do local radio stations offer useful new information to improve food security indicators' accuracy and/or timeliness?"

Local radio stations in Somalia could potentially provide crucial real-time information about food security conditions that traditional monitoring systems miss due to data gaps, delayed reporting, and limited access to conflict zones. By tapping into community broadcasts through an AI-powered approach, humanitarian organizations might gain timely insights about local agricultural conditions, market prices, and population movements in remote areas, potentially enabling faster and more effective interventions before food security situations deteriorate to emergency levels.

It all depends on the specific activities of radio stations and what they cover in their radio broadcasts; we will try to check if there are programs that can provide us with such information.

1.3 Objectives

This preliminary phase evaluates whether Somali radio broadcasts can serve as a novel data source for enhancing food security indicators through an AI-powered pipeline. I am identifying relevant radio stations, establishing data collection methods, and testing speech-to-text technologies specifically optimized for Somali language. After completing the 4-week plan, the results will include:

- List of Somali radio stations with relevance to food security.

- Sample data of audio recordings.
- Transcriptions using speech-to-text tools.
- Notebook with working code for testing.
- Summary report with insights and recommendations.

---

2. AI-Powered Pipeline for Extracting Food Security Insights

2.1 Proposed Workflow

Radio Broadcast Collection → Speech-to-Text (STT) Conversion → Large Language Model (LLM) Processing (translation and summarizing) → Food Security Insights Integration

The proposed four-step AI-powered workflow begins with collecting radio broadcasts from Somali stations, converting speech to text using specialized tools, processing the transcribed content through Large Language Models, and finally extracting actionable food security insights also using LLM. In this preliminary phase, I will focus on the first two steps: finding appropriate radio stations and broadcasts, establishing methods for radio broadcast collection and testing speech-to-text conversion capabilities optimized for the Somali language.

3. Four-Week Plan for a Preliminary Phase

During the first week, I conducted research on Somali radio stations to evaluate their relevance to food security. I compiled a list of online stations, identified available live streaming options, and assessed whether they covered important topics such as food security, agriculture, or climate conditions.
In the second week, I focused on evaluating the feasibility of accessing radio broadcasts and collecting sample audio data. I developed a script to download and store the samples of recordings from one of the stations, setting the stage for subsequent analysis. The prototype was developed using Python due to its capabilities for simplicity and also for fast implementation of speech-to-text models. To facilitate access, testing speed, and free GPU access, the prototype was created on Google Colab with data storage on Google Drive.
By the third week, I employed speech-to-text tools like Whisper to transcribe the recordings. I also evaluated transcription accuracy and compared different models for Somali language processing to determine the most reliable option for our specific needs.
In the fourth week, I synthesized all my findings and outlined comprehensive recommendations for further research. I provided detailed insights into potential next steps for the project, ensuring a clear path forward for continued investigation into how Somali radio broadcasts can contribute to food security monitoring and response.

---

4. A Preliminary List of Online Somali Radio Stations Categorized by Relevance to Food Security

4.1 Overview of relevant radio stations

Below is a list of radio stations in Somalia that provide information related to food security and those that potentially may provide such information. Only stations accessible via the internet have been included. The stations have been divided into three main categories: Highly Relevant to Food Security which have dedicated programs for this purpose, Moderately Relevant to Food Security which have covered food security topics in the past, and Less Relevant to Food Security where information related to food security appears sporadically. Each radio station has a brief overview, a link to its website, a link to its archive or live stream, its connection to food security, and its coverage area.

4.2 Highly Relevant to Food Security (Dedicated Programs on Food Security)

Radio Ergo

- **Overview**: A humanitarian radio station in Somalia delivering programs on health, farming, and livestock to support local communities.
- **Website**: Radio Ergo
- **Archive**: Radio Ergo Archive
- **Food Security**: Provides comprehensive coverage on food security, featuring news and discussions on agriculture, climate conditions, market prices, and humanitarian aid updates.
- **Coverage area**: Nationwide in **Somalia**, including **rural and hard-to-reach areas** via **shortwave radio** and **local partner stations**.

