



# ATTICUS

*Ambient-intelligent Tele-monitoring System*

Codice progetto: ARS01\_00860

**Deliverable D6.4**

**Studio di metafore per la visualizzazione dei dati**



**UNIVERSITÀ  
DEGLI STUDI  
DEL MOLISE**



## Data: Marzo 2020

Leader OR	DISC
Obiettivo realizzativo (OR)	OR6 – Realizzazione del DSS
Attività	6.4 – Progettazione dell'architettura del sistema
Descrizione dell'attività <i>(come da capitolato)</i>	Questa attività si pone come obiettivo la definizione di metafore per una visione semplice, immediata ed efficace del dato rilevato, sia esso inteso come dato grezzo sia come dato elaborato sia come notifica o allarme. Le metafore dovranno tenere conto anche della capacità di visualizzazione del dispositivo hardware (e.g., smartphone, monitor da scrivania). Saranno quindi definiti i tipi di dati da mostrare consistenti e compatibili con gli scenari di utilizzo del sistema ATTICUS e scelte le interfacce grafiche per la loro visualizzazione. Saranno definiti inoltre i metodi per interagire con i dati, sistemi di taggatura, esportazione ed analisi programmata.
Partner coinvolti nell'attività	<ul style="list-style-type: none"> <li>● REGMOL</li> <li>● UNIMOL</li> <li>● CON-ITA</li> </ul>
Stato del documento	Final draft
Confidenzialità	Ristretta

## Indice del documento

<b>1</b>	<b>Introduzione</b> .....	<b>7</b>
<b>2</b>	<b>Stato dell'arte</b> .....	<b>8</b>
<b>3</b>	<b>Dashboard per l'analisi dei dati</b> .....	<b>32</b>
3.1	<i>Amministratore</i> .....	32
3.2	<i>Caregiver</i> .....	33
3.3	<i>Medico Specialista e Responsabile di Monitoraggio</i> .....	35
3.3.1	Allarmi: I miei assistiti .....	36
3.3.2	Allarmi: tutti gli assistiti .....	39
<b>4</b>	<b>Interazione con i dati</b> .....	<b>59</b>
4.1	<i>Active learning</i> .....	62
4.2	<i>Approcci di labeling</i> .....	63
4.2.1	In-house labeling .....	63
4.2.2	Crowdsourcing .....	64
4.2.3	Outsourcing a privati .....	66
4.2.4	Outsourcing a società .....	66
4.2.5	Synthetic labeling .....	66
4.2.6	Data programming.....	68
4.3	<i>Tool di data labeling</i> .....	68
4.4	<i>Etichettatura dei dati</i> .....	73
4.5	<i>Esportazione programmata</i> .....	78
	<b>Riferimenti bibliografici</b> .....	<b>79</b>

## Indice delle tabelle

Tabella 1. Dashboard – Patologie.....	46
Tabella 2. Dashboard – Patologie: dettaglio Mensile.....	47
Tabella 3. Dashboard – Fascia Oraria .....	54
Tabella 4. In/outdoor .....	57
Tabella 5. Campi della maschera Etichettatura dei Dati .....	75
Tabella 6. Campi del pop-up del task da etichettare.....	76



## Indice delle figure

Figura 1. Dati sanitari mensili in Tominiski et al. [34].....	9
Figura 2. Mappa di tutti i ginecologi delle regioni della Slovenia.....	10
Figura 3. Immagine dell'applicazione. Sulla sinistra l'ambiente virtuale in 3D e sulla destra la cartella clinica elettronica in 2D.....	11
Figura 4. Architettura del sistema LOHD.....	12
Figura 5. Visualizzazione delle serie temporali sulla prevalenza di HIV in Etiopia dal 1990 al 2010. ....	13
Figura 6. Visualizzazione spaziale con sistemi GIS.....	14
Figura 7. Visualizzazione della correlazione tra la prevalenza di HIV e l'ART coverage basata su indicatore. ....	15
Figura 8. Differenti tipologie di rappresentazione dei dati. ....	16
Figura 9. Un'interfaccia della Community Health Map. ....	17
Figura 10. Grafico riportato nello studio di Poswner e Tufte [21]. ....	20
Figura 11. Interfaccia grafica di Lifelines [22]. ....	21
Figura 12. Framework di VISITORS [25].....	22
Figura 13. Visualizzazione di un singolo concetto in VISITORS [25]. ....	23
Figura 14. Visualizzazione di concetti multipli per un singolo gruppo all'interno di un periodo di tempo in VISITORS [25]. ....	23
Figura 15. Codifica visuale di DICON [23]. ....	24
Figura 16. Rappresentazione tramite ontologie.....	25
Figura 17. Prototipo dello stato di salute corrente di Meyer et al. [31].....	26
Figura 18. Prototipo dello stato di salute passato di Meyer et al. [31] .....	26
Figura 19. Mappa tematica della presenza di varie strutture sanitarie. ....	27
Figura 20. Scheletro della visualizzazione specifica per paziente [33]. ....	28
Figura 21. TEMPO.....	29
Figura 22. Interfaccia utente di Eidolon. ....	30
Figura 23. Schermata dell'applicazione MeD-Lights. ....	31
Figura 24. Esempio di stanza fisica.....	31
Figura 25. Home - Amministratore di sistema .....	32
Figura 26. Home - Caregiver.....	33
Figura 27. Home - Caregiver - Segnaposto .....	35
Figura 28. Dashboard – Menù - Medico Specialista – Responsabile di Monitoraggio .....	36
Figura 29. Dashboard - i miei assistiti - Medico Specialista – Responsabile di Monitoraggio .....	36
Figura 30. Dashboard – Medico Specialista – Responsabile di Monitoraggio – Segnaposto – i miei assistiti .....	38
Figura 31. Dashboard - Tutti gli assistiti - Medico Specialista/Responsabile di Monitoraggio .....	39
Figura 32. Dashboard - Tutti gli assistiti - Medico Specialista - Responsabile di Monitoraggio .....	40
Figura 33. Dashboard - Medico Specialista - Responsabile di Monitoraggio .....	40
Figura 34. Dashboard - Tutti gli assistiti - Patologie - Aree Geografiche - Medico Specialista - Responsabile di Monitoraggio.....	43
Figura 35. Dashboard - Tutti gli allarmi - Patologie - Mesi - Medico Specialista - Responsabile di Monitoraggio.....	44
Figura 36. Dashboard - Tutti gli assistiti - Patologie - Medico Specialista - Responsabile di Monitoraggio ....	45
Figura 37. Dashboard – Patologie - Dettaglio Mensile.....	48
Figura 38. Dashboard - Tutti gli assistiti - Patologie - Confronto - Medico Specialista - Responsabile di Monitoraggio.....	49

