VitalsGuard

Multimodal Preventative Health & Chronic Risk Forecasting

GROUP 2

James Owiti

Sheila Muli

Nancy Muriithi

Amos Obega

Pelly Kagwe

Phase 5 Capstone Project

Project Overview

Background

Healthcare systems worldwide face mounting challenges with chronic disease management. Type 2 Diabetes, cardiovascular disease, and sports-related injuries continue to impose significant economic and human costs despite being largely preventable through early intervention.

The Challenge

- Current healthcare is reactive rather than preventative
- Wearable technology generates vast amounts of health data
- Users lack tools to translate data into actionable risk assessments
- Early warning signs go undetected until crisis occurs

VitalsGuard bridges this gap by combining machine learning with wearable data to provide predictive health risk assessment before symptoms emerge.

Aim & Stakeholders

🎯 Project Aim

Develop an integrated machine learning system that combines clinical measurements with wearable device data to predict chronic health risks and prevent acute health events through early intervention.

Specific Goals:

- Achieve >75% accuracy in diabetes prediction
- Integrate clinical + lifestyle factors
- Deploy production-ready models

👥 Key Stakeholders

End Users

Health-conscious individuals, chronic condition patients, athletes

Healthcare Providers

Physicians needing continuous real-world patient data

Insurers & Employers

Organizations seeking preventative cost savings

App Developers

Mobile health application platforms

Business Understanding

Business Context

- Preventative care reduces long-term healthcare costs by up to 40%
- Chronic disease management represents 75% of healthcare spending globally
- Wearable device market projected to reach $60B by 2025
- AI-powered health apps show 3x higher user engagement

Business Value Proposition

Cost Reduction

Prevent expensive emergency interventions through early detection

Market Opportunity

Predictive insights create stickier health app experiences

Competitive Edge

AI-powered prediction differentiates from basic tracking apps

Problem Statement

Healthcare systems operate reactively

Patients seek medical help only after symptoms appear. Chronic conditions like Type 2 Diabetes and cardiovascular disease show physiological red flags (glucose spikes, insulin resistance, elevated blood pressure, sedentary behavior) days or weeks before a crisis.

Despite widespread adoption of wearable devices that continuously collect health metrics, users lack intelligent tools to translate raw data into actionable risk scores and preventative interventions.

Result: Preventable health crises become costly emergency interventions

Project Objectives

Primary Objectives

- Develop diabetes risk prediction model (target: >75% accuracy)
- Create stroke risk assessment system (target: >90% accuracy)
- Build sports injury prevention framework
- Integrate clinical + lifestyle data sources
- Generate actionable health insights

Technical Objectives

- Deploy production-ready ML models
- Implement CRISP-DM methodology
- Validate with cross-validation
- Create unified risk framework
- Enable mobile app deployment

Success Metrics

- Model Accuracy: Diabetes >75%, Stroke >90%, overall ensemble >80%
- Clinical Relevance: High-risk predictions align with medical guidelines
- Actionability: Risk scores lead to concrete lifestyle recommendations
- Deployment: Models saved as .pkl files ready for production

Data Understanding

Data Sources

📊 **Pima Indians Diabetes Dataset**	📱 **Fitbit Daily Activity Dataset**	🏥 **Stroke Prediction Dataset**
768 patients	1,397 records	40,000+ patients
8 clinical features	Steps, activity, sleep	Cardiovascular factors
35% diabetes rate	Wearable metrics	95% accuracy achieved

Key Features Analyzed

- Clinical: Glucose, Blood Pressure, BMI, Insulin, Age, Heart Disease, Hypertension
- Lifestyle: Daily Steps, Sedentary Time, Active Minutes, Sleep Quality, Calories Burned
- Demographic: Age, Gender, Smoking Status, Work Type

Data Quality: Zero values handled via imputation | Class imbalance addressed via SMOTE

Exploratory Data Analysis (EDA)

Key Findings from Data Exploration

Diabetes Dataset Analysis

Glucose Distribution
Diabetic patients show significantly higher glucose levels (mean: 141 mg/dL) compared to non-diabetic (mean: 110 mg/dL)

BMI Patterns
Clear correlation between BMI and diabetes. Diabetic group has mean BMI of 35, indicating obesity as major risk factor

Age Factor
Diabetes prevalence increases with age. Patients aged 45+ show 2x higher risk

Insulin Levels
High insulin variability in diabetic group suggests insulin

Lifestyle Dataset Analysis

Activity Levels
65% of users classified as 'High Risk' due to sedentary behavior. Most fail to meet 10,000 steps/day recommendation