Radio Daar Dheer

- **Overview**: Community-based station in Dhusamareb, supported by the United Nations, covering news, health, education, and culture.
- **Website**: Radio Daar Dheer
- **Live Broadcast**: Listen Live
- **Archive**: Radio Daar Dheer Archive
- **Food Security Program**: **"Barnaamijka Horumarinta Wax Soo saarka Beeraha iyo Xoolaha"** ("Program on the Development of Agricultural and Livestock Production").
- **Coverage area**: Dhusamareb and surrounding areas.

Isnaay Radio

- **Overview**: Community radio station in Jowhar, Middle Shabelle, delivering news, entertainment, and music.
- **Website**: Isnaay Radio
- **Archive**: Isnaay Radio Archive
- **Food Security Program**: **"Barnaamijka Hormarinta Wax Soosaarka Beeraha"** ("Program on Agricultural Production Development").
- **Coverage area**: Jowhar and surrounding areas.

4.3 Moderately Relevant to Food Security (Occasionally Covers Food Security Topics)

Radio Waamo

- **Overview**: Community radio station in Kismayo covering local news and educational programs.
- **Website**: Radio Waamo
- **Live**: Listen Live
- **Archive**: Radio Waamo Archive
- **Food Security Program**: Occasionally covers food security topics through climate change discussions and business programs.
- **Coverage area**: Kismayo and surrounding regions.

Radio Galmudug

- **Overview**: State-run broadcaster in Galmudug focusing on news, education, and cultural programs.
- **Website**: Radio Galmudug
- **Live**: Listen Live
- **Archive**: Galgaduud TV Archive
- **Food Security Program**: Participates in educational content on agriculture, livestock, and nutrition.
- **Coverage area**: Galmudug region, including Dhusamareb and Galkayo.

Radio Kismaayo

- **Overview**: State-run radio station of Kismaayo, Jubaland, covering news, education, and culture.
- **Website**: Radio Kismaayo
- **Live**: Listen Live
- **Food Security Program**: Occasionally covers food security through news reports and discussions on drought resilience.
- **Coverage area**: Kismayo and surrounding areas.

Radio Daljir

- **Overview**: Established in 2003, the largest FM network in Somalia operating in multiple locations.
- **Website**: Radio Daljir
- **Live**: Listen Live
- **Food Security**: Reports on food security topics but does not have dedicated programs.
- **Coverage area**: Puntland, Galmudug, eastern Somaliland, Somali region of Ethiopia.

4.4 Less Relevant to Food Security (Minimal to No Regular Coverage)

Radio Arlaadi

- **Overview**: Community engagement radio station in Baidoa.
- **Website**: Radio Arlaadi
- **Live**: Listen Live
- **Archive**: Radio Arlaadi Archive
- **Food Security Program**: No dedicated programs, but occasionally covers related topics.
- **Coverage area**: Baidoa and surrounding regions.

Radio Sanguuni

- **Overview**: Community radio in Dhobley focusing on local issues.
- **Website**: Radio Sanguuni
- **Archive**: Radio Sanguuni Archive
- **Food Security**: No dedicated programs but covers occasional topics in broader discussions.
- **Coverage area**: Dhoobley and surrounding regions.

Radio Dalsan

- **Overview**: Leading Somali news station with social and political discussions.
- **Website**: Radio Dalsan
- **Live**: Listen Live
- **Food Security Program**: Occasionally covers related topics but has no dedicated program.
- **Coverage area**: Mogadishu, Jowhar, Baidoa, Adado, Wanlaweyn.

Radio Mogadishu (Radio Muqdisho)

- **Overview**: Established in 1951, Somalia's state-run broadcaster.
- **Website**: Radio Muqdisho
- **Live**: Listen Live
- **Food Security**: Occasionally participates in food security initiatives.
- **Coverage area**: Mogadishu and surrounding areas.

4.5 Conclusions and choice

Based on the radio stations listed above, the choice was simple - Radio Ergo was selected as the primary source for our food security monitoring prototype. Radio Ergo offers several distinct advantages that make it ideal for this research:

1. It provides comprehensive coverage on food security topics, featuring dedicated programs on agriculture, climate conditions, market prices, and humanitarian aid updates.
2. Its nationwide reach across Somalia, including rural and hard-to-reach areas via shortwave radio and local partner stations, ensures broad geographical coverage.
3. The station has an accessible online archive of past broadcasts, particularly their farming programs, making historical data collection feasible.
4. As a humanitarian radio station explicitly focused on delivering programs related to health, farming, and livestock to support local communities, Radio Ergo's content directly aligns with our research objectives.

