**Phase 0 — Problem framing & guardrails**

* **Goal (exploratory):** Determine whether skin color/appearance in images contains *usable signal* correlated with diabetes status (e.g., HbA1c-defined) for **risk screening**, not diagnosis.
* **Primary endpoints:**
  1. AUC/PR-AUC for classifying diabetes (HbA1c ≥ 6.5%) vs. non-diabetes;
  2. Calibration (Brier score);
  3. Subgroup fairness gaps (age, sex, Fitzpatrick type, ethnicity).
* **Key risks:** Confounding (skin tone, lighting, camera model, jaundice/anemia, vascular disease, smoking, sun exposure, cosmetics). Mitigate up front (below).
* **Ethics/IRB:** Obtain IRB approval, HIPAA-compliant data handling, explicit consent for imaging + AI use, clear non-diagnostic disclosures.

**Phase 1 — Study design**

* **Population:** Adults ≥18 across diverse Fitzpatrick types I–VI. Aim for balanced representation across sex, age bands, BMI categories, and ethnic groups. Include both diabetic and non-diabetic participants; optionally prediabetes (HbA1c 5.7–6.4%).
* **Ground truth:** Same-day **HbA1c** (primary). Optionally fasting plasma glucose and OGTT where feasible.
* **Sample size (order-of-magnitude):** Start with N≈1,500–3,000 participants across ≥3 clinical sites to enable robust held-out site validation and subgroup analyses. (Power analysis finalized with your biostatistician.)
* **Inclusion/exclusion examples:**
  + Include: Stable chronic conditions.
  + Exclude/flag: Acute dermatologic disease at imaging sites, active jaundice, recent self-tanner, heavy makeup, strong recent sunburn; capture as metadata if not excluded.

**Phase 2 — Imaging protocol (standardization is everything)**

* **Anatomical sites (practical + informative):** Dorsal hand, volar forearm, lower leg/shin, face (cheek). Record exact site.
* **Lighting:** Controlled booth or portable light box at ~D65 color temperature; constant lux; no mixed lighting.
* **Camera:** Fixed smartphone model *and* a secondary model to test generalization. Capture **RAW/DNG** when possible; otherwise highest-quality JPEG with all “beauty” filters off.
* **Color calibration:** Include a **24-patch color checker** in the first frame of each session; use it to build a per-session color transform (white balance + color constancy).
* **Geometry:** Fixed distance (ruler/stand), fixed focal length, polarizing filter pair (cross-polarization) to reduce specular glare.
* **Hygiene:** Remove lotions/makeup on imaged area; 5-minute acclimatization to room temperature to normalize perfusion.
* **Replicates:** 2–3 shots per site to estimate within-session variance.

**Phase 3 — Data & metadata capture**

* **Clinical metadata (min set):** Age, sex, ethnicity, Fitzpatrick type (self-report + assessor), BMI, smoking status, blood pressure, known anemia/liver disease/thyroid disease, medications (e.g., steroids), time since last sun exposure, recent cosmetics.
* **Device metadata:** Phone model, lens, ISO, exposure, white balance, RAW vs JPEG, lighting ID, site ID, operator ID, timestamp.
* **De-identification:** Crop to skin ROIs; remove facial identifiers unless face is a planned site (then store separately with restricted access). Hash participant IDs; store key in a separate enclave.

**Phase 4 — Preprocessing pipeline (analysis-ready images)**

* **Quality checks:** Sharpness/blur detection, exposure clipping checks; auto-flag and re-capture if needed.
* **Color correction:** Per-session white balance + color transform using the calibration card; convert to device-independent space (CIELAB).
* **Skin segmentation:** Classical (thresholding in normalized color space) + sanity-checked by a lightweight model; exclude nails, hair, veins if possible.
* **Artifact handling:** Mask specular highlights, ink/tattoos, scars.
* **Normalization:** Standardize crop size, spatial resolution, and dynamic range; store masks and transforms for auditability.

