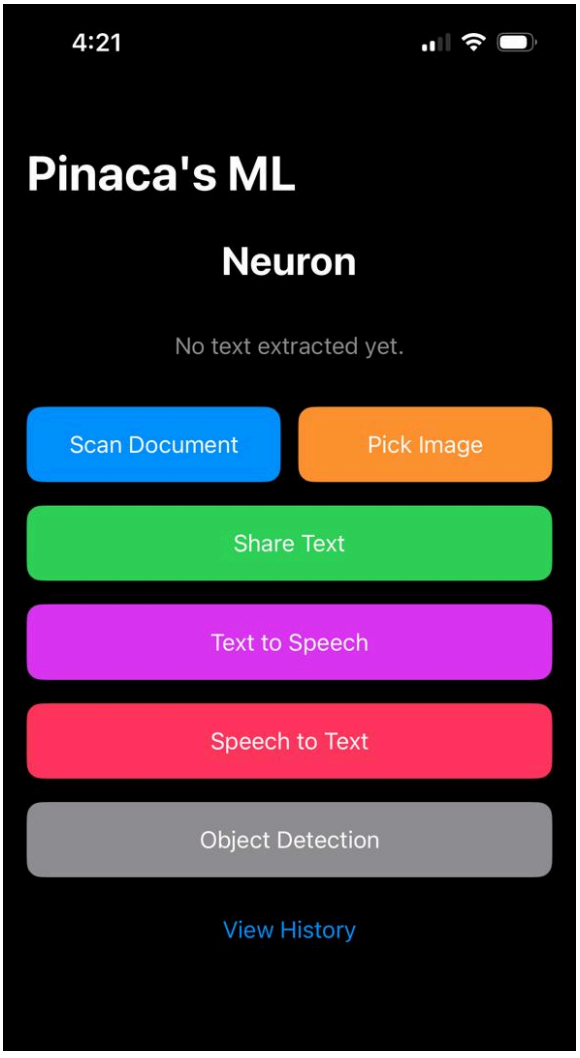
Apple’s On-Device Machine Learning Capabilities

Evaluation on iPhone 13 (iOS 18)

Objective:

To assess the effectiveness of Apple’s native, on-device machine learning (ML) capabilities in reducing latency, minimizing server dependency, and supporting real-time intelligent tasks directly on iOS hardware.

Device Tested: iPhone 13
OS Version: IOS18
Focus Areas: Optical Character Recognition (OCR), Text-to-Speech (TTS), Speech-to-Text (ASR), Object Detection



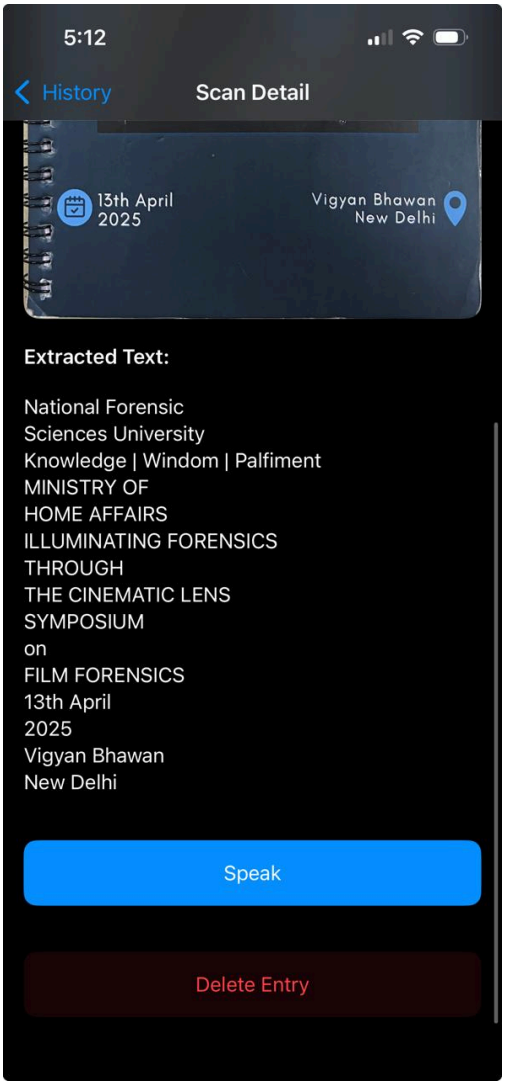
1. Optical Character Recognition (OCR)

Apple's OCR pipeline, built into the Vision framework, provides high-performance, low-latency text recognition through the VNRecognizeTextRequest class. This implementation supports both printed and handwritten content natively.

- Performance: OCR completes in under **0.5 seconds per image** with consistently high accuracy.
- Capabilities: **Supports multiple languages and handwritten content** without external APIs.
- Use Cases: Document scanning, data extraction, real-time camera overlays.

