

AI Voice-to-Token Triage: Optimizing Patient Flow at Aravind Eye Hospital

Prithvi Parthasarathy
August 1st, 2025



Patient Flow Challenges in General OPD

Guided Entry

Invisible Queue

Station Circuit

- **Escorted by MLOP:** After registration, every patient must wait for an MLOP to physically guide them from the lobby to the General Unit.
- **Uncertain Waiting:** Patients sit in a crowded hall with no visual queue indicator for individual stations—clinical care is essentially a “black box”
- **Name Call-Out System:** MLOPs call out names when it’s time to attend next station; soft-spoken calls or noise make many patients anxious about missing their turn; name callouts can also be ambiguous when multiple occurrences exist
- **Complaint Voiced Late:** The first full discussion of symptoms happens only after all stations, inside the doctor’s room, limiting targeted triage earlier in the visit.

Opportunity: AI-Guided Kiosk Triage



Kiosk Function

- Captures **Tamil speech** at entry, auto-transcribes to English within seconds
- AI maps complaint → produces **tailored station list**
- Prints a clear token + on-screen queue position at each station—**no name call-outs needed**



Patient Benefits

- **Self-directed flow:** patient knows exactly where to go next
- **Visible queue status:** reduces anxiety, supports elderly & hard-of-hearing
- **Faster visit**—time saved by avoiding unnecessary stations



Clinic Benefits

- **MLOPs freed** from escort & shout duties; focus on clinical tasks
- **Shorter bottlenecks:** low-acuity cases fast-tracked, senior doctors see the right patients sooner
- Structured data captured for **real-time analytics** (wait-time dashboards, pathway audits)

Research Snapshot: NLP & LLM Tools for Clinical Triage


The Use of a Kiosk-Model Bilingual Self-Triage System in the Pediatric Emergency Department

Sinha, Madhumita MD, MHSM[†]; Khor, Kai-Ning MPH[†]; Amresh, Ashish PhD[†]; Drachman, David PhD[†]; Frechette, Alan MD^{*}

Author information 

Pediatric Emergency Care 30(1):p 63-68, January 2014. | DOI: 10.1097/PEC.0000000000000037

Comparison between the Smart Triage model and the Emergency Triage Assessment and Treatment guidelines in triaging children presenting to the emergency departments of two public hospitals in Kenya

Stephen Kamau, Joyce Kigo, Paul Mwaniki, Dustin Dunsmuir, Yashodani Pillay , Cherri Zhang, Brian Nyamwaya, David Kimutai, Mary Ouma, Ismael Mohammed, Keziah Gachuhi, Mary Chege, Lydia Thurairara, J Mark Ansermino, Samuel Akech

Pilot study of ASSORT (AI-based Symptom Stratification in Ophthalmology for Rapid Triage): a triage tool for ophthalmic emergencies

Claudio Xompero , Lorenzo Rossi , Francesca Amoroso, Antonio Bechara Ghobril, Diana Elena Ionita, Eric H. Souied, Carl-Joe Mehanna ^{*} 

Department of Ophthalmology, Centre Hospitalier Intercommunal de Créteil, Créteil, France

- **AI models rival clinicians:** Across high-volume centres, machine-learning and large-language-model triage tools regularly match—or out-perform—junior doctors while meeting accepted safety thresholds.
- **Voice-first kiosks cut front-end delay:** Self-service, speech-guided check-in systems typically maintain data quality and score high on patient usability—even in low-literacy groups.
- **Free-text NLP boosts prediction everywhere—including LMICs:** Tablet-based implementations in resource-constrained hospitals show the approach is feasible and accurate at the bedside.

V0 Prototype Overview

Why build a prototype?

- Provides **proof-of-feasibility** before presenting to IT team
- Allows us to understand baseline metrics in order to implement possible Phase-1 study KPIs
- Can give a starting point for physician-guided inputs and clinical improvement

What does V0 address?

- Can we accurately record, store, and analyze patient complaint audio?
- What is a possible source of training data for an AI triage model?
- What are the challenges associated with verbal triage?

What is in V0?

