

Tao PhD Thesis - Introduction Summary

"Speech-based Automatic Depression Detection"

University of Glasgow, 2023

The Problem

Depression is massive:

- 300+ million people affected globally
- Leading cause of disability worldwide
- Only 1 in 3 receive treatment
- Diagnosis is slow, subjective, and expensive

Current diagnosis problems:

- Relies on clinical interviews (Hamilton Scale, Beck Depression Inventory)
 - Requires trained professionals
 - Time-consuming and costly
 - Subjective — different clinicians may disagree
 - Patients may hide symptoms or lack insight
-

The Opportunity: Speech as a Biomarker

Why speech?

- Depression affects cognition → cognition affects speech
- Changes are often **involuntary** (hard to fake)
- Recording speech is **cheap and non-invasive**
- Can be done **remotely** (phone, app, telehealth)

What changes in depressed speech?

- **Prosody:** Flatter intonation, less pitch variation
 - **Timing:** Slower speech rate, longer pauses
 - **Voice quality:** Breathier, less energy
 - **Articulation:** Less precise consonants
-

The Research Gap

Previous work focused on:

- Achieving high accuracy (the "black box" approach)
- Using complex deep learning models
- Single datasets, single languages

What's missing:

1. **Interpretability** — WHY does the model predict depression?
 2. **Biomarker identification** — WHICH features matter clinically?
 3. **Cross-task comparison** — Does read vs spontaneous speech matter?
 4. **Generalisability** — Do findings transfer across datasets?
-

Tao's Thesis Contributions

Contribution 1: The ANDROIDS Corpus

- Created new publicly available dataset
- 118 participants (64 depressed, 54 controls)
- **Professional psychiatric diagnosis** (not questionnaires!)
- Both **read speech** AND **spontaneous speech** from same people
- Italian language, in-the-wild recording conditions

Contribution 2: Speech Duration & Silences

- Analysed timing patterns as depression markers
- Found significant differences in pause behaviour

Contribution 3: Feature Correlation Matrices

- Novel approach: look at relationships BETWEEN features
- Depressed speakers show different correlation patterns
- "Stability" of features differs between groups

Contribution 4: Multi-Local Attention (MLA)

- Novel deep learning architecture
- Focuses on most depression-relevant parts of speech
- Reduces recording time needed for detection

Contribution 5: Cross-Data Multilevel Attention (CDMA)

- Combines read AND spontaneous speech
 - Uses attention to weight task-specific vs shared features
 - Best results in the thesis
-

Key Results

Task	Method	Accuracy
Read Speech	LSTM Baseline	83.4%
Spontaneous	LSTM Baseline	81.6%
Combined	CDMA	~85%

Comparable to General Practitioners who achieve 57.9-73.1% accuracy!

Why This Matters For Your Dissertation

Tao's thesis focused on **achieving accuracy** with novel architectures.

Your dissertation can focus on **understanding WHY** — which features actually drive these predictions?

This is complementary, not competing:

- Tao built the tools and dataset
- You analyse what they reveal about depression

Your angle fills a gap Tao explicitly identifies:

"Future work should focus on interpretability and identifying which speech markers are most clinically relevant."

Supervisor Connection

Tao's supervisor: **Professor Alessandro Vinciarelli** (Glasgow)

Your supervisor may know him — worth asking! Could provide:

- Additional guidance
- Access to unpublished findings
- Potential examiner connection

Summary created for Nile's dissertation project, Feb 2026