# Virtual Care Assistant from home

Presentation Part Two

Presented By :

Shay Sason
Gabrielle Maor

# Project Description

**Project**

Virtual Care Assistant from home : predict patient status from vitals & text

**Task**

- Multi-modal classifier for Normal / Improvement / Deterioration
- Fuse structured vitals + free-text complaints

**Data**

- Synthetic data
- 700 patient-episodes (HR, $SpO_2$, Temp, RR + free text)
- Preprocessing: impute & normalize vitals, TF-IDF & RandomForest or BERT embeddings

**Evaluation**

- Stratified 5-fold CV
- Metrics: Accuracy, Macro-F1
- Baseline vs. Deep-Learning comparison

# Prior Art

| Source / Title (Year) | Approach / Model | Data (size) | Metrics | Results |
|---|---|---|---|---|
| “Deep learning for early warning of inpatient deterioration” (Nature Digit. Med., 2021) https://did.li/1otx5 | LSTM with attention on vital-sign time-series | 2,100 patient episodes (minute-by-minute vitals) | AUC, Macro-F1 | AUC 0.87, Macro-F1 0.71; attention maps showed $SpO_2$ drops and HR spikes as strongest early-warning signals; model predicts deterioration up to 12 h in advance. |
| “Fusion of clinical notes and vitals using BERT + MLP” (J. Gen. Intern. Med., 2023) https://did.li/fAVaa | BERT encoder on free-text notes + MLP head for vitals | 800 home-hospitalization episodes | Accuracy, Precision, Recall | Accuracy 0.76; Precision 0.74; Recall 0.72; text+vitals fusion improved Macro-F1 by 8 pp over vitals-only baseline; ablation confirmed synergy of both modalities. |
| “Combining vital signs and free-text for deterioration detection” (PMC, 2019) https://did.li/dR7CN | TF-IDF vectorization on clinical notes + LightGBM | 1,200 inpatient stays (notes + vitals) | ROC-AUC, Sensitivity, Specificity | ROC-AUC 0.79; Sensitivity 0.81; Specificity 0.75; top predictive text tokens included “pain,” “breathlessness,” “fatigue”; combining features outperformed single-modality. |

# Market Solutions

**Examples of Existing Market Solutions:**

**Current Health (USA)**

- AI-powered remote monitoring platform that combines wearable sensors and patient-reported data to detect early clinical deterioration. www.currenthealth.com

**Datos Health (Israel)**

- Automated remote-care system integrating physiological signals and patient self-reports to personalize treatment plans using AI. www.datos-health.com

**Our Approach:**

**Active, Proactive Classification**

- Real-time prediction of patient status (Stable / Improving / Deteriorating), not just passive monitoring.

**Fusion of Free-Text Reports & Vitals**

- Applying NLP to unstructured patient complaints and descriptions alongside physiological measurements.

**Clinical Decision Support Focus**

- Prioritizing patients who need immediate intervention and aiding medical teams on the ground with rapid, data-driven decisions.

# Data Construction Process:

**During the project, a Synthetic data was developed to handle free-text inputs and patient-reported clinical measures. The goal was to extract meaningful clinical, emotional, and functional insights from the data. The process involved building, processing, and organizing the data in a way that enables effective and insightful analysis.**

**1. Data Ingestion and Sources:**

The data consisted of free-text entries written by patients as part of a home hospitalization service. A dedicated ingestion interface was developed, integrating with the OpenAI API to enable automated submission of texts for processing.

**2. Technical Setup and Work Environment**

Essential helper libraries were imported, including: os, json, time, random, and openai.

The API key was configured securely using environment variables (os.getenv) to ensure data protection and privacy.

**3. Natural Language Processing (NLP)**

Each text was analyzed using OpenAI language models.

The processing aimed to:

- Summarize the texts to identify key points
- Extract clinical indicators, emotions, and patient-reported experiences
- Detect recurring patterns such as functional decline, worsening conditions, or signs of improvement

**4. Feature Construction for Insight Extraction**

Following the processing phase, the output was structured into a standard format (JSON/CSV) for further analysis.