- Takes Tamil audio recordings and converts to English text
- Trains AI classifier model on 4130 patient complaints sourced from EMR (May records)
- Provides skeleton for backend JSON elements
- Determines accuracy of a basic AI rule engine

AI Complaint Classifier (How it Works)

Data Cleaning

Using EMR data for 1 week in May, complaint entry for each patient was cleaned: shorthand expansion (“c/o” → “complains of”), unification of terms (“reduced vision” = “blurred vision”), etc.

Complaint Assignment

Each EMR record was assigned a matched complaint (validated through GPT) based on keywords present in data

LinearSVM Algorithm

Keywords that appeared in rare scenarios (“redness”, “pain”) were weighted heavily, while common words (“it”, “patient”) were given low weight

Model Training

Based on what keywords were present in the EMR complaint, common phrases in data were correlated with an assigned complaint value

Code Snippets

EMR Data Cleaning

```
ABBREV_MAP = {  
    # common ophthalmic shorthand expansions  
    "c/o": "complains of ",  
    "h/o": "history of ",  
    "no h/o": "no history of ",  
    "re": "right eye ",  
    "le": "left eye ",  
    "od": "right eye ",  
    "os": "left eye ",  
    "ou": "both eyes ",  
    "va": "vision ",  
    "iop": "intraocular pressure ",  
    "dov": "diminution of vision ",  
    "fb": "foreign body ",  
    "glare": "glare ",  
    "photophobia": "photophobia ",  
    "watering": "watering ",  
    "redness": "redness ",  
    "pain": "pain ",  
    "itching": "itching ",  
    "irritation": "irritation ",  
    "discharge": "discharge ",  
    "floaters": "floaters ",  
    "flashes": "flashes ",  
    "trauma": "trauma ",  
    "injury": "injury ",  
    "chemical": "chemical ",  
    "post op": "postoperative ",  
    "post-op": "postoperative ",  
}
```

Tamil speech-to-text via Whisper Model

```
import whisper  
model = whisper.load_model("medium")  
#English translation from Tamil speech  
res_en = model.transcribe(audio_wav_path, task = "translate", language = "ta")  
print("English translation:", res_en["text"])
```

Assigned complaints to EMR data

```
expanded_keywords = [  
    'blurred vision', 'blurred', 'blurring', 'dim vision',  
    'eye pain', 'ocular pain', 'pain',  
    'redness', 'red eye',  
    'irritation', 'burning', 'stinging',  
    'itching', 'itchy',  
    'watering', 'tearing', 'discharge',  
    'floaters', 'black spots',  
    'flashes', 'flashing lights',  
    'photophobia', 'light sensitivity',  
    'glare', 'halos',  
    'drooping eyelid', 'ptosis', 'drooping',  
    'twitching', 'eyelid twitching',  
    'dryness', 'dry eye', 'dry eyes',  
    'headache',  
    'foreign body', 'foreign body sensation',  
    'deviation of eye', 'squint', 'eye turn',  
    'bulging eye', 'proptosis', 'bulging',  
    'night blindness',  
    'double vision',  
    'lasik',  
    'general', 'review'  
]
```

EMR Complaint	Assigned Complaint Value
c/o persistence of redness re since 1 month.	Redness
patient complains of re leaf injury 2 days ago followed by re discharge	Discharge
came for power check up	Check-up

Sample Token Generation (General Check-up)



```
resp = openai.audio.translations.create(  
    model = "whisper-1",  
    file = AUDIO_PATH.open("rb"),  
    response_format = "text",  
    temperature = 0.0,  
)  
english_text = resp  
print(english_text)
```



I came here for a regular check-up and after 3 years this is my first check-up.



```
# 1) Complaint text  
sample = "I came here for a regular check-up and after 3 years this is my first check-up."  
  
# 2) Predict Token Function  
label, conf, top = predict_token(sample)  
  
print("Predicted label :", label)  
print("Confidence      :", round(conf, 3))
```



Predicted label : check_up
Confidence : 0.94

TOKEN 08	
Aravind Eye Hospital – General OPD	
Date: 30-Jul-2025	Time: 10 : 15 AM
MRN : AEX77844	
Name : _____	
Age : _____	Sex : _____
Complaint : GENERAL CHECK-UP	
Assigned Doctor : Dr. S. Meena (Gen. Ophthalmology)	
PLEASE VISIT STATIONS IN THIS ORDER	
1. Vision Assessment (VA)	Room VA-2
2. Doctor Examination (Slit-Lamp)	Room SL-10
3. Follow-Up Desk / Next Steps	Counter F-1
Priority : Routine	
Instructions	
• Show this ticket at each station and wait to be called.	
• இந்த சீட்டினை ஒவ்வொரு நிலையிலும் காட்டவும்.	
• உங்கள் பெயர் அழைக்கப்படும் வரை காத்திருக்கவும்.	