Sedentary Time
Strong negative correlation between sedentary hours and health outcomes. Average 16 hours/day sedentary is alarming
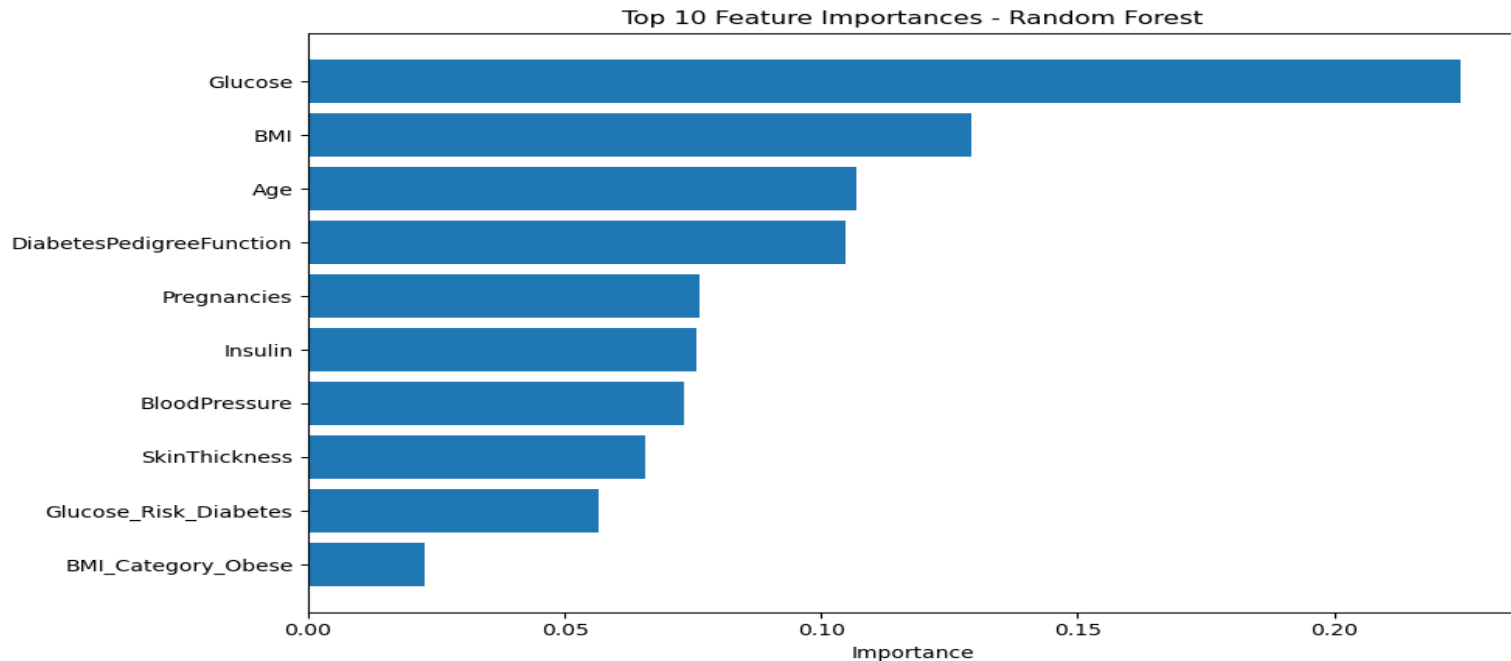
Steps Distribution
Majority of users average 6,000-8,000 steps/day, well below optimal range