Figura 39. Dashboard – Tutti gli assistiti - Area Geografica – Medico Specialista – Responsabile di Monitoraggio.....	50
Figura 40. Dashboard - Tutti gli assistiti - Heatmap - Medico Specialista - Responsabile di Monitoraggio.....	51
Figura 41. Dashboard - Tutti gli assistiti - Area Geografica - Cluster - Medico Specialista - Responsabile di Monitoraggio.....	52
Figura 42. Dashboard-Area Geografica - Confronto - Medico Specialista - Responsabile di Monitoraggio....	52
Figura 43. Dashboard – Tutti gli assistiti - Area Geografica – Confronto - Dettaglio –Medico Specialista – Responsabile di Monitoraggio.....	53
Figura 44. Dashboard – Tutti gli assistiti - Fascia Oraria – Medico Specialista – Responsabile di Monitoraggio .....	55
Figura 45. Dashboard – Tutti gli assistiti – In/outdoor – Medico Specialista – Responsabile di Monitoraggio .....	58
Figura 46. Tre comuni approcci per ridurre i costi di annotazione, prima, durante e dopo il processo di training.....	61
Figura 47. Illustrazione dell'etichettatura delle immagini.....	65
Figura 48. Annotorious.....	69
Figura 49. LabelMe.....	70
Figura 50. Tagtog.....	72
Figura 51. Etichettatura dei dati – Battiti .....	74
Figura 52. Etichettatura Dati – Battiti .....	76
Figura 53. Pop up etichettatura dati .....	77

## 1 Introduzione

---

Il presente documento offre una definizione di metafore per una visione semplice, immediata ed efficace del dato rilevato. Il dato rilevato può essere visualizzato grezzo, elaborato o notificato. Le metafore tengono conto dell'interfaccia grafica del dispositivo hardware.

Il documento è organizzato in quattro capitoli, incluso il presente:

- il Capitolo 2 riporta lo stato dell'arte sulle metafore visuali, tendendo conto di simulazioni virtuali e di interazioni utenti;
- il Capitolo 3 descrive la Dashboard e la Home di ATTICUS, mostrando la visualizzazione aggregata delle informazioni in base al ruolo dell'utente loggato;
- il Capitolo 4, infine, descrive i metodi a disposizione degli utenti per interagire con i dati

## 2 Stato dell'arte

In letteratura non sono presenti molte metafore visuali a supporto della visualizzazione dei dati biomedici e sanitari.

Althoff *et al.* [5] hanno effettuato uno studio comparativo valutando le interazioni dell'utente su un'applicazione desktop di realtà virtuale (DVA – *desktop Virtual-Reality application*) e un'applicazione automotive infotainment (AIA – *automotive infotainment application*), quindi una combinazione basata su touchscreen e sulle gesture delle mani e della testa e un'altra basata su touchscreen e una console con mouse e tastiera. Gli autori hanno voluto studiare la tendenza degli esperti e degli utenti normali con l'uso di questi dispositivi. In questo modo hanno potuto vedere che gli esperti sono più propensi ad utilizzare i dispositivi tattili, mentre gli utenti normali preferiscono la combinazione di più dispositivi.

Tominiski *et al.* [34] riportano un metodo di visualizzazione dei dati spazio – temporale sulle mappe, che è basato sulla visualizzazione dell'informazione 3D e sull'information hiding. Nello specifico, su una mappa del sistema GIS vengono posizionate delle icone 3D che vogliono riportare dei dati spazio-temporali. Ad esempio, nella Figura 1 vengono visualizzati i dati sanitari mensili con delle icone 3D sulla mappa. Precisamente, Figura 1(a) vengono riportati i dati di 6 tipi di malattia (influenza, problemi legati alle vie respiratorie superiori, problemi legati al tratto digestivo, problemi lievi, malattie relative alla psiche, unioni) attraverso le icone a forma di matita. Ogni faccia della matita rappresenta una malattia. Questa forma permette di visualizzare i dati ottenuti nel tempo, codificati dalla punta della matita su tutta la lunghezza della matita. Il numero delle facce di una matita può dipendere dal numero di attributi da visualizzare. Invece, nella Figura 1(b), i dati vengono visualizzati attraverso delle icone a spirale, rivelando le caratteristiche cicliche di 2 malattie selezionate. Per ogni icona ad elica viene creato un nastro. Per ogni passo temporale il nastro si estende in angolo e altezza, i quali dipendono sul numero di primitive temporali per ciclo e il numero di passi ciclici. Il colore è utilizzato per codificare i valori dei dati lungo il nastro.

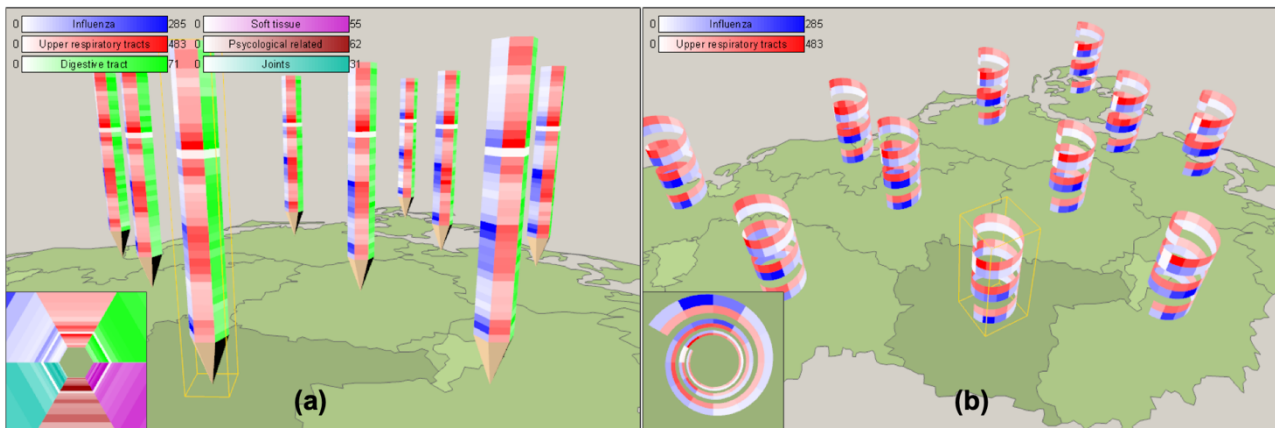


Figura 1. Dati sanitari mensili in Tominiski et al. [34].

