Phase 2 Feasibility Report: AI-Powered Chronic Disease Risk Screener Using Non-Invasive Signals

Team Members: Niyati Nikunj Kapadia, Teli Sumit Laxmanbhai, Sarah Pradhan, Patel Jugal Kunteshkumar
Course: CECS 551 – Term Project
Date: October 17, 2025

---

1. Market Research & Existing Solutions

Chronic diseases such as diabetes, hypertension, and kidney disease account for approximately 75% of global deaths. Yet, early detection still depends on invasive tests, costly imaging, or self-reported symptoms, methods that remain inaccessible to 4.5 billion people worldwide who lack full access to essential health services.

Our goal is to enable **early-stage screening** using **non-invasive inputs** such as smartphone images and minimal patient data.

Current ecosystem:

- **Google DeepMind**'s diabetic retinopathy AI requires fundus cameras costing thousands of dollars.
- **IDx-DR**, the first FDA-approved tool, focuses only on retinopathy and requires clinical hardware.
- **Cardiogram** and **SkinAI** leverage wearables or image data but limit predictions to single parameters such as heart rate or skin lesions.

There is currently **no low-cost, multi-disease screener** capable of running on consumer devices. This is the gap our project fills, particularly valuable for community clinics and under-resourced regions.

---

2. Deep Learning Approach & Feasibility

We selected lightweight yet effective architectures that balance **accuracy, interpretability, and deployability**:

- **MobileNet V3** and **EfficientNet-B0** for retinal and nailbed image analysis, optimized for limited compute.
- **Tabular Transformer** for demographic and lifestyle data such as age, BMI, smoking, and activity.
- **Fusion Layer** merging visual and tabular embeddings into interpretable probability scores for multiple diseases.
- **Explainability:** SHAP for tabular data and Grad-CAM heatmaps for images to clarify prediction rationale.

Preliminary trials on sample subsets already achieved **>80% accuracy** in identifying early disease markers, confirming both feasibility and promise for further scaling in Phase 3.

---

3. Data Sources & Preprocessing Workflow

Ensure both visual and demographic coverage, we integrated six open-access datasets:

1. **Mendeley Fundus Images** – retinal photos labelled for diabetic and hypertensive indicators.
2. **Kaggle Nail Disease Dataset** – nailbed images showing pigmentation and vascular variations.
3. **Zenodo Nailbed Health Dataset** – additional color and texture diversity.
4. **Kaggle Health & Lifestyle Dataset** – self-reported habits and physical activity data.
5. **Synthetic Population Demographics Dataset** – demographic balancing and feature diversity.
6. **CDC NHANES Dataset** – real U.S. health survey linking lifestyle and clinical outcomes.

Preprocessing steps:

- Merged datasets using *age* as a key and excluded entries under 18 years.

- Dropped columns with >60% missing data and removed redundant variables (e.g., waist highly correlated with weight).
- Filled missing numerical values with median and categorical with mode to preserve distributions.
- Filtered implausible records (e.g., height >250 cm or weight <30 kg).
- Normalized all numerical fields to [0, 1] and removed personal identifiers for privacy compliance.

After cleaning, the **final dataset includes 15,609 valid tabular records and ~60,000 filtered images**. Exploratory analysis confirmed expected trends, e.g., higher BMI and poor sleep correlated with elevated diabetes risk.

---

4. Business Applicability & Impact

The proposed tool serves **three main user groups:**

- **Community Clinics:** Identify high-risk patients within two minutes using only smartphone images.
- **Telehealth Platforms:** Integrate as an API to enable automatic screening during virtual consultations.
- **Insurance Providers:** Analyse anonymized risk trends to inform preventive-care incentives.

A pilot in a local clinic could **reduce screening costs by up to 90%**, providing clinicians a ranked list of patients needing lab confirmation. With minimal hardware requirements, the tool extends preventive healthcare to underinsured and remote populations.

---

5. Feasibility Assessment

Dimension	Status	Supporting Evidence
Technical	Feasible	Clean multimodal dataset: CNNs and Transformer evaluated successfully
Operational	Practical	Runs on mid-tier devices; integrable with web/mobile interfaces
Market Need	Strong	Targets major affordability and accessibility gaps
Ethical Compliance	Ensured	All datasets open-access and anonymized

---

Conclusion

The feasibility phase confirms that the **AI-powered chronic disease screener** is technically achievable, operationally practical, and socially impactful. By merging multimodal datasets, leveraging interpretable deep-learning architectures, and prioritizing affordability, this project establishes a durable foundation for scalable early detection of chronic conditions across underserved communities.

References

- Grand View Research, *AI in Healthcare Market Report* (2024) https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market
- Mendeley Retinal Fundus Images (2023) https://data.mendeley.com/datasets/s9bfhswzjb/1 https://data.mendeley.com/datasets/trghs22fpg/4
- Kaggle Nail Disease Image Classification Dataset (2024) https://www.kaggle.com/datasets/josephrasanjana/nail-disease-image-classification-dataset
- Zenodo Nailbed Health Dataset (2023) https://www.mdpi.com/2306-5729/8/2/29
- Kaggle Health & Lifestyle, Synthetic Population, and NHANES Datasets (2023–24) https://www.kaggle.com/datasets/mahdimashayekhi/health-and-lifestyle-dataset