5. The consistency and reliability of their food security programming creates a more standardized dataset for our AI-powered speech-to-text analysis pipeline.

These factors combined make Radio Ergo the most pragmatic starting point for testing our hypothesis about extracting valuable food security insights from Somali radio broadcasts. While other stations like Radio Daar Dheer and Isnaay Radio also offer dedicated food security programs, Radio Ergo's broader coverage and more extensive archive provide optimal conditions for our initial prototype development.

---

5. Data Collection

5.1 Radio Ergo

Radio Ergo is a specialized humanitarian radio service that broadcasts critical information across Somalia and Somali-speaking regions through a network of 25 local correspondents, covering essential topics like health, agriculture, displacement, and education. Through its programming, Radio Ergo amplifies the voices and stories of ordinary Somalis, including farmers, pastoralists, women, and the displaced, enabling them to participate in public discussions about humanitarian crises and development challenges. The station employs innovative communication channels, including shortwave transmission and the Freedom Fone call-in platform, allowing listeners in remote areas to both receive vital information and provide feedback on local challenges, which is then analyzed and shared weekly with humanitarian organizations to inform their response efforts.
Also, a great thing about Radio Ergo is that on their website you can find officially translated English documentation of conversations with affected/vulnerable people. Their updates from 2022 provided critical humanitarian information and documented firsthand accounts from affected communities across Somalia during a particularly challenging period. The 2022 updates covered a range of pressing issues including drought conditions, disease outbreaks, displacement, and conflict situations, offering valuable insights directly from local Somali correspondents based in regions most impacted by these crises.

I think this is another great resource which could be used for Real-Time Food Security Insights.

5.2 Availability of Radio Ergo Broadcasts

Radio Ergo provides an archive of past broadcasts, which can be accessed through their official website: Radio Ergo Archive and on their **SoundCloud profile**: Radio Ergo on SoundCloud.

After multiple attempts to access the audio files directly from the frontend, I discovered that the broadcasts are stored on **SoundCloud**, making it possible to download them directly from there. However, Python libraries such as **yt-dlp, requests,** and **BeautifulSoup4** did not work for direct downloads due to site restrictions.

To address this, I developed a function that successfully retrieves these files. The function, along with other related functionalities, is available in an easily accessible **Google Colab notebook**: Google Colab Notebook. This prototype allows for easy testing and verification of the file download process.

5.3 Implementation of Audio File Downloading

I implemented a function to **download audio files from Radio Ergo's SoundCloud profile** based on a **specified date range**. The function works as follows:

- It scrapes the SoundCloud profile page and extracts track URLs.
- It filters the extracted URLs based on the **broadcast date** using **regular expressions**.
- It downloads the matching **MP3 files** using **yt-dlp**.
- If running in **Google Colab**, it can optionally save the downloaded files to **Google Drive** for easy access.

For a step-by-step guide on how to download the broadcasts, refer to the **Google Colab notebook** linked above.

---

6. Implementing Speech-to-Text for Somali Language Recordings

6.1 Initial Approach: Standard Whisper Model

The implementation begins with OpenAI's Whisper model, which offers multi-language support. The function:

- Processes MP3 files automatically from a specified directory
- Creates and organizes transcript files systematically

However, evaluation reveals a critical issue: while Whisper detects Somali language, it incorrectly transcribes it using Arabic script instead of Latin script (which Somali uses in writing). The transcription consists almost entirely of the repeated Arabic word "موضوع" (meaning "subject"), making the output completely unusable.

6.2 Specialized Somali Model

Then I found and implemented a fine tuned version of Whisper model specifically for somali language form Hugging Face [steja/whisper-small-somali] which shows significant improvement:

Key advantages:

- Correctly transcribes in Latin script as used in written Somali
- Produces recognizable Somali vocabulary and content related to drought conditions, humanitarian aid, and local initiatives
- Successfully identifies proper nouns, locations, and contextual terms

6.3 ElevenLabs: Advanced Somali Transcription with Scribev1

Following the limitations encountered with Whisper-based models, the implementation of the **ElevenLabs Scribev1** model represents a **breakthrough** in Somali speech-to-text transcription.