This general check-up patient can be routed to a specific fast-track line with a junior doctor

Sample Token Generation (Irritation)



```
resp = openai.audio.translations.create(  
    model      = "whisper-1",  
    file       = AUDIO_PATH.open("rb"),  
    response_format = "text",  
    temperature = 0.0,  
)  
english_text = resp  
print(english_text)
```



There was a small irritation in my eye. I thought it was because of the increased power. That's why I came to check at Aravind Hospital.



```
# 1) Complaint text  
sample = "There was a small irritation in my eye. I thought it was because of the increased power. That's why I came to check at Aravind Hospital."  
  
# 2) Predict Token Function  
label, conf, top = predict_token(sample)  
  
print("Predicted label :", label)  
print("Confidence      :", round(conf, 3))
```

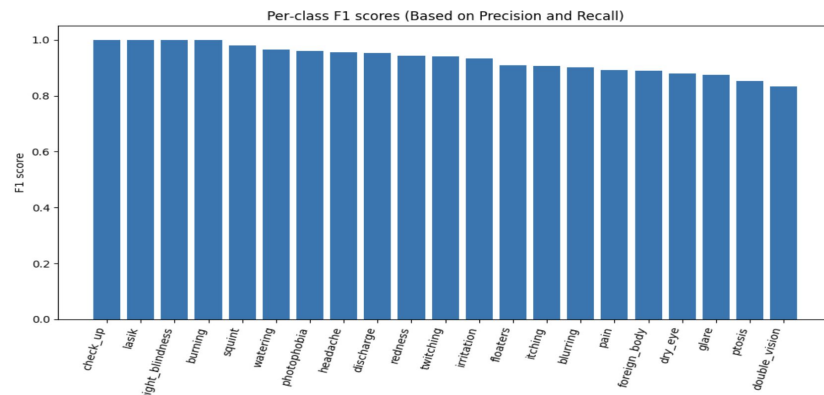
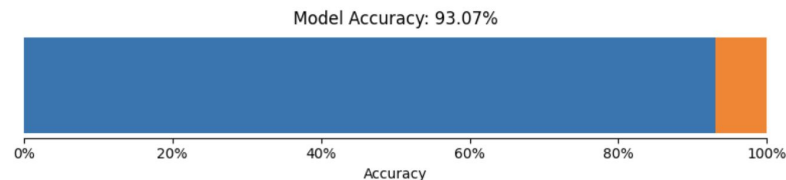
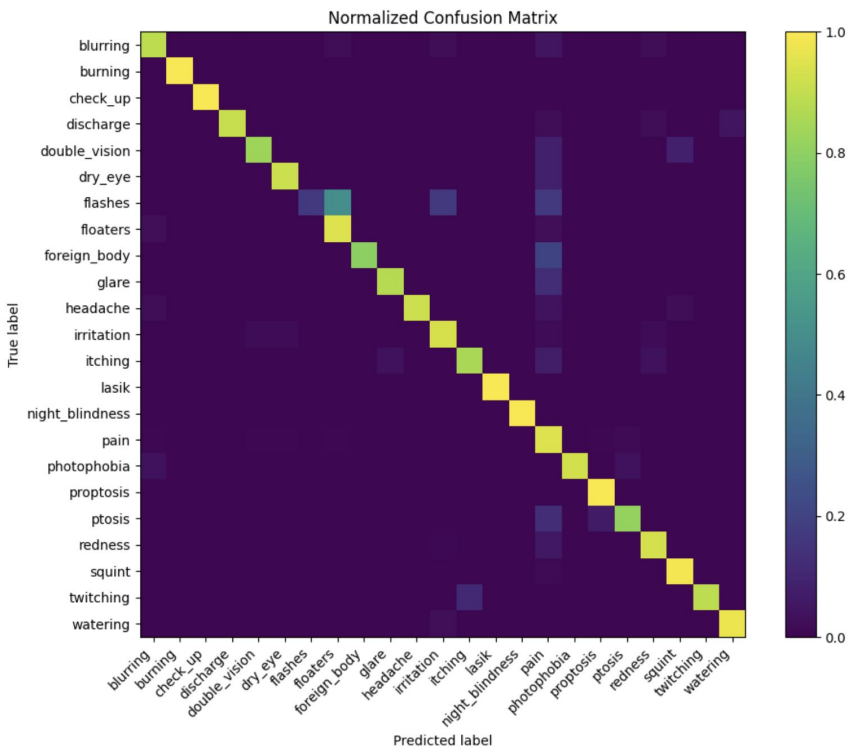