Lavrač *et al.* [29] sfruttano il data mining e le tecniche di visualizzazione come supporto alle decisioni per la pianificazione e la gestione a livello regionale della sanità pubblica in Slovenia. In questo modo è possibile comprendere quali sono le aree che sono atipiche in termini di accessibilità e disponibilità dei servizi di sanità pubblica per la popolazione. I risultati sono utili nel processo di decisione delle autorità sanitarie locali e regionali. Lavrač *et al.* [29] hanno creato MediMap, un'applicazione di data mining e di supporto alle decisioni nella sanità pubblica slovena. In particolare, vengono usati dei metodi di data mining descrittivo, che combinati con tecniche di supporto alle decisioni multicriterio migliorano la gestione dei dati e la conoscenza dell'Istituto di Sanità Pubblica della regione Celje. Come strumento di visualizzazione vengono usati i dati GIS, *Geographic Information System*, in modo da localizzare i centri di salute sulla mappa della Slovenia (Figura 2).



Figura 2. Mappa di tutti i ginecologi delle regioni della Slovenia.

In letteratura esiste anche lo studio di Fukazawa *et al.* [6], in cui vengono comparate le tecniche di interazione multimodale in un ambiente multi-display basato su perspective-corrected (MDE - *multi-display environment*). Per effettuare questa comparazione sono state misurate le performance di interazioni multimodali basandosi sulle gesture, sulla direzione della testa e sullo sguardo degli occhi, comparandoli con l'uso del mouse. Un altro studio comparativo è stato effettuato da Bowman *et al.* [7] che propongono uno studio sistematico sull'applicazione delle tecniche di interazione in un ambiente virtuale (VE - *virtual environment*).

Invece, Dukes *et al.* [4] si sono focalizzati su una simulazione sanitaria in modalità multitasking, in cui sono presenti due attori importanti: gli infermieri e i pazienti. In questo modo gli autori hanno potuto riprodurre la scena reale, utilizzando un ambiente virtuale 3D, registrando i dati in un sistema di cartelle cliniche elettroniche basato su un'interfaccia grafica 2D. Precisamente, hanno voluto mostrare se una metafora con un'interazione tradizionale basata su mouse e tastiera per

sottosistemi virtuali e interfaccia grafica era superiore, in termini di preferenze dell'utente, e più performante rispetto a un'interazione ibrida, ottenuta attraverso l'uso di touch screen per l'ambiente virtuale e mouse e tastiera per l'interfaccia grafica. Con questo lavoro, gli autori vogliono fornire una soluzione basata sul touch screen che sia il più naturale possibile per i propri tirocinanti, avendo un paziente virtuale, l'ambiente e un'interazione familiare attraverso l'utilizzo delle cartelle elettroniche sanitarie. In Figura 3 è mostrata una figura dell'applicazione fornita da Dukes *et al.* [4].



Figura 3. Immagine dell'applicazione. Sulla sinistra l'ambiente virtuale in 3D e sulla destra la cartella clinica elettronica in 2D.

Tilahun *et al.* [30] hanno analizzato l'uso delle tecnologie dei Linked Data per vedere se possono essere usate come opzioni di visualizzazione. Linked Data è un metodo per pubblicare e rendere accessibili dati strutturati usando le tecnologie Web standard, che usano gli HTTP URI – *Uniform Resource Identifier* – per l'identificazione degli elementi, il *Resource Description Framework* (RDF) per descrivere i dati e i link per descrivere le relazioni. Da questo studio preliminare, loro hanno creato il sistema Linked Open Health Data (LOHD) per arricchire i dati sanitari spaziali e statistici da varie fonti. La struttura di questo sistema è composta da 4 livelli:

1. **Data layer:** componente che permette la conversione e il collegamento tra i dati;
2. **Transformation layer:** componente che processa ogni query SPARQL con il crawler dei pattern per localizzare i dati dal Web dei Linked Data;
3. **Service layer:** componente che verifica l'accesso ai dati e collega il client e il server tramite i protocolli di servizio;

4. **Presentation layer:** componente che consente agli utenti di interagire con i servizi usando tool di visualizzazione o di recupero.

È possibile visionare la struttura in Figura 4.