Framework: Vision
API/Class: VNRecognizeTextRequest
On-Device: Yes

Text extraction



Extracted Text output

2. Text-to-Speech (TTS)

Text-to-Speech is executed through the AVSpeechSynthesizer class within the AVFoundation framework. It supports a broad set of languages with automatic language detection and fast synthesis.

- Performance: **Real-time response, low-latency playback**.
- Accuracy: Effective in single-language contexts; **auto-detection generally reliable**.
- Limitations: Voices retain a **synthetic (robotic) tone; multilingual text synthesis reduces contextual clarity.**

Framework: AVFoundation
API/Class: AVSpeechSynthesizer, AVSpeechUtterance
On-Device: Yes
Use Cases: Accessibility tools, voice feedback, language learning apps.

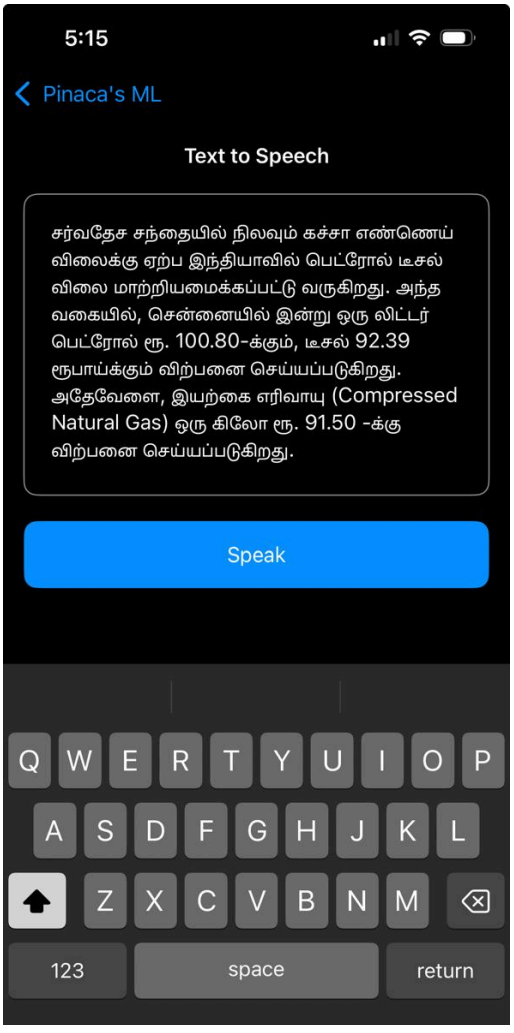
Supported TTS Languages

Indian Languages

Language	Code
Hindi	hi-IN
Tamil	ta-IN
Telugu	te-IN
Bengali	bn-IN
Marathi	mr-IN
Gujarati	gu-IN
Kannada	kn-IN
Malayalam	ml-IN
Urdu	ur-IN

Global Languages

Language	Code
English (US)	en-US
English (UK)	en-GB
French	fr-FR
German	de-DE
Spanish	es-ES
Japanese	ja-JP
Korean	ko-KR
Chinese	zh-CN



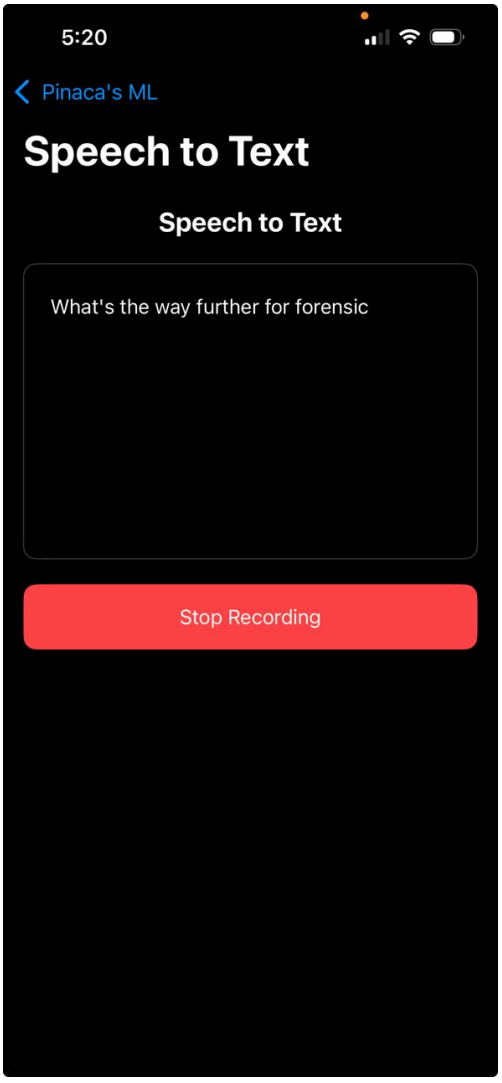
Speech to Text on Multilingual Inputs

3. Speech-to-Text (ASR)

Speech recognition is handled by the SFSpeechRecognizer class in the Speech framework. It performs reasonably well in global language contexts but lacks robustness for Indian languages and mixed-language inputs.

