



manas

Every Mind deserves Care, because silence should never be a sentence

Describe Your Solution Through the Four Phases of Machine Learning

CTO Program Group 2

01 Understand Data

02 Make Predictions

03 Make Decisions

04 Causal Inference

CTO Program Group 2 presents innovative AI-driven mental health support solutions

The Four Phases of ML Applied to MANAS

How the complete ML framework powers the AI-driven mental health companion

01 Understanding Data

Voice: speech rate, pitch variability, pause frequency

Video: facial action units, micro-expressions, psychomotor cues

K-Means clustering by behavioral stress profile

PCA dimensionality reduction of voice/video feature space

Target: 5 validated signal categories

02 Making Predictions

Regression: continuous risk score (0–100) from voice/video inputs

Classification: intervention type needed (4-level triage)

RNNs: temporal deterioration patterns in speech streams

AUROC 0.70 → 0.85 | Lead-time: 6–12 months

03 Making Decisions

MPC: stable users with acoustic baselines established

Multi-Armed Bandit: new users, low voice/video history

MDP: user state shifts detected via prosodic change

Reinforcement Learning: novel cohorts, no prior data

Retention: 1.0× → 2.0×

04 Causal Inference

RCT across 3 pilot sites (500+ users)

T-test: validates voice/video intervention outcomes

A/B measurement for intervention intensity comparison

Null hypothesis testing against control group

Relapse reduction: ~25% (1.0 → 0.75 index)

PHASE 1 — Understanding the Data

Voice & Video as the Primary Signal · Data Exploration · Clustering

P1 FOM: 0 → 5 validated signal categories

Primary Signal Streams

VOICE

Speech rate, pitch variability, pause frequency, vocal energy decay, micro-tremor captured passively via smartphone microphone

VIDEO

Facial action units, eye contact frequency, blink rate, micro-expressions, psychomotor agitation/retardation via front-facing camera check-ins

TEXT

Sentiment polarity, hopelessness vocabulary, linguistic complexity decay layered over acoustic & visual signals

BEHAVIOR

App engagement cadence, response latency, sleep proxies contextualizes and validates voice/video findings over time

ML Techniques Applied

Histograms & Box Plots

Visualize acoustic feature distributions; detect sudden vocal energy drops (early-warning signal)

Scatter Plots

Correlate prosodic & facial features with clinical screening scores (PHQ-9, GAD-7)

PCA

Reduce high-dimensional voice/video feature space to 2D makes cluster patterns legible to clinicians

K-Means Clustering

Segment users by behavioral stress profiles population-specific baselines for accurate anomaly detection

PHASE 2 — Making Predictions

Regression · Classification · Neural Networks · Voice & Video as Primary Inputs AUROC: 0.70 → 0.85 | Lead-time: 0 → 6–12 months

Regression

Risk trajectory score

- Continuous risk score (0–100) from voice/video inputs
- MSE minimized; R^2 validates acoustic signal explanatory power
- LASSO: selects strongest prosodic & facial features, prevents overfitting on sparse rural data
- Predicts deterioration 6–12 months ahead — the core lead-time model

Classification

Intervention category

- Binary: intervention needed now (yes/no)
- Multi-class: nudge / peer support / referral / emergency
- Vocal stress severity + video psychomotor signals drive triage
- ROC curve tuned to minimize missed crises (false negatives)

Neural Networks

Temporal pattern detection

- RNNs: designed for sequential data, speech and video streams over time
- Hidden layers fuse acoustic embeddings + facial action unit sequences
- Detects gradual deterioration in vocal patterns over weeks, impossible with snapshot models
- On-device inference protects voice/video privacy (federated learning)

Key insight: RNNs see the trend over time — not just a snapshot. For voice & video, the trajectory IS the signal.

PHASE 3 — Making Decisions

MPC · Multi-Armed Bandit · Markov Decision Process · Reinforcement Learning

Target: Retention 1.0x → 2.0x

The 2x2 Decision Making Framework Applied to MANAS

Case 1 — Model Predictive Control

Fixed · High info

Stable users with 90+ days of voice/video history.
Uses acoustic baselines to schedule optimal micro-interventions.
Environment: static; data: abundant.

Case 2 — Multi-Armed Bandit

Fixed · Low info

New users with little to no voice/video history.
Balances testing new interventions vs. repeating proven ones.
Environment: stable; data: scarce.

Case 3 — Markov Decision Process

Dynamic · High info

User whose vocal patterns shift suddenly from stable baseline.
Models risk state transitions; selects best next action in real time.
Environment: dynamic; data: available.

Case 4 — Reinforcement Learning

Dynamic · Low info

New cohorts or geographies with no historical acoustic data.
Agent learns from engagement & wellbeing feedback over time.
Environment: dynamic; data: scarce.

PHASE 4 — Determining Causal Inferences

Did MANAS Actually Cause the Improvement?

Target: Relapse -25% (1.0 → 0.75 index)

A user's vocal patterns improve after a MANAS intervention. Did MANAS cause the recovery, or were they going to recover anyway?

Cause 1 (Desired)

MANAS detected early vocal stress patterns → triggered timely intervention → user recovered. The system caused the outcome. Confirms the technology is clinically valid and worth scaling.

Cause 2 (Confound)

User naturally recovered, or a new therapist / resolved life event drove improvement. MANAS acoustic signals correlated with outcomes but did not cause them.

Solution: Randomized Control Trial Design

3 Pilot Sites

University campus (200 students) + Veterans healthcare (150) + Frontline workers, each randomized into MANAS vs. control group

T-test Validation

Compare mean PHQ-9/GAD-7 scores between groups at 30, 60, and 90 days to test statistical significance

Null Hypothesis

H₀: MANAS voice/video-driven interventions have no causal effect on mental health outcomes vs. control

A/B Measurement

Parallel cohorts with different intervention intensities, attribute outcome differences to intervention variable

Summary: The Four Phases Applied to MANAS

Complete synthesis — voice & video as the primary signal thread across all phases

Phase	ML Concepts	MANAS Application	Key Metric	FOM Link
01 Understand Data	PCA, K-Means Clustering, Histograms, Scatter Plots	Voice acoustics + video facial cues as primary signals. Cluster users by stress profiles.	5 signal categories P1 FOM $t_0 \rightarrow t_1$	P1: 0→5 signals
02 Make Predictions	Regression (MSE/R ² /LASSO), Classification (Logistic/LDA/KNN/ROC), Neural Networks (RNN)	Risk trajectory from voice/video inputs. RNNs detect temporal deterioration over weeks.	AUROC 0.70→0.85 6–12 mo lead-time	P3: AUROC target
03 Make Decisions	MPC, Multi-Armed Bandit, Markov Decision Process, Reinforcement Learning	Framework selected based on acoustic data availability and user state dynamics.	Retention 1.0×→2.0×	P5: Retention KPI
04 Causal Inference	Randomized Control, Null Hypothesis, T-test, A/B Measurement	RCT across 3 pilot sites validates that voice/video interventions cause health improvement.	Relapse −25% 1.0→0.75 index	P5: Relapse KPI

MANAS Targets: Early detection +6–12 mo • Rural reach +30–40% • Retention 2× • Relapse −25%