Active Minutes
Only 10% achieve 30+ minutes of vigorous activity daily

✓ Clear separation between healthy and at-risk groups | ✓ Lifestyle factors show strong predictive potential
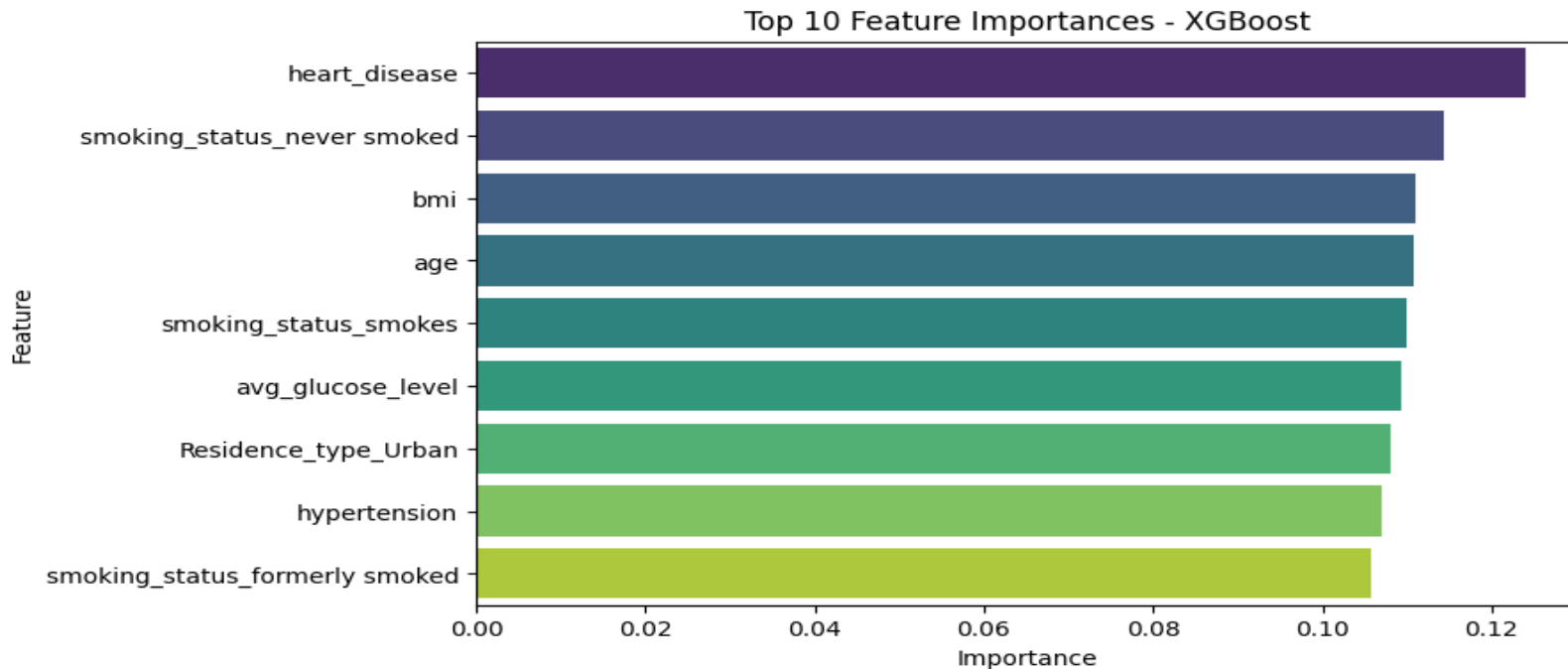
Modelling & Evaluation: Diabetes