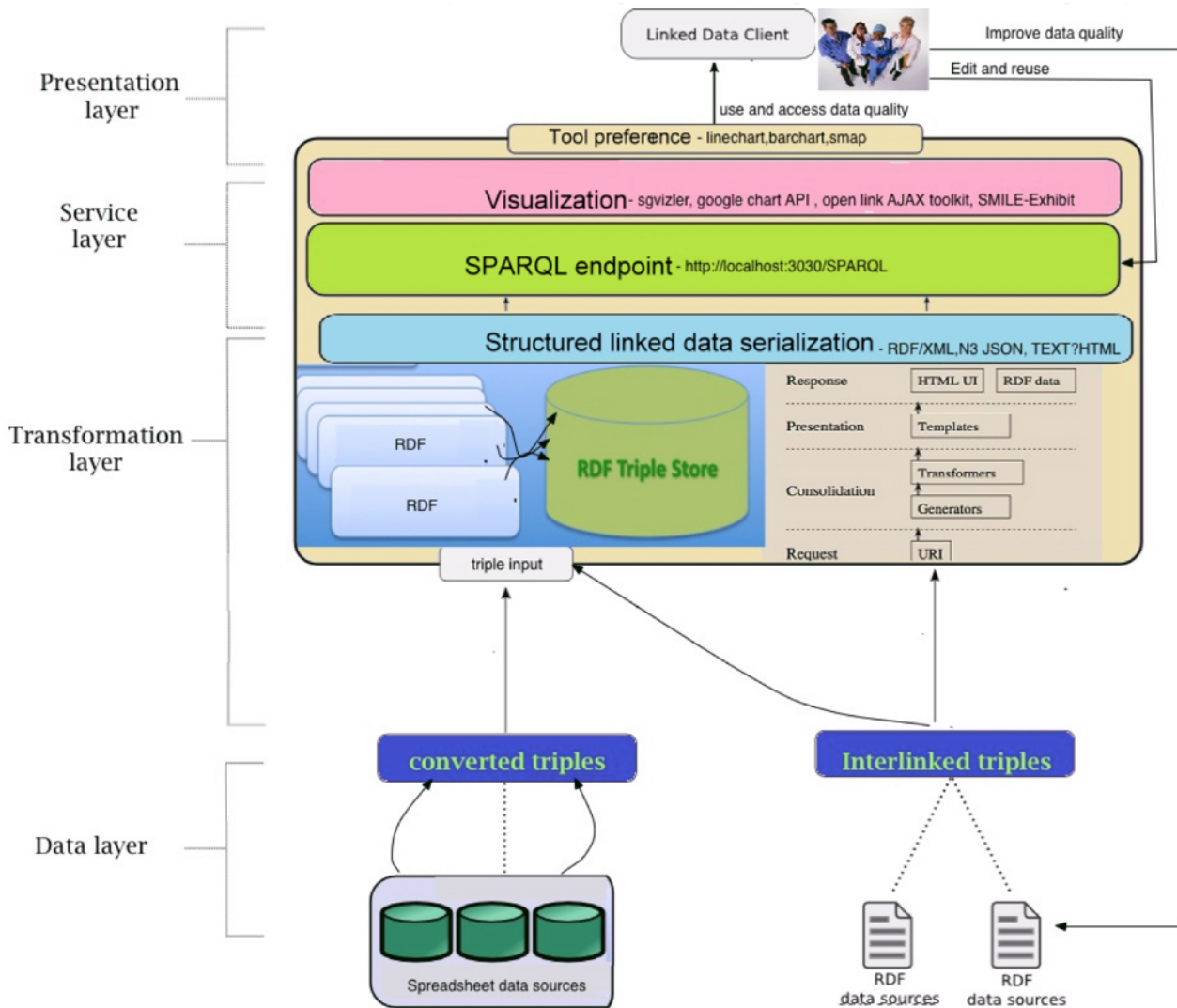


Figura 4. Architettura del sistema LOHD.

Tilahun *et al.* [30] applicano delle visualizzazioni considerando delle query spaziali o temporali. Nel caso delle serie temporali, esse permettono di far capire quali sono i trend e i pattern. Ad esempio, in Figura 5 viene mostrato il trend sulla prevalenza di HIV in Etiopia.

```

PREFIX lohd: <http://localhost:3030/lohd/data#>
PREFIX loh: <http://localhost:3030/lohd/d#>
PREFIX qb: http://purl.org/linked-data/cube#

SELECT ?year xsd:decimal(?prevalence)
where
{
?lohd loh:year ?year;
loh:countryname "Ethiopia";
qb:prevalence ?prevalence.
}
order by ?year

```

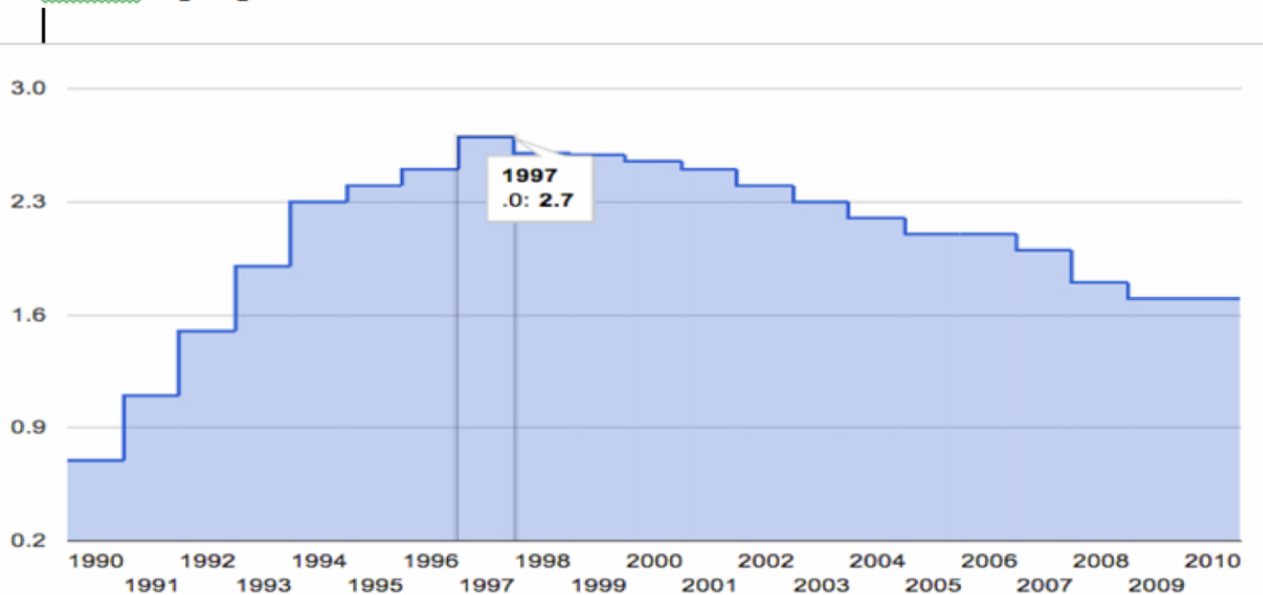


Figura 5. Visualizzazione delle serie temporali sulla prevalenza di HIV in Etiopia dal 1990 al 2010.

Invece, se consideriamo la visualizzazione spaziale, Tilahun *et al.* [30] usano i sistemi GIS con le visualizzazioni location-based che aiutano nella scrittura di query e nella scelta del metodo di visualizzazione, come, ad esempio, in Figura 6 è possibile visualizzare la diffusione dell'HIV per ogni paese.

```

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX lohd: <http://localhost:3030/lohd/data#>
PREFIX loh: <http://localhost:3030/lohd/d#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX wgs84: <http://www.w3.org/2003/01/geo/wgs84_pos#>

SELECT xsd:decimal(?lat) xsd:decimal(?lon) ?name ?text ?url ?image
WHERE {
  ?lohd wgs84:lat ?lat;
    wgs84:long ?lon;
    geo:name ?name.

  OPTIONAL {
    ?lohd rdfs:isDefinedBy ?url;
    geo:image ?image;
    loh:label ?text;
    geo:image ?image . }
}

