REFERENCES

1. Systematic Reviews & Scoping Studies

1. **Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches** (MDPI)
   mdpi.com
   Systematic review of multimodal passive sensing for mental health.
2. **Passive Sensing for Mental Health Monitoring Using Machine Learning With Wearables and Smartphones: Scoping Review** (Journal of Medical Internet Research)
   jmir.org
   Broad review of machine learning + passive data for mental health.
3. **Beyond Detection: Towards Actionable Sensing Research in Clinical Mental Healthcare** (ACM UbiComp/IMWUT)
   acm.org
   Argues for actionable insights beyond raw detection.
4. **Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring** (Frontiers in Digital Health)
   frontiersin.org
   Overview of multi-source sensing (wearables, smartphones, environment).

---

2. Machine Learning Approaches

1. **Machine-learning detection of stress severity expressed on a continuous scale using acoustic, verbal, visual, and physiological data: lessons learned** (Frontiers in Psychiatry)
   frontiersin.org
2. **Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies** (PLOS One)
   plos.org
3. **Prediction of Mood Instability with Passive Sensing** (ResearchGate preprint)
   researchgate.net

---

3. EMA (Ecological Momentary Assessment) & Protocol Studies

1. **JMIR Research Protocols - Identifying Person-Specific Drivers of Depression in Adolescents: Smartphone-Based EMA and Passive Sensing Study**
   researchprotocols.org
2. **Possible Application of Ecological Momentary Assessment to Older Adults' Daily Depressive Mood: Integrative Literature Review** (JMIR Mental Health)
   jmir.org
3. **The utility of smartphone-based EMA for depressive symptoms** (ScienceDirect)
   sciencedirect.com

---

4. Clinical Validation & Key Case Studies

1. **StudentLife: Mobile Sensing of Depression & PHQ-9 Validation** (Dartmouth Thesis / Rui Wang 2018)
   PDF link
2. **Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students** (Frontiers in Psychiatry)
   frontiersin.org
3. **Mental Health and Behavior of College Students During Early COVID: Longitudinal Smartphone + EMA Study** (JMIR)
   jmir.org

---

5. CrossCheck & Digital Psychiatry Studies

1. **CrossCheck: Toward Passive Sensing and Detection of Mental Health Changes in People With Schizophrenia** (ACM / ResearchGate / PubMed)
   acm.org | researchgate.net | pubmed
2. **CrossCheck: Integrating Self-Report, Behavioral Sensing, and Smartphone Use to Identify Digital Indicators of Psychotic Relapse** (PubMed)
   pubmed

---

6. Supporting Tools & Measures

1. **PHQ-9: Validity of a Brief Depression Severity Measure** (NIH / PMC)
   □ PMC
2. **PHQ-9 Modified for Teens** (AACAP)
   □ aacap.org
3. **MDCalc PHQ-9 Clinical Tool**
   □ mdcalc.com

Part A — Literature Review (concise synthesis)

1. Field summary

- **Passive sensing** (smartphone + wearable sensors + environmental sensors) can yield predictive signals for mood, depression, stress, and relapse risk when combined with ML approaches (feature engineering + temporal models). Evidence across multiple longitudinal studies supports this potential. MDPI+1
- **Multimodal fusion** (accelerometer, GPS/location entropy, phone usage, device interaction, heart rate/HRV, sleep, conversation/activity proxies) outperforms single-modality approaches in most reviewed work. Systematic reviews show multimodal pipelines and careful preprocessing are common best practices. MDPI
- **Generalization is a core challenge**: models trained on one longitudinal dataset often fail to generalize to others without adaptation (domain shift, sensor heterogeneity, demographics). Cross-study generalization efforts show mixed success and call for personalization and transfer-learning strategies. PubMed Central

2. Key datasets / deployments & clinical validation

- **StudentLife (Dartmouth)** — large college-student longitudinal dataset linking smartphone sensing to GPA, stress, depression measures (PHQ-9), social and sleep behavior. Demonstrates feasibility of week-level depression prediction. studentlife.cs.dartmouth.edu+1
- **CrossCheck** — deployed clinical study with people with schizophrenia; integrates EMA, passive sensing, and clinician-verified outcomes; provides lessons about engagement, privacy, and clinical workflow integration. PubMed Central+1

- Multiple smaller/clinical trials (JMIR/Journals) and more recent scoping/systematic reviews (JMIR 2025, Sensors 2024) summarize evidence and note the need for stronger clinical validation pathways. PubMed Central+1

3. Typical ML architectures and performance measures

- **Feature extraction**: time-series summarization (hourly/daily buckets), activity counts, sleep duration, circadian regularity, location entropy, call/SMS/interaction counts, phone unlock patterns, HRV metrics when available.
- **Models**: Random Forests / Gradient-boosted trees for interpretability and baseline; LSTM / CNN-LSTM and transformer-style temporal models to capture dynamics; personalized models (fine-tuning) for better individual performance. MDPI+1
- **Evaluation**: AUC, F1, precision/recall, MAE for continuous severity. Reported accuracies vary widely depending on task/time-window/labels (week-level detection tends to be easier than day-level). Cross-study AUCs typically range from modest (~0.65) to strong (~0.9) depending on label quality and personalization. PubMed Central+1