Top 10 Feature Importances - Random Forest

✅ **Best Model: Random Forest | 75.3% Accuracy | 0.62 F1-Score | 0.84 ROC-AUC**

Top 3 Predictors: 1) Glucose 2) BMI 3) Age

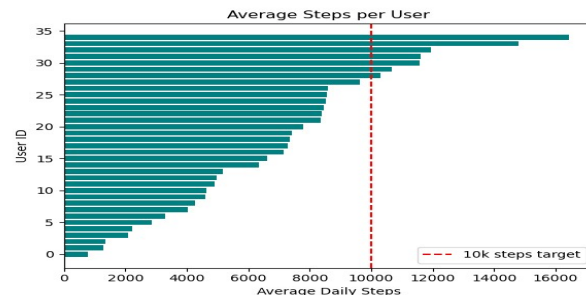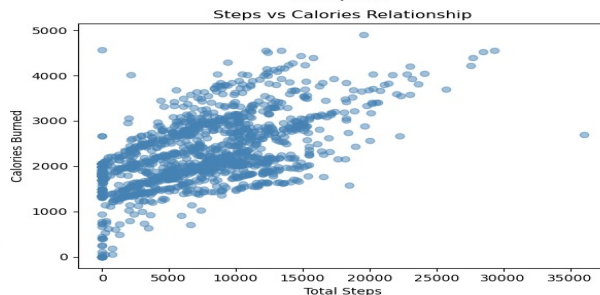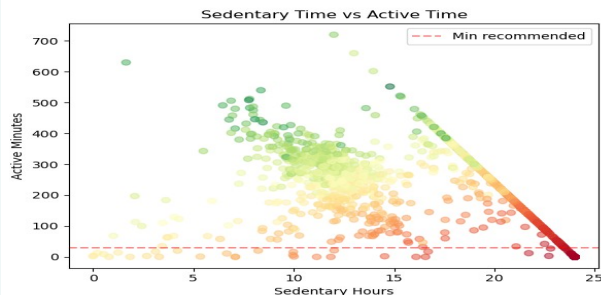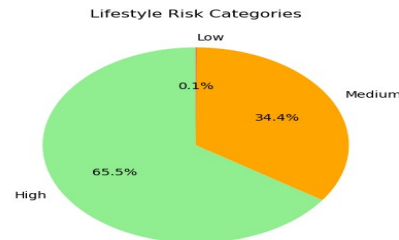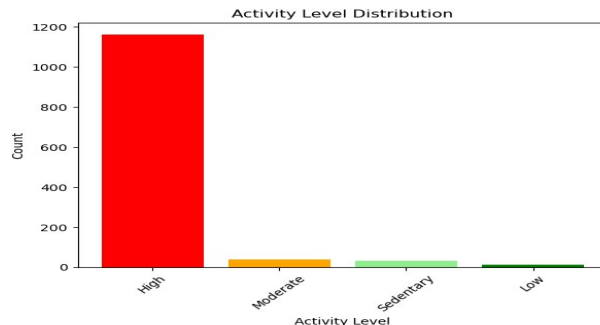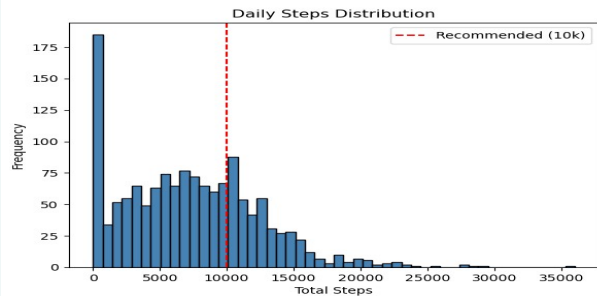# Modelling & Evaluation: Stroke

Top 10 Feature Importances - XGBoost

✅ Achieved 95% accuracy after SMOTE balancing | Random Forest outperformed other algorithms