```

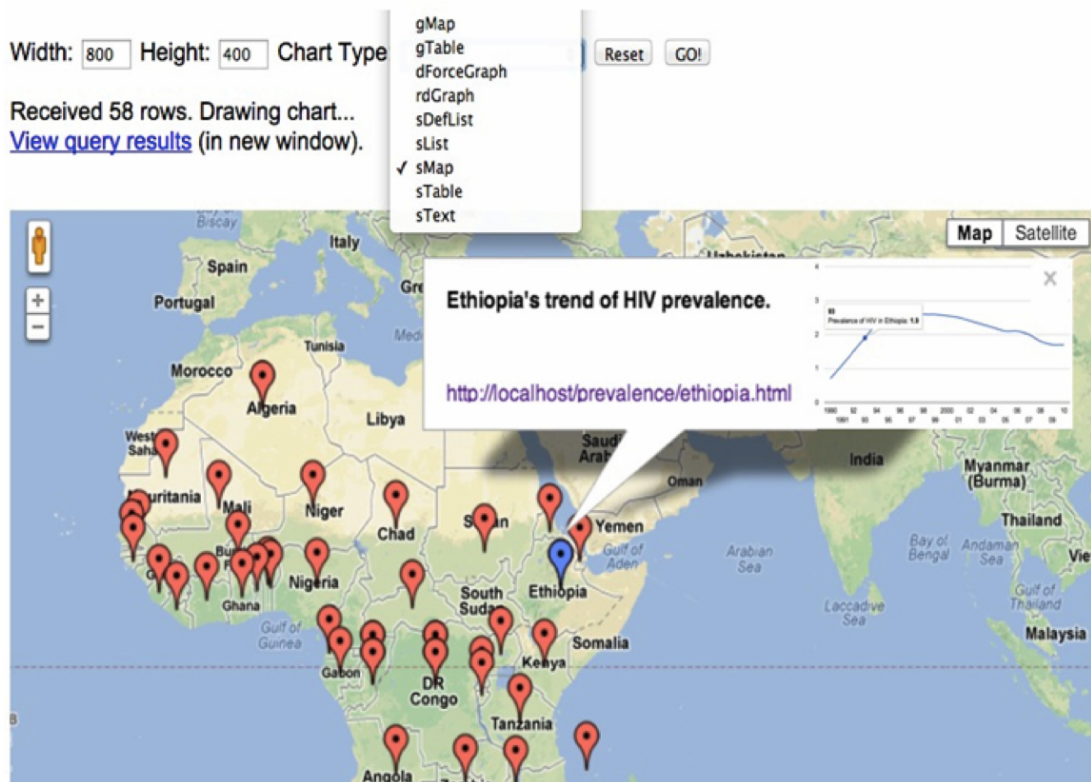


Figura 6. Visualizzazione spaziale con sistemi GIS.

Un altro tipo di visualizzazione è quella basata sugli indicatori dei Linked Data (*Indicator-Based Visualization of Linked Data*), dove ogni indicatore è un dato sanitario. Un esempio è possibile

visionarlo in Figura 7, ossia una visualizzazione della correlazione tra la prevalenza di HIV e l'ART (*antiretroviral therapy* – terapia antiretrovirale)<sup>1</sup> coverage basata su indicatore [30].

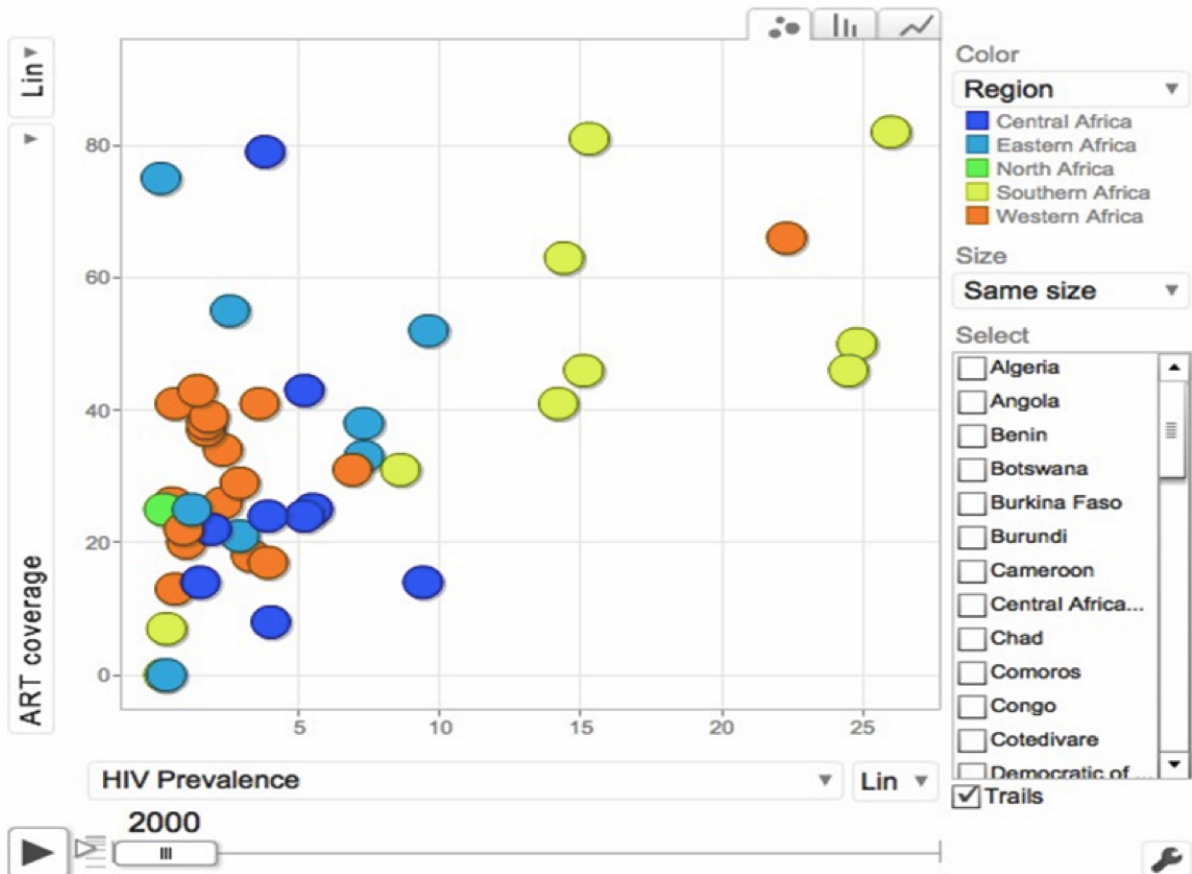


Figura 7. Visualizzazione della correlazione tra la prevalenza di HIV e l'ART coverage basata su indicatore.