Each record included the following features:

- **day1_note and day2_note:** First-person, simple, everyday-style text
- **Daily vitals:** Heart Rate (HR), Blood Pressure (BP), Body Temperature (Temp), Respiratory Rate (RR)
- **Status change direction:** Aligned with both the free-text entries and vital signs
- **Reasoning:** Objective explanation of the change observed between Day 1 and Day 2
- **Randomized parameters**: Age (25–110), gender, medical diagnosis
- **Labeling & Classification:** Improvement/ Deterioration/ No Change

# Steps

**1.Preprocessing:**

**Dataset:**

home-hospitalization episodes

**Vitals**

- Impute missing values
- Smooth sensor spikes & cap extreme outliers (e.g. HR <40 or >180, $SpO_2$ <80)
- Normalize with StandardScaler

**Text**

- Lowercase, strip punctuation & stopwords
- Option A: TF-IDF vectorization
- Option B: extract contextual embeddings (e.g. BERT)

**2.Labeling**

- **Classes:**
  - no change : 37.2%
  - Improvement: 31.5%
  - Deterioration: 31.3%
- **Preparation:**
  - Ensure correct mapping of status codes
  - Stratify splits to preserve class ratios

**3.Model Comparison**

- **Vitals-only (baseline)**
  - Logistic Regression
  - Random Forest
- **Text-only**
  - TF-IDF → Logistic Regression
  - Frozen BERT embeddings → Random Forest
- **Classical Fusion**
  - [TF-IDF + Vitals] → Random Forest
- **Deep-Learning Fusion**
  - Fine-tune BERT on text + MLP head on vitals features

**4.Evaluation**

- Stratified 5-fold CV
- **Metrics:** Accuracy | Macro-F1 |
- **Ablation studies:** compare
  a. Vitals-only
  b. Text-only
  c. Combined modalities

# Exploration & Baseline

**Exploration**

**Label distribution:**

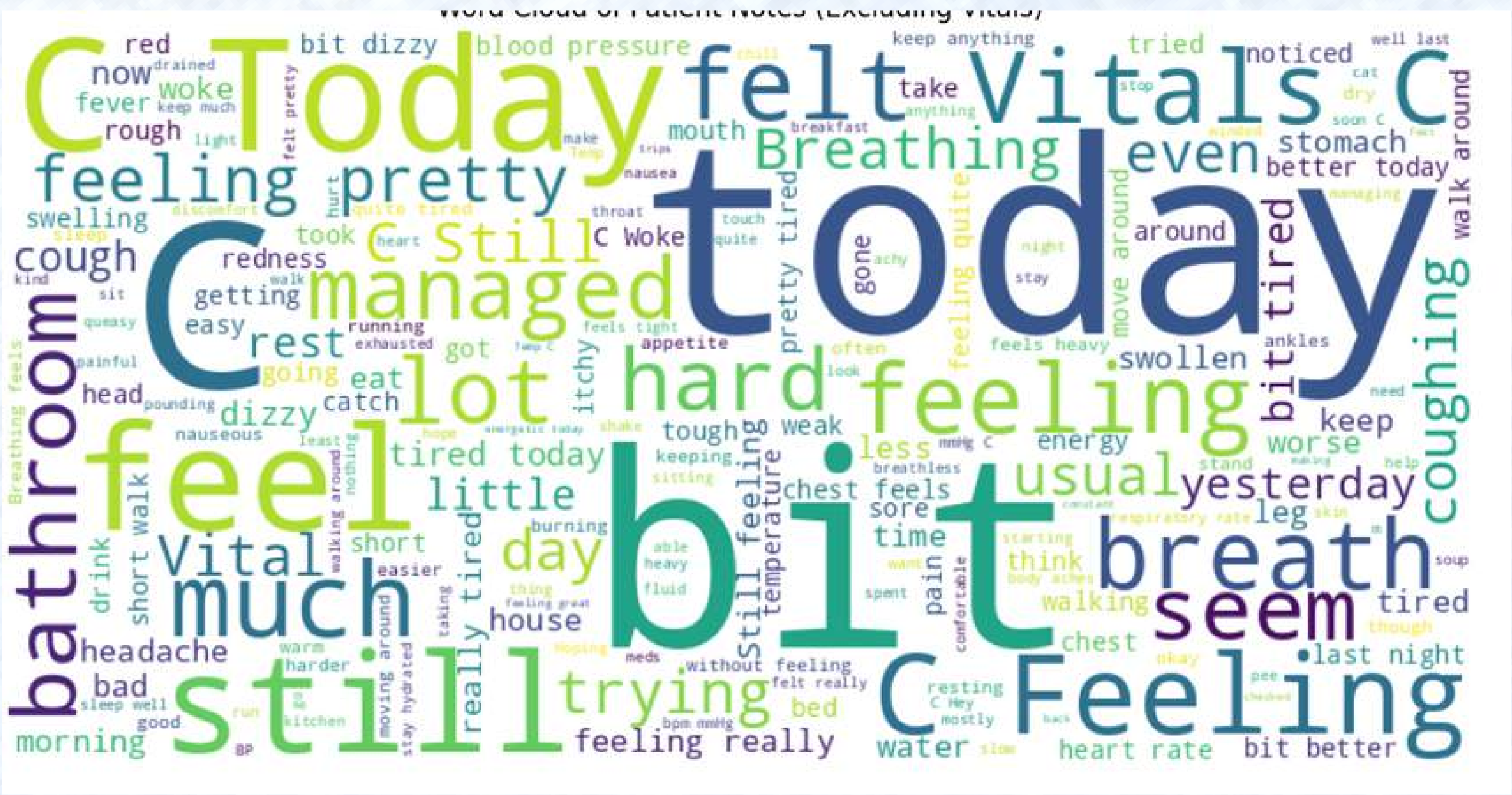
- no change : 37.2%
- Improvement: 31.5%
- Deterioration: 31.3%