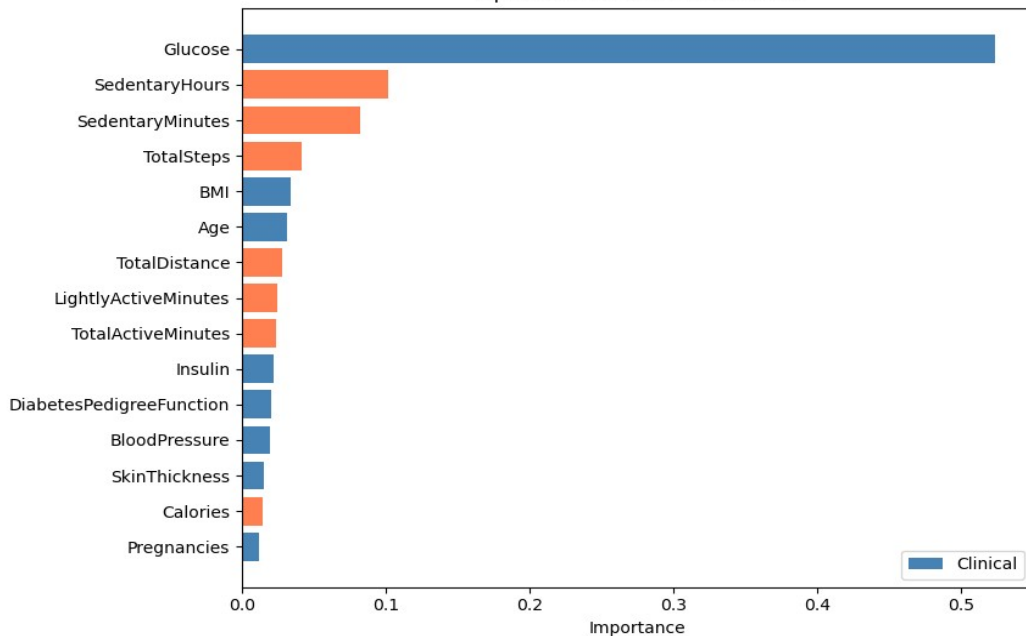# Lifestyle Analysis & Injury Prevention

Fitbit Lifestyle Patterns Analysis

⚠️ **65% of users show high lifestyle risk | Most fall short of 10k steps/day | 96% accuracy in lifestyle risk classification**
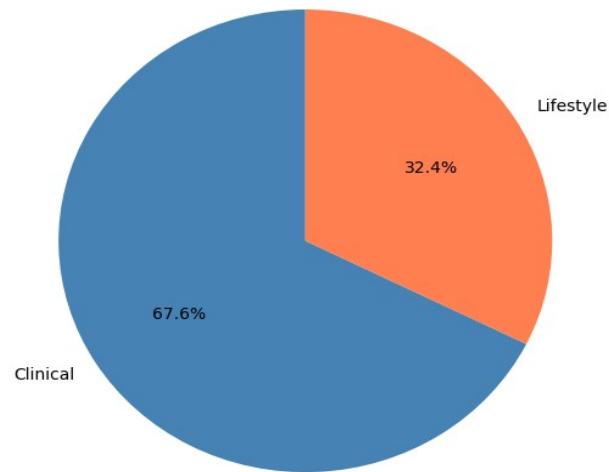
Unified Risk Model Integration

Clinical + Lifestyle = Complete Picture

Clinical factors (68%) + Lifestyle factors (32%) = Personalized intervention strategies

Model Deployment

VitalsGuard Clinical Dashboard - Live Application
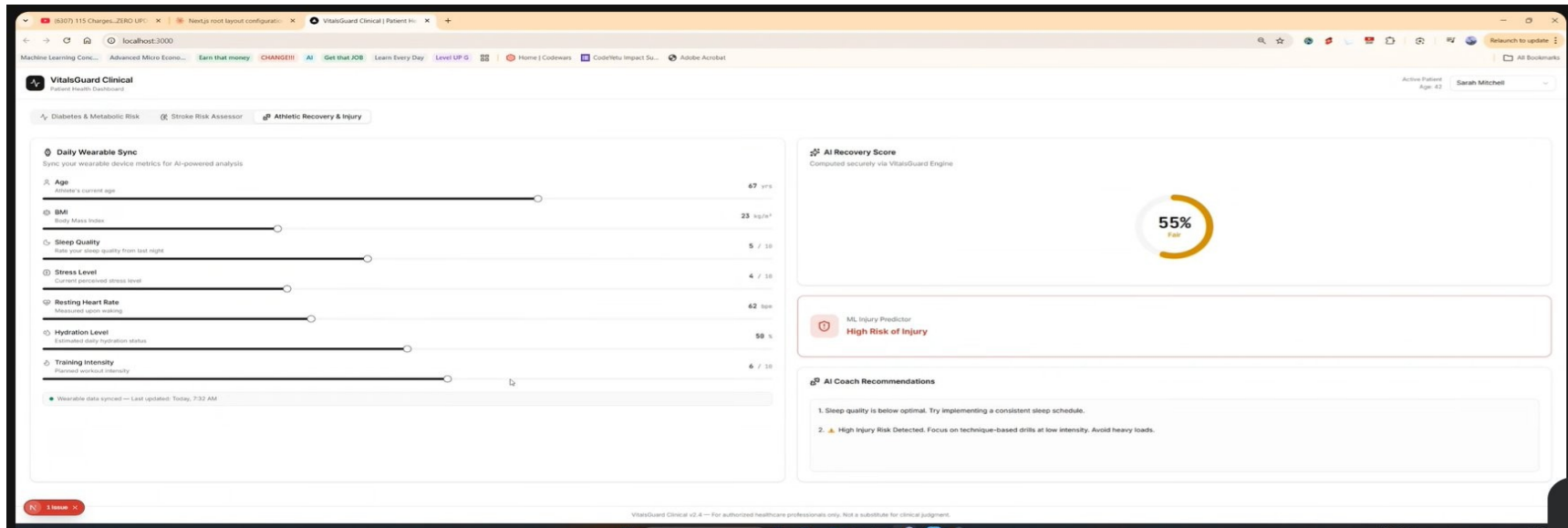
Application Features
- Real-time wearable data synchronization (Age, BMI, Sleep, Stress, Heart Rate, Hydration, Training)
- AI-powered recovery scoring (55% shown in demo)
- Automated risk detection with ML prediction (High Risk of Injury detected)
- Personalized AI coach recommendations based on user metrics