O'Donoghue *et al.* [15] riportano le diverse metodologie per la visualizzazione dei dati biomedici presentate in letteratura analizzando in che modo queste tecniche di visualizzazione vengono adoperate nelle aree di ricerca biomedica, prestando attenzione al sequenziamento di RNA a singole-cellule (RNA-seq) e all'analisi della struttura delle proteine o la chirurgia assistita dalla realtà aumentata (VR). Nonostante ciò, le tecniche di visualizzazione di dati biomedici sono utilizzate molto poco oppure vengono utilizzati strumenti poco validi, contribuendo a possibili problematiche sulle diagnosi, come la mancata identificazione di particolari anomalie. Infatti, la maggior parte degli

<sup>1</sup> **Antiretroviral therapy:** consiste della combinazione di minimo 3 farmaci antiretrovirali per rimuovere il virus HIV al massimo e rallentare la sua progressione.

errori diagnostici sorgono come conseguenza di un'errata elaborazione e di un errato processamento dei dati. Una corretta applicazione delle tecniche di visualizzazione può aiutare ad arginare tali problemi [16, 17]. Per avere però una visualizzazione dei dati più precisa ed efficace, bisognerebbe avere grandi volumi di dati. O'Donoghue *et al.* [15] si sono focalizzati nel campo della genomica e dell'epigenetica, evidenziando un avanzamento delle tecnologie usate per il sequenziamento del DNA che porta però alla generazione di una grande quantità di dati. In Figura 8 sono visualizzati i differenti tipi di rappresentazione dei dati. Nel campo della biologia, la sequenza del RNA deve essere visualizzata per capire se ci sono eventuali problemi. Ad esempio, se il numero di righe e colonne aumenta o la dimensione della cella è ridotta rendendo impossibile la visualizzazione di tutti i risultati significativi come, ad esempio, una heat map molto grande.

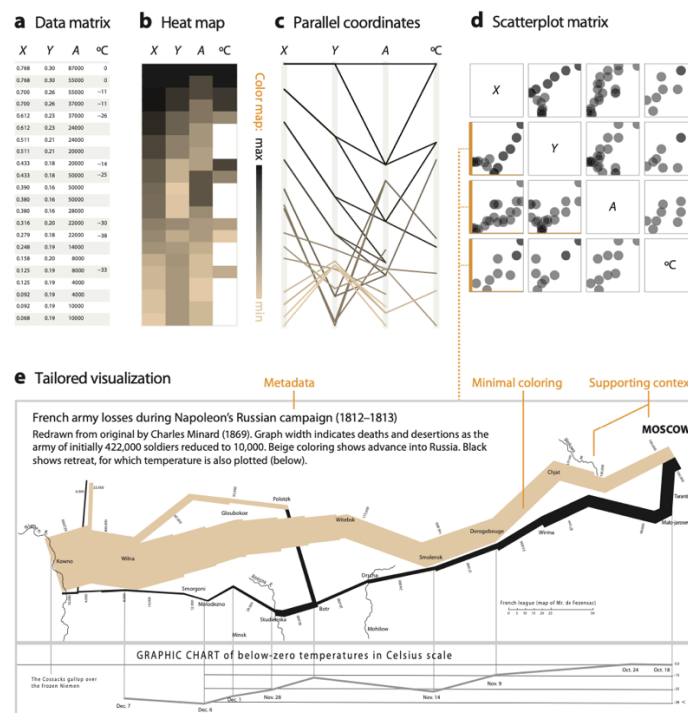


Figura 8. Differenti tipologie di rappresentazione dei dati.

Sopan *et al.* [32] hanno sviluppato la Community Health Map, un'applicazione web per visualizzare i dati delle cure sanitarie in uno spazio ampio. Le visualizzazioni vengono fatte attraverso mappe interattive in modo da comparare due o più variabili usando i grafici e le tabelle. La Community Health Map è creata per gli utenti con esperienza nell'industria sanitaria per poter: (i) visualizzare

le prestazioni sanitarie, l'accesso e gli indicatori di qualità per la loro regione; (ii) filtrare le regioni attraverso le caratteristiche demografiche (e.g., età, reddito, povertà, o educazione); (iii) produrre risultati visuali in modo da agire per migliorare la salute e la qualità dell'assistenza sanitaria nelle loro regioni. In Figura 9 è mostrata un'interfaccia della Community Health Map.

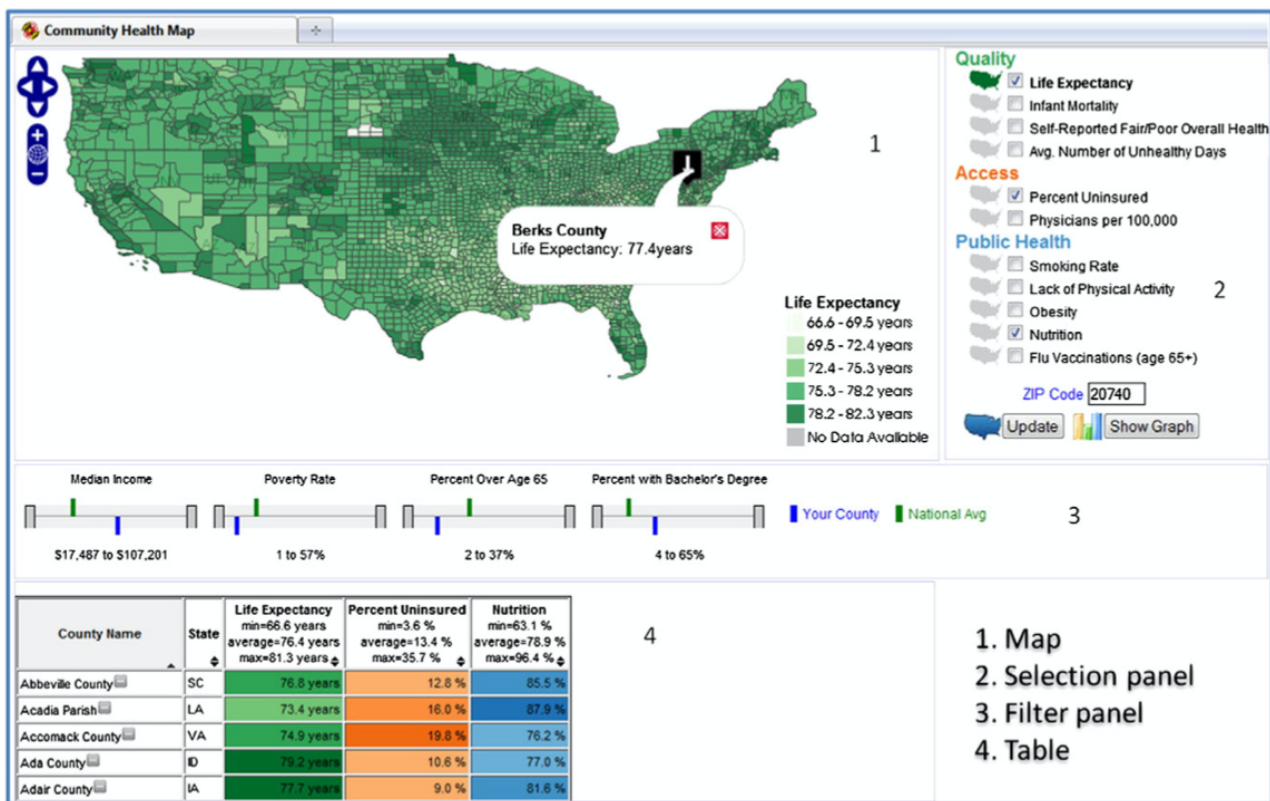


Figura 9. Un'interfaccia della Community Health Map.

