

Breaking the 70% Accuracy Barrier in Voice-Based Depression Detection

Dan Vilenchik | Aya Biderman | Elissa Freedman | Temira Feinsilver

Ruslan Sergienko | Yuval Ratzabi | Julie Cwikel

Ben-Gurion University | Clalit Health Services South District, Beer-Sheva, Israel

Clinical Problem

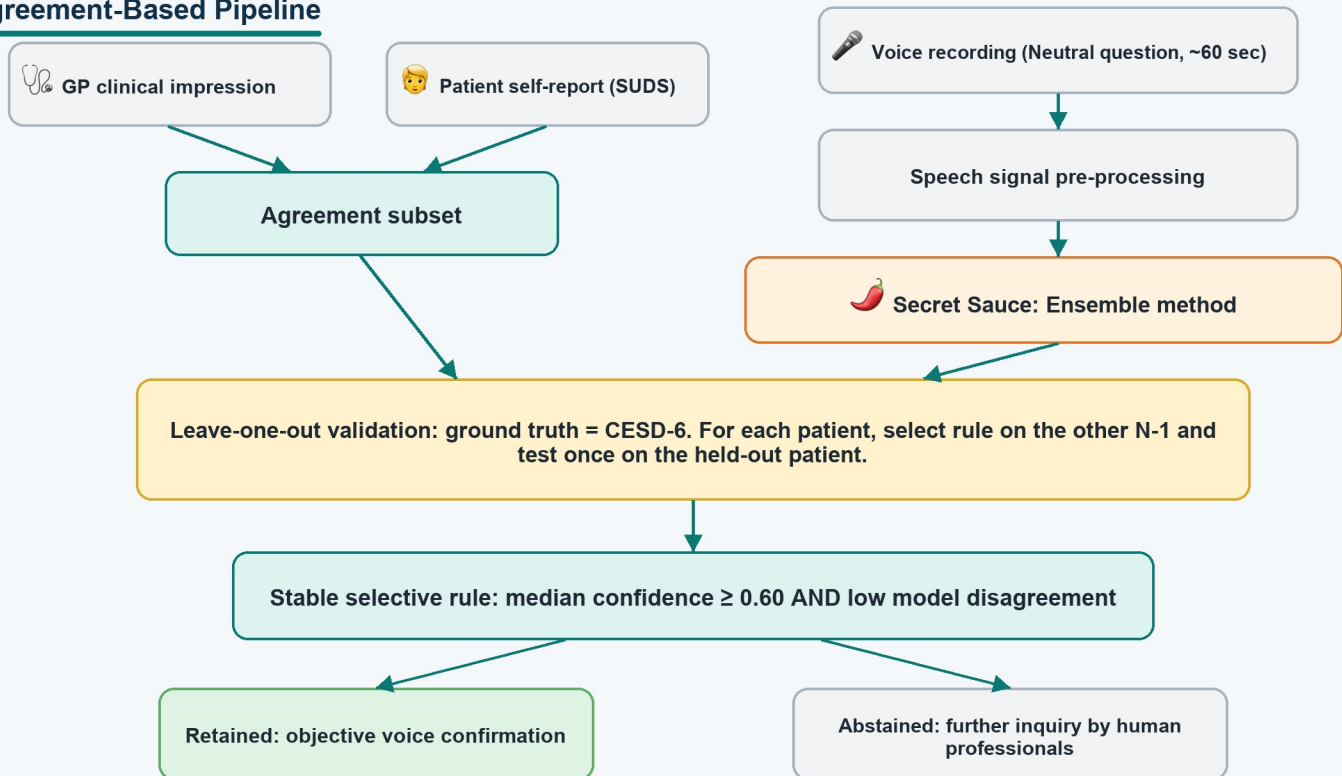
- No objective biomarker exists in psychiatry — diagnosis relies entirely on subjective clinician and patient ratings
- At 5–10% prevalence with ~70% voice-model accuracy, PPV is only ~20–25%: most positive screens are false alarms
- Non-compliance with depression treatment (medication or behavioral) is a major burden; uncertain diagnoses further reduce patient adherence

Reframing

A new way to use voice-based screening

- Activate only when GP and patient independently agree on subjective severity (same SUDS distress bucket: low / mid / high)
- Voice provides an objective signal to confirm — not override — the shared clinical impression
- When confidence is low, the system ABSTAINS — responsible AI by design**

Agreement-Based Pipeline



Agreement-Based Results

Metric	Result
Retained patients	21 / 59
Actual coverage	35.6%
Overall accuracy	20 / 21 = 95%
Depressed retained	5 / 6 = 83.3%
Non-depressed retained	15 / 15 = 100%
Statistical test	p = 0.017

Binomial test vs majority-class prior (67.8%); 59-patient pilot cohort.

Disagreement-Based Results

Retained	Coverage	Accuracy	Dep acc	Non-dep	p-val
8/31	25.8%	7/8 = 87.5%	3/4 = 75.0%	4/4 = 100%	0.062
15/31	48.4%	11/15 = 73.3%	5/8 = 62.5%	6/7 = 85.7%	0.118

Study Datasets

Dataset	N	Female %	Role
PAIN (cross-domain)	129	86%	Data augmentation
CLALIT no-MD	67	62%	Pipeline optimization
CLALIT MD-agreement	59	69%	Evaluation

Conclusions

Voice screening should serve as a confirmatory support tool — not a standalone screener. When GP impression and patient self-report agree on severity, voice provides objective weight to the shared clinical judgement. In a 59-patient pilot, this selective approach achieved 95% accuracy at 35.6% coverage (p = 0.017), validating the agreement-based design.