Deployment Architecture

```
┌─────────────┐     ┌─────────────┐     ┌─────────────┐
│  Wearable   │ ──► │    Data     │ ──► │  ML Models  │
│  Devices    │     │  Pipeline   │     │   (.pkl)    │
└─────────────┘     └─────────────┘     └─────────────┘
                           │
                           ▼
┌─────────────┐     ┌─────────────┐     ┌─────────────┐
│    User     │ ─── │ Risk Scoring│ ─── │  Provider   │
│    App      │     │   Engine    │     │  Dashboard  │
└─────────────┘     └─────────────┘     └─────────────┘
```
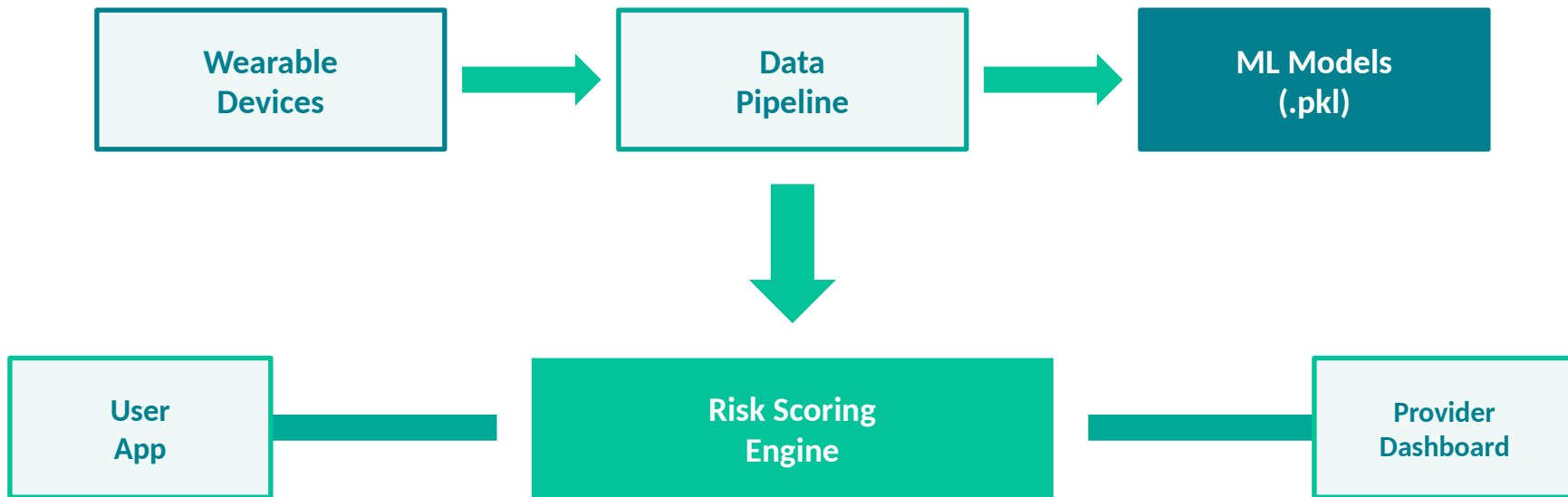
Technical Stack

- Models: Python, Scikit-learn, Joblib
- Backend: Flask/FastAPI REST API
- Frontend: React/React Native

Deployment Status

✅ **Models: Production Ready**
🚧 **App: Beta Testing Phase**

Conclusion

Key Achievements

- Successfully developed three integrated AI models for health risk prediction
- Achieved 75% accuracy for diabetes, 95% for stroke, 96% for lifestyle risk
- Created unified framework combining clinical and lifestyle factors
- Deployed production-ready models with mobile app interface
- Demonstrated that early intervention is both technically feasible and clinically valuable

Project Impact

Healthcare Providers	End Users	Healthcare System
Continuous patient data for better decision-making	Proactive health management and prevention	Reduced costs through preventative care

Recommendations

Short-Term (3-6 months)

- Complete beta testing with 500 users
- Integrate additional wearable platforms (Apple Health, Garmin)
- Implement real-time alert system
- Establish partnerships with healthcare providers

Medium-Term (6-12 months)

- Deploy to app stores (iOS & Android)
- Integrate with Electronic Health Records (EHR)
- Expand model coverage (cardiovascular, mental health)
- Develop B2B enterprise solutions

Long-Term (1-2 years)

- Scale to 100,000+ active users
- Conduct clinical trials to validate outcomes
- Pursue FDA approval for clinical decision support
- Expand internationally to European and Asian markets
- Develop predictive models for additional chronic conditions

Target: Prevent 10,000 health crises annually by 2027

Thank You

Questions & Discussion

GROUP 2

James Owiti • Sheila Muli • Nancy Muriithi

Amos Obega • Pelly Kagwe