4. Privacy, ethics, and deployment concerns

- **Privacy-first design** is essential: collect minimum necessary features, on-device preprocessing, differential privacy or federated learning where possible, robust consent and data governance. Clinical deployments (CrossCheck) require strict IRB and clinical workflows. PubMed Central
- **Sensor heterogeneity & missing data**: device differences (Android vs iOS, wearable vendors) and adherence dropouts require robust imputation and confidence/uncertainty modeling. MDPI

5. Gaps & research needs (from reviews)

- Larger multisite clinical trials with clinician-verified outcomes and prospective evaluation. PubMed Central
- Standardized benchmarks / open datasets to compare methods and assess generalization. PubMed Central
- Clear regulatory and validation pathways for clinical-grade use (medical device classification, CE/FDA when moving to clinical interventions). PubMed Central

---

Part B — Final Documentation: Software Requirements Specification (SRS)

(High-level actionable SRS for an MVP "Passive Mood Detection System" suitable for research-to-clinic pipeline)

1. Purpose & Scope

Deliver an application and backend that collects consenting users' passive sensor data (smartphone + wearables), processes it securely, and produces per-user mood risk estimates and clinician-facing summaries with alerts for high-risk events. The system supports research and clinical pilots.

2. Stakeholders

- End users (patients/participants)
- Clinicians / mental health teams
- Researchers / data scientists
- Product owners / site admins
- Security & compliance officers

3. Definitions

- **EMA**: Ecological Momentary Assessment (short self-report surveys).
- **Bucket**: aggregation window (e.g., daily).
- **Model Score**: continuous risk/probability of depression/stress on a 0–1 scale.

4. High-level Functional Requirements (FRs)

FR1 — User onboarding & consent

- FR1.1: Provide informed consent UI with plain-language description of data recorded, storage, sharing, and opt-out.
- FR1.2: Collect demographic metadata (age bracket, gender, diagnosis if clinical pilot, language) with optional fields.

FR2 — Data collection (mobile + wearables)

- FR2.1: Passive smartphone sensors: accelerometer, gyroscope, GPS coarse location (hashed), screen on/off, app usage, call/text metadata (counts, NOT content), battery level, Bluetooth device presence.

- FR2.2: Wearable integration: heart rate / HRV, step count, sleep stages when available via vendor APIs (Fitbit, Garmin, Oura, Apple Health). Use vendor OAuth flows.
- FR2.3: EMA administration: deliver short surveys (mood, PHQ-2/PHQ-9 short form) at configurable schedules.

FR3 — Edge preprocessing

- FR3.1: On-device preprocessing to compute derived features and reduce raw-data upload (feature extraction windows: 5m / 1h / 1d).
- FR3.2: Local encryption of cached data; upload over HTTPS when on Wi-Fi (policy configurable).

FR4 — Backend ingestion & storage

- FR4.1: Ingest batched feature uploads; store encrypted-at-rest.
- FR4.2: Support multi-tenant separation (research studies/clinical sites).
- FR4.3: Implement retention and deletion policies per consent (e.g., delete raw audio immediately; retain derived features for X months).

FR5 — ML pipeline

- FR5.1: Feature normalization and missing data handling modules.
- FR5.2: Baseline models: Random Forest + temporal LSTM ensemble. Support per-user personalization (fine-tune using 2-week calibration window).
- FR5.3: Provide uncertainty/confidence metric with each prediction; withhold alerts unless confidence > threshold.
- FR5.4: Versioned models + A/B testing support.

FR6 — Alerts & clinician dashboard

- FR6.1: Dashboard showing longitudinal trends, risk score, EMA responses, and explanationable features driving alerts (top-5 features).
- FR6.2: Alerting workflows: email/SMS/push notification to clinician or care team; include risk level and suggested next steps.
- FR6.3: Audit logs for all clinician actions and alerts.

FR7 — APIs & integrations

- FR7.1: REST API endpoints for retrieving aggregated data, scores, and reports.
- FR7.2: FHIR/HL7 integration for EHR interoperability (optional for clinical pilots).
- FR7.3: Vendor connectors: Fitbit, Oura, Garmin, Apple HealthKit, Google Fit / Health Connect.

5. Non-functional Requirements (NFRs)

- NFR1: **Security** — All data encrypted in transit (TLS 1.2+) and at rest (AES-256). Role-based access control (RBAC).
- NFR2: **Privacy & Compliance** — Support HIPAA-equivalent safeguards for US deployments; GDPR compliance for EU users (data subject access/deletion).
- NFR3: **Availability** — 99.5% system uptime for core ingestion/data access.
- NFR4: **Latency** — Feature batches processed within 5 minutes of upload; dashboard updated hourly.
- NFR5: **Scalability** — Support scaling to 10k active users in pilot, with horizontal scaling of ingestion and ML inference.
- NFR6: **Explainability** — For any clinical alert, include top contributing features and confidence level.