Key Advantages of Scribev1

The Scribev1 model delivers superior transcription quality for Somali audio recordings, especially from humanitarian media like **Radio Ergo**:

- **High-fidelity Latin script transcription** – Faithfully reflects the Somali writing system, avoiding script errors
- **Accurate capture of idiomatic expressions and named entities** – Retains subtle linguistic and cultural meaning
- **Extremely low hallucination rate** – Eliminates the repetitive artifacts observed in previous models
- **Excellent domain alignment** – Effectively handles broadcast themes like health, food security, religion, and pastoralist culture

6.3 Evaluation Results

After conducting initial tests, it is clear that transcription of Somali language audio remains challenging for general-purpose models.

Metric	Whisper Base	steja/whisper-small-somali	ElevenLabs Scribev1
Script accuracy	❌ Arabic	✅ Latin	✅ Latin + fluent
Repetition errors	Severe	Moderate	Minimal
Named entity recognition	Poor	Good	Excellent
Domain vocabulary precision	Low	Moderate	High
Sentence coherence	Fragmented	Partially coherent	Fully coherent

While Whisper models show some promise, hallucinations like repeated phrases (“iyo iyo iyo”, “dhul dhul dhul”) persist even in specialized variants. ElevenLabs Scribev1, in contrast, offers a drastic quality improvement, with only 8 repetitions—many of which reflect natural repetition in Somali speech.

This evaluation confirms the need for **language-specific model development**. General-purpose models often underperform on low-resource languages like Somali, making dedicated model fine-tuning essential.

6.4 Key Food Security-Related Topics Identified in the Transcription

Highlights from the **March 15th Radio Ergo** broadcast reveal serious food security concerns:

- Families are traveling **up to 40 kilometers** for water, paying as much as **seven dollars per barrel**, forcing many into debt

- Testimony from **Abdullahi Mohammed Ali** describes dried-up water sources and dependence on distant, unreliable water points
- Livestock—key to the pastoralist economy—are dying due to lack of water and forage
- Warnings about imminent **"macluul" (famine)** conditions dominate the broadcast
- Humanitarian programs (nutrition, mobile health, feeding initiatives) are mentioned, but appear **insufficient given the scale of need**
- Crop failures from multiple missed rainy seasons have **depleted food reserves** and driven up commodity prices

This illustrates the **immense value** of local radio as a source of real-time, ground-level data for monitoring food insecurity.

---

7. Challenges & Considerations

7.1 Data

A major advantage in Somalia is the existence of **Radio Ergo**, which provides **accessible, high-quality recordings**. Without it, analyzing food security trends would require a more labor-intensive approach, such as:

- Identifying famine periods from historical records
- Manually collecting and transcribing broadcasts from those timeframes
- Evaluating correlations between broadcast themes and real-world food security indicators

7.2 Speech-to-Text Models

As highlighted earlier, the main technical bottleneck is the **lack of highly accurate Somali transcription models**. While progress has been made, options remain limited. Future exploration should include:

- Additional fine-tuned models from Hugging Face
- Collaborative development with researchers focused on low-resource languages
- Potential custom training or fine-tuning using aligned Somali speech-text datasets

---

8. Conclusion

This preliminary phase demonstrates both the **promise** and **obstacles** of using AI for food security monitoring through Somali radio broadcasts.

- The system for downloading and organizing broadcast data from **Radio Ergo** works well
- However, transcription accuracy remains the **primary barrier** to extracting reliable insights
- While ElevenLabs Scribev1 outperforms existing models, **native speaker validation** remains essential to confirm qualitative improvements

Overcoming this technical hurdle is critical to unlocking the potential of local radio as a real-time, community-driven sensor for humanitarian monitoring.

---

9. Recommended Next Steps

1. **Find or fine-tune speech-to-text models** with improved performance on Somali language
2. **Evaluate model output with native speakers** to ensure linguistic and cultural fidelity
3. Once a reliable model is confirmed, **scale up transcription experiments** using larger Radio Ergo datasets and test integration with food security indicators
4. Explore existing **documentation and transcripts** on ReliefWeb
5. **Engage with Radio Ergo, SIDA, and DANIDA** to explore potential collaborations and data-sharing opportunities

By advancing speech-to-text solutions and forging local partnerships, we can move toward **real-time monitoring tools** for food insecurity based on voices from the ground.