- Performance: Real-time **transcription in English and supported languages is accurate**.
- Limitations: **Multilingual audio and slang-heavy input frequently reduce reliability. Indian language support remains minimal.**
- Use Cases: Voice command recognition, live transcription, note dictation.

Framework: Speech
API/Class: SFSpeechRecognizer, SFSpeechAudioBufferRecognitionRequest
On-Device: Partially (some scenarios may fall back to server processing).



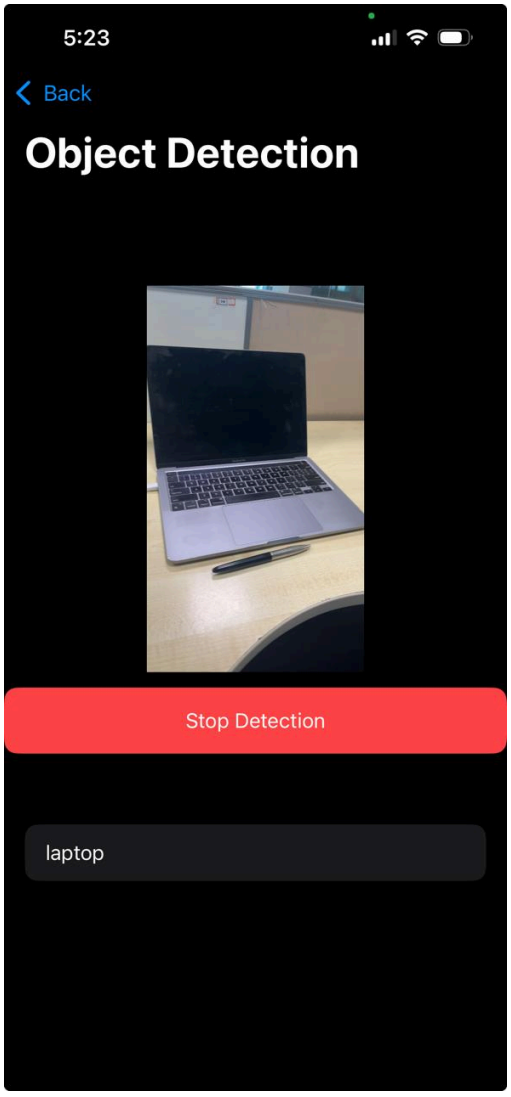
Speech to Text

4. Object Detection

Object detection is powered by CoreML and the Vision framework using VNCoreMLRequest. A YOLOv3 model is currently in use, offering practical object detection capabilities across common items.

- Performance: Detects most everyday objects with around **80% confidence**.
- Accuracy: Reliable for single-object detection; **performance degrades with multiple targets in frame.**
- Limitations: Causes noticeable **device heating; multitarget detection remains limited**.

Frameworks: Vision, CoreML
API/Class: VNCoreMLRequest + YOLOv3 CoreML model
On-Device: Yes
Use Cases: Smart camera, AR features, product identification



Object detection of basic products

Native ML Capabilities in IOS (on-device)

Capability	Framework	API/Class	Use Case Example
OCR	Vision	VNRecognizeTextRequest	Document scanning, text extraction
Text-to-Speech (TTS)	AVFoundation	AVSpeechSynthesizer	Accessibility, virtual assistants
Speech-to-Text (ASR)	Speech	SFSpeechRecognizer	Dictation, voice commands
Object Detection	Vision/CoreML	VNCoreMLRequest + YOLOv3	Real-time detection, camera apps
Face Detection	Vision	VNDetectFaceRectanglesRequest	Filters, facial tracking
Language Detection	NaturalLanguage	NLLanguageRecognizer	Auto-detect for dynamic content
Text Classification	NaturalLanguage	NLModel, NLTagger	Sentiment analysis, auto-tagging
Barcode Scanning	AVFoundation	AVCaptureMetadataOutput	QR code scanning, logistics
Pose Detection	Vision	VNDetectHumanBodyPoseRequest	Fitness apps, gesture recognition
Handwriting OCR	Vision	VNRecognizeTextRequest	Whiteboard capture, note digitization
Image Similarity	Vision/CoreML	VNGenerateImageFeaturePrintRequest	Visual deduplication, content match

Way Further

1. Utilize Apple Intelligence from Iphone 15 pro (A17 Bionic chip)'s native textual model (3B parameter) to content summarization, email completion, grammar check (helps in preprocessing) and other textual tasks.

2. To create custom models and onboard its performance in domain-specific tasks.