6. Data & Privacy Specifications

- Collect minimum necessary; never collect raw audio content; extract only voice activity counts and short hashed features if needed (and only after IRB approval). CrossCheck & StudentLife practices recommend **remove/avoid raw content**. PubMed Central+1
- Use pseudonymization: store PII separately from sensor streams with key management.
- Provide user controls for data export and deletion.

7. ML Evaluation & Clinical Validation Plan

- **Phase 0 — Retrospective validation**: use existing datasets (StudentLife, CrossCheck-like datasets, and in-house collected datasets) to benchmark (AUC, F1, calibration). studentlife.cs.dartmouth.edu+1
- **Phase 1 — Pilot (n≈100)**: 4–8 week passive + EMA data collection; compare model outputs to validated clinical scales (PHQ-9 weekly) and clinician assessment. Use pre-registered analysis plan. PubMed Central
- **Phase 2 — Clinical trial**: multisite prospective validation with clinician endpoints and safety protocols; plan for regulatory consultation if using for clinical decision support.

8. Implementation Notes & Best Practices (from reviews)

- **Personalization** is often required — plan for a short calibration period per user (commonly 1–2 weeks). PubMed Central

- **Cross-study generalization**: include domain-adaptation or meta-learning strategies for better transfer. PubMed Central
- **Clinician workflow**: present concise, actionable summaries — raw risk probability alone is not enough. CrossCheck experience shows integration with clinicians is crucial for real-world usefulness. PubMed Central

9. Risk Analysis & Mitigations

- **False positives**: produce unnecessary alarm — mitigate via thresholds, confidence gating, human-in-the-loop review.
- **Privacy breach**: implement strong encryption, key rotation, and breach response plan.
- **Bias & fairness**: test models across demographics; include fairness metrics in reporting.

10. Acceptance Criteria (MVP)

- Successful enrollment and consent flow.
- Continuous passive data ingestion from smartphone + one wearable vendor.
- Baseline ML model produces daily mood score, with documented AUC ≥ 0.70 on held-out internal pilot dataset.
- Clinician dashboard with alerts, explanations, and audit logs.
- IRB approval and privacy review completed for pilot.

Appendices

Appendix A — Short summarized findings (by paper category)

- **Systematic reviews (Sensors 2024, JMIR 2025)**: summarize trends, sensors used, typical pipelines, and highlight need for better clinical trials. MDPI+1
- **Cross-study generalization (PLOS ONE 2022)**: demonstrates domain shift issues and the value of personalization/transfer learning. PubMed Central
- **StudentLife and deployment studies**: successful college-student longitudinal predictions linking behavior to PHQ-9 / stress; practical lessons on adherence and label collection. studentlife.cs.dartmouth.edu+1
- **CrossCheck and schizophrenia deployments**: clinical integration lessons, EMA + passive sensing synergy, and ethical/operational considerations for severe mental illness contexts. PubMed Central+1

Appendix B — Recommended immediate next steps (practical)

1. **Select pilot population & IRB**: define inclusion/exclusion and get approvals (clinical vs. general research).
2. **Instrument apps**: build mobile SDK with on-device feature extraction + secure upload.
3. **Select wearable vendor** for pilot (Fitbit or Oura) and implement OAuth connector.
4. **Assemble baseline datasets**: obtain StudentLife-style datasets (public or collaboration) and re-run baseline models to reproduce performance. studentlife.cs.dartmouth.edu
5. **Run 8-week internal pilot (n=50–100)** and measure model metrics & clinician feedback.

Sources / Suggested reading (clickable)

- Khoo LS, Lim MK, Chong CY, McNaney R. *Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches.* Sensors. 2024. MDPI
- ShiYing Shen et al. *Passive Sensing for Mental Health Monitoring Using Machine Learning With Wearables and Smartphones: Scoping Review.* JMIR. 2025. PubMed Central
- Adler DA, Wang F, Mohr DC, Choudhury T. *Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies.* PLOS ONE. 2022. PubMed Central
- Rui Wang et al. *StudentLife / Tracking Depression Dynamics in College Students.* Dartmouth (thesis & papers). 2014–2018. studentlife.cs.dartmouth.edu+1
- Ben-Zeev D. *CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse.* JMIR / ACM (CrossCheck publications). PubMed Central+1

Notes about coverage & limitations of this document

- I synthesized the document from the primary reviews and representative study papers you provided. For an **exhaustive** academic systematic review or a full SRS with contractual/regulatory language, we should iterate with direct access to full-text PDFs of each paper and any internal policy/regulatory constraints (HIPAA, GDPR, local regs).
- The SRS is intentionally scoped for an *MVP/pilot* to get to validated clinical utility; elements for production medical-device certification (e.g., IEC 62304, FDA SaMD guidance) are mentioned at a high level but not expanded into certification checklists.