Shneiderman *et al.* [19] si focalizza sulla letteratura di tre domini, ossia sui programmi di salute personale (*personal health programs*), sulla consegna clinica sanitaria (*clinical healthcare delivery*) e sulla definizione delle politiche in materia di sanità pubblica (*public health policy making*). Inoltre, esso descrive alcune sfide per la visualizzazione dell'informazione e per i ricercatori di analisi visuale. Nel caso di un cancro, sono stati fatti grandi progressi nel sequenziamento dell'intero genoma per l'identificazione di specifici errori di trascrizione. L'Institute of Medicine (IOM) afferma che la tecnologia dell'informazione sanitaria consentirà potenti analisi dei dati visivi in modo da ottenere una maggiore assistenza e scoprire più velocemente la presenza del cancro [40]. Anche l'American Society of Clinical Oncology (ASCO) informa che l'uso di un sistema di analisi sanitaria visual-

analytics-infused può aiutare nello sviluppo di un sistema di apprendimento rapido per la cura del cancro [40]. Con la visualizzazione i pazienti, i medici e tutti gli operatori sanitari possono accedere al dato per il monitoraggio delle loro condizioni, per supporto, per riportare i sintomi o per comunicare asincronamente con tutto il team dedicato alla cura. Inoltre, loro hanno riportato le seguenti 7 sfide:

1. **Offrire ai medici disponibili informazioni tempestive nel formato giusto:** con la graduale raccolta dei dati è necessario avere degli strumenti di analisi visiva in modo da visionare completamente le funzioni sanitarie. Infatti, la progettazione visiva del flusso di lavoro e gli strumenti di presentazione sono usati per adattare i loro sistemi EHR alle loro pratiche;
2. **Muoversi attraverso un ecosistema di tool visuali:** con il progresso tecnologico tutti gli operatori sanitari e i familiari possono monitorare lo stato di salute dell'assistito e allertare il medico di famiglia in caso di anomalia, dato che sono stati creati degli strumenti visivi di facile comprensione. Infatti, saranno fornite delle interfacce utente ben progettate con presentazioni visive per i pazienti e i loro assistenti in modo da riuscire a migliorare la qualità della vita dei pazienti di tutte le età;
3. **Facilitare il processo di decision-making del team:** data l'interoperabilità dei sistemi biomedici, il processo di decision-making passa da individuale a processo di gruppo. Normalmente, il processo medico di decision-making orientato al team è complicato perché bisogna identificare la responsabilità legale. Utilizzando questa tipologia di sistemi, si può dare la possibilità a tutti i membri del team di esprimere il proprio consenso su una eventuale decisione.
4. **Caratterizzare e comprendere la somiglianza:** la visualizzazione dovrà essere strumento nell'aiuto dei medici per risolvere i casi difficili (come le malattie rare).
5. **Visualizzazione dell'efficacia comparativa e delle relazioni causa ed effetto:** le visualizzazioni aiuteranno a caratterizzare le differenze e faciliteranno gli studi dell'efficacia comparativa per abilitare i ricercatori e i responsabili della sicurezza dei pazienti di valutare metodi e trattamenti.

6. **Presentare rischio e incertezza:** la presentazione di risorse incerte per dati sanitari personali sono dei casi eccezionali. Le visualizzazioni interattive appaiono utili per la presentazione di rischi di trattamenti, medicazioni o errori.
7. **Evaluation:** I disegnatori di interfacce visuali hanno la necessità di apprendere molto sulla salute e sulla sanità. Ad esempio, delle interfacce progettate male potrebbero fornire una cattiva interpretazione.

West *et al.* [20] riporta tutte le tecniche di visualizzazione riportate dal 1996 al 2013. Da questo lavoro riassuntivo, è possibile notare che Powsner e Tufte [21] e Plaisant *et al.* [22] sono stati i primi a visualizzare i dati dei pazienti tramite un grafo. In Figura 10 viene mostrato il grafico di un paziente che è stato curato per la sua malattia maniaco depressiva riportato nello studio di Powsner e Tufte [21].

Plaisant *et al.* [22] hanno creato LifeLines che sfrutta i dati dal Maryland Department of Juvenile Justice per fornire una visualizzazione per ogni singolo paziente. In particolare, viene creata una linea temporale di tutti gli eventi di un singolo paziente, dove sull'asse orizzontale è riportato il tempo e sull'asse verticale sono visualizzati gli eventi che riguardano, ad esempio, le allergie, le diagnosi e le medicazioni. In Figura 11 è mostrata una schermata di LifeLines. Successivamente, c'è stata una evoluzione di LifeLines in LifeLines2 [24] per visualizzare più record di pazienti insieme. Inoltre, possono essere visualizzati i dati numerici e categorici.