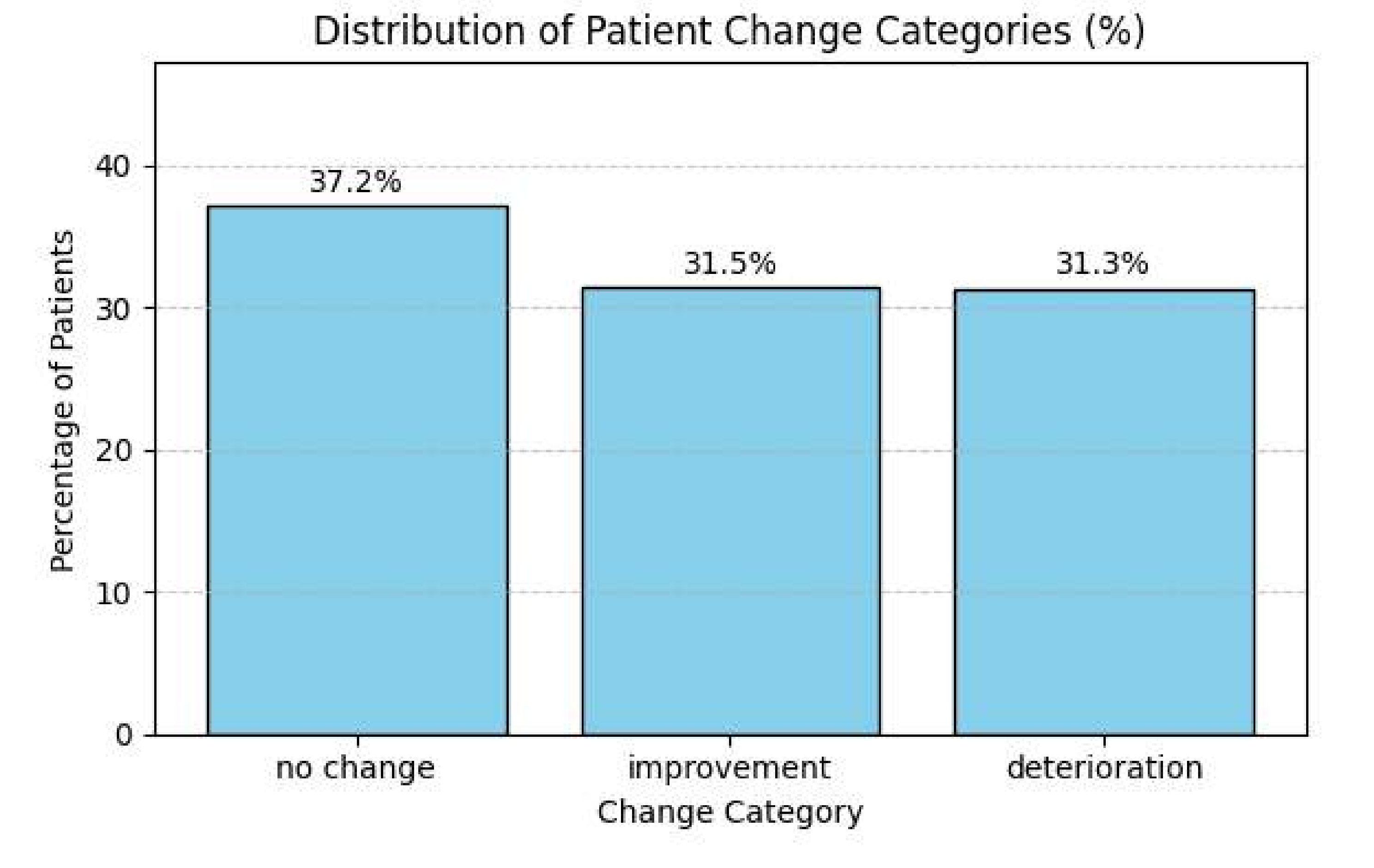
**Vitals (mean ± std):**

- HR: 74 ± 15 bpm
- $SpO_2$: 96 ± 2 %
- Temp: 36.8 ± 0.4 °C

**Age:**

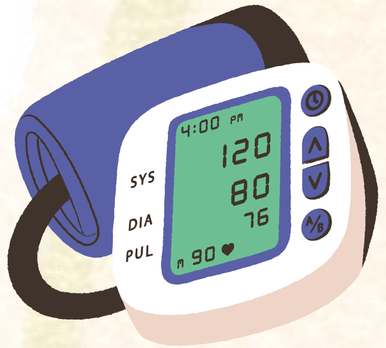
- Mean: 67.6 ± 25 yrs, range: 25–110
- Text insights
- WordCloud tokens: “pain”, “breath”, “chest”, “tired”




**Baseline**

- **Features**
  - Standardized vitals
  - TF-IDF(500 features) on patient complaints
- **Model**
  - Logistic Regression (class_weight=balanced)
- **Evaluation**
  - Stratified 5-fold CV
  - Accuracy = 0.92
  - Macro-F1 = 0.90
- **Takeaway**
  - The results are impressive, indicating that even baseline simply manages to learn the change label well from the text and metrics.




# Insights & Recommendations

**Free-text patient reports provide valuable yet noisy information**

- EDA revealed variability in report length and phrasing.
- Introducing more realistic linguistic noise and structure in synthetic data may improve generalization

**Traditional models outperformed BERT in this setting**

- Random Forest + TF-IDF + vitals outperformed the BERT-based model (Macro-F1 ~0.49 vs. ~0.40).
- Likely due to limited dataset size and class imbalance.

**Class imbalance negatively affects model performance**

- Rare classes such as Deterioration saw notably lower F1 scores.
- Applying class weights improved results marginally; data augmentation may be necessary.

**Current multimodal fusion is too simplistic**

- Simple feature concatenation was used.
- Future models should explore more advanced fusion techniques: cross-attention, co-training, or learned joint representations.

**Multitask learning could boost both accuracy and relevance**

- Simultaneous prediction of both status and recommended_action may better reflect clinical reasoning.
- Could also improve efficiency and robustness in deployment scenarios.

# Thank You for listening