Predicted label : irritation
Confidence : 0.921

TOKEN 27	
Aravind Eye Hospital – General OPD	
Date: 30-Jul-2025	Time: 11 : 28 AM
MRN : BAH56231	
Name : _____	
Age : _____	Sex : ____
Complaint : IRRITATION	
Assigned Doctor : Dr. R. Kumar (Gen. Ophthalmology)	
PLEASE VISIT STATIONS IN THIS ORDER	
1. Refraction	Room REF-3
2. Vision Assessment (VA)	Room VA-2
3. Intra-ocular Pressure (IOP)	Room IOP-1
4. Doctor Examination (Slit Lamp)	Room SL-14
5. Counselling / Pharmacy	Counter C-1
Priority : Routine	
Instructions	
• Show this ticket at each station and wait to be called.	
• இந்த சீட்டினை ஒவ்வொரு நிலையிலும் காட்டவும்.	
• உங்கள் பெயர் அழைக்கப்படும் வரை காத்திருக்கவும்.	



Model Performance Metrics



Backend Schema for IT Development

```
{
  "mrn": "AEX12345",
  "token_number": "76",
  "visit_id": "2025-07-30-000317",
  "created_ts": "2025-07-30T10:42:08+05:30",
  "language": "ta",
  "transcript_text": "இடது கண்ணில் சிவப்பு, கண்ணீர் வடிதல் மூன்று நாட்களாகத் தொடர்கிறது",
  "complaint_label": "redness",
  "pathway_version": "opd_rules_v0.7",
  "priority": "urgent",
  "risk_flags": ["photophobia"],
  "stations": [
    { "code": "VA", "room": "OPD-VA-2", "seq": 1 },
    { "code": "DOC_SLIT", "room": "OPD-2-12", "seq": 2 },
    { "code": "IOP", "room": "IOP-1", "seq": 3 }
  ],
  "human_confirmation": "TRUE",
  "confirmed_by": "ML0P187",
  "token_id": "TOK-7K49F",
  "audit": [
    { "who": "ML0P187", "action": "confirmed", "ts": "2025-07-30T10:42:30+05:30" }
  ]
}
```

- Every time the kiosk prints a ticket, this JSON object is stored / sent to backend system.
- It captures ID, complaint, routing logic, and audit fields in a single, standardized format.
- This allows for easy tracking of analytics and token performance

Phase 1 Pilot Study and Associated KPIs

4-5 week pilot study at AEH Thanjavur OPD (possibly followed in other centers based on results)



Accuracy

- Doctor-approved pathway matches token in **≥ 95 %** of cases

Adoption

- Staff override rate **≤ 10 %**
- Patient “knew next step” score **≥ 90%** (exit survey, n ≥ 50)

Data Yield

- **≥ 95 %** of tokens saved with full logs (label, confidence, timestamps)
- **Audit-ready logs:** Model + rule versions stored per token → enables KPI tracking and post-pilot retraining.

Efficiency

- MLOP confirmation time **≤ 10 s** (median)
- Wait-time for simple cases drops **≥ 30 %** vs baseline

Safety

- Missed urgent/red-flag cases = **0** (weekly audit)

Thank You!