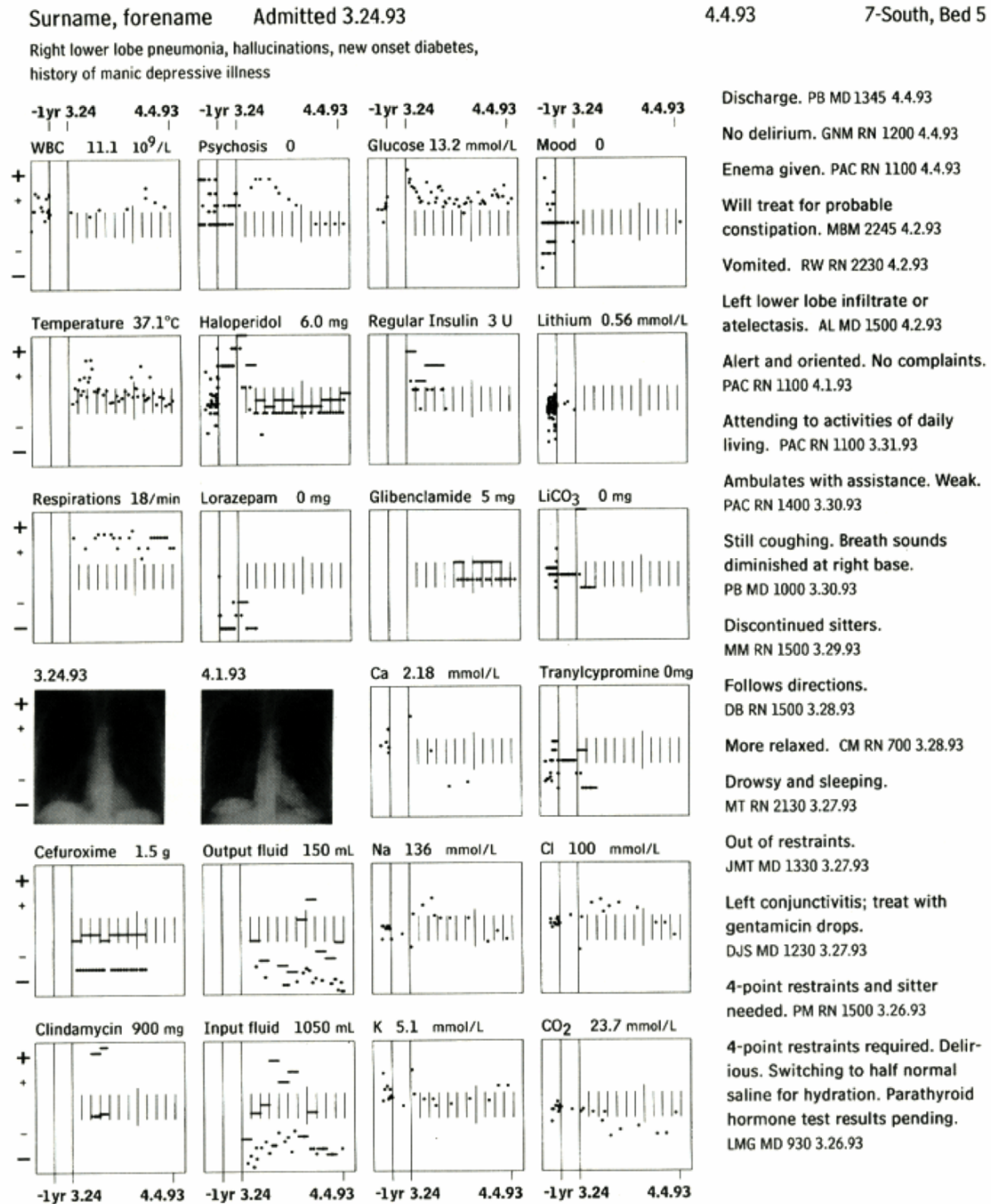


Figure 2: Graphical summary of patient status

Notes on right show initials and professional qualification (MD = medical doctor, RN = registered nurse), time, and date (American format). — = period of dosing.

Figura 10. Grafico riportato nello studio di Poswner e Tufta [21].

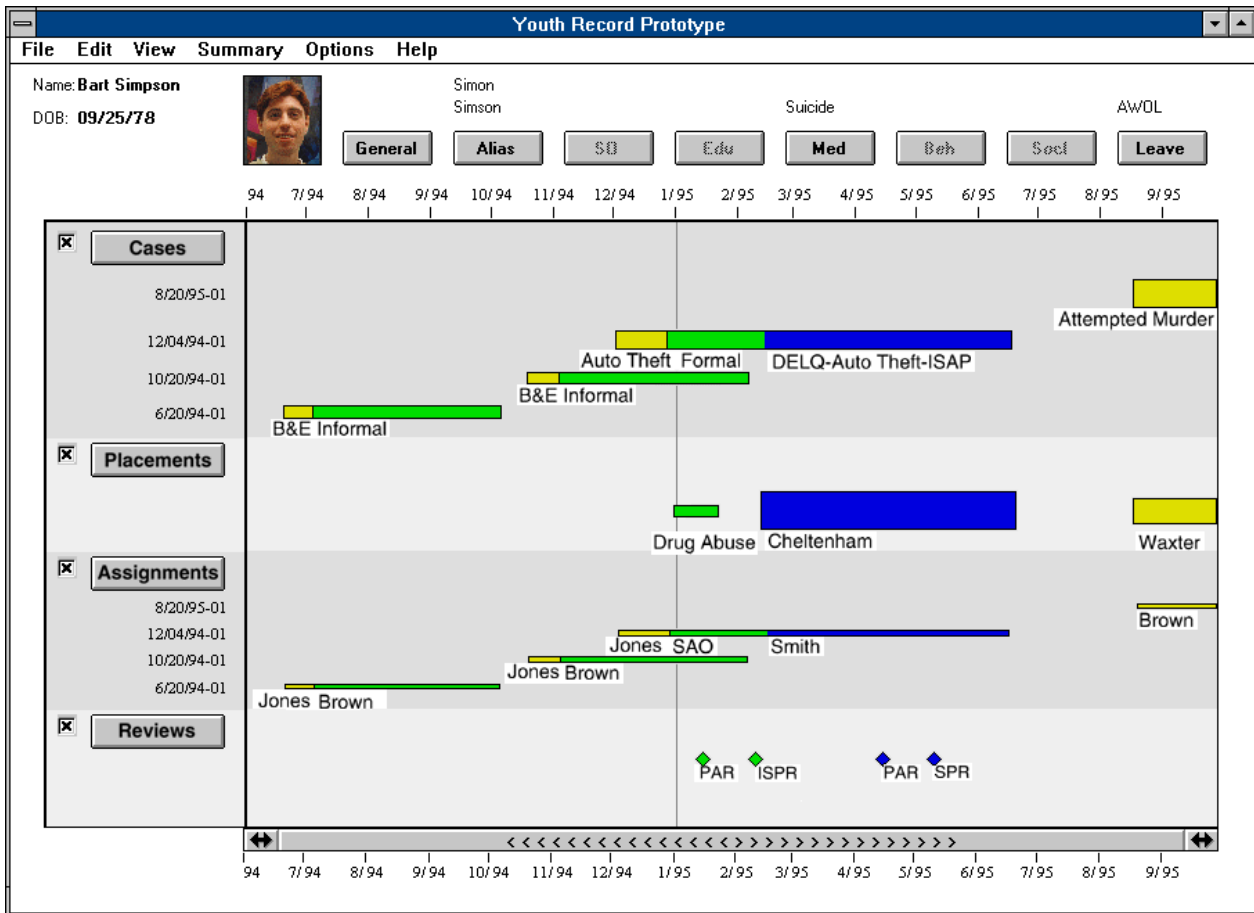


Figura 11. Interfaccia grafica di LifeLines [22].

Un'altra visualizzazione innovativa è VISITORS (*VISualization of Time-Oriented Records*) [25], che viene usato per visualizzare ed esplorare i dati grezzi e i concetti astratti (ad esempio, la temperatura), che possono essere ottenuti dai dataset clinici orientati al tempo.

In Figura 12 è mostrato il framework di VISITORS che si distingue per tre caratteristiche: trattamento di record multipli, trattamento della dimensione temporale e interfaccia utente. VISITORS fornisce vari tipi di visualizzazione, ad esempio