**Phase 5 — Feature strategy (progressive, interpretable → complex)**

* **Tier 1 (interpretable baselines):**
  + Color statistics in RGB/HSV/CIELAB over skin mask (means, percentiles).
  + **Melanin index** and **erythema index** proxies from visible channels.
  + Texture descriptors (local binary patterns), simple vascularity proxies.
  + Site-specific models (hand/forearm/leg/face) and a pooled model with site as a feature.
* **Tier 2 (learning-based):**
  + Compact CNN/ViT features with heavy color-constancy augmentation and **domain adversarial** training to reduce device/lighting leakage.
  + Multi-instance learning to combine multiple sites per person.
* **Tier 3 (robustness):**
  + Domain generalization across camera models and sites; test-time adaptation limited to color calibration only (no label leakage).

**Phase 6 — Modeling & validation plan**

* **Splits:** Person-level splits.
  + Dev: 60% (stratified) for model selection.
  + Internal test: 20% held-out.
  + **External test:** 20% from a never-seen clinical site (leave-one-site-out) to assess generalization.
* **Metrics:** AUC, PR-AUC, sensitivity at fixed specificity (e.g., 90%), calibration (reliability plots, ECE, Brier).
* **Fairness:** Report metrics by Fitzpatrick type, sex, age, BMI, ethnicity, camera model, and site; **equalized odds gap** and **TPR disparity**.
* **Confounding audits:**
  + Predict camera model/site from raw images; if high, incorporate adversarial debiasing or stricter preprocessing.
  + Negative control tasks (e.g., predict handedness) to estimate spurious learning.
  + Sensitivity analyses with/without anemics/jaundiced participants.

**Phase 7 — Statistical analysis & clinical utility**

* **Primary analysis:** Association between image-derived risk score and HbA1c (logistic for diabetes status; linear for HbA1c as continuous).
* **Secondary:** Net reclassification improvement (NRI) over simple baselines (age+BMI) and over non-image vitals if available.
* **Decision analysis:** Threshold selection for screening; PPV/NPV at realistic prevalences in primary care vs community settings.

**Phase 8 — Bias, safety, and governance**

* **Bias monitoring:** Pre-register subgroup analyses; publish subgroup performance table with CIs.
* **Human factors:** Clear UI labels (“experimental risk estimate, not diagnostic”), no individual result returned to participants unless IRB-approved process exists.
* **Data governance:** Secure storage, audit trails, model cards and data sheets, reproducible pipelines with versioned configs.

**Phase 9 — Reproducibility & documentation**

* **Artifacts to maintain:**
  + Imaging SOP (one-pager with photos of setup), calibration SOP, operator checklist.
  + Data dictionary & codebook for metadata.
  + Model card (intended use, limitations, populations, metrics).
  + Full experiment logs: seeds, split manifests, preprocessing parameters.
* **Blinded re-read:** Random subset independently re-processed to quantify pipeline repeatability (ICC).

**Phase 10 — Extensions (optional after baseline)**

* **Longitudinal sensitivity:** Same participants at 3–6-month intervals to see if risk score tracks HbA1c change.
* **Multimodal add-ons:** Combine image features with easily captured signals (e.g., PPG from phone camera, mmWave vitals you already collect) to test incremental value.
* **Open science:** Release a de-identified, bias-balanced benchmark (with permissions) plus SOPs to catalyze external validation.

**Concrete deliverables & milestones**

* **M0–M1:** IRB approval, imaging booth setup, SOPs finalized, operator training.
* **M2–M5:** Data collection across 3 sites (target ≥1,500 participants), weekly QA dashboards.
* **M6:** Locked preprocessing pipeline; Tier-1 baseline results.
* **M7–M8:** Tier-2 modeling, external-site test, fairness audit.
* **M9:** Clinical utility analysis, write-up (methods + model card + data sheet).
* **M10:** External advisory review (dermatology/endocrinology/biostatistics), preprint.

**Minimal kit list**

* Portable light booth or standardized LED panel with diffuser.
* Smartphone(s) with RAW capture, fixed mounts, cross-polarizers.
* Color checker card + ruler/stand.
* Consent forms, REDCap/ODK or equivalent for metadata capture.
* Secure storage with access controls; pipeline for RAW → calibrated CIELAB tiles.

**Go/no-go criteria (early)**

* If after Tier-1/2 the *externally tested* AUC < 0.70 or subgroup gaps >10–15 percentage points persist despite mitigation, **stop or pivot** (e.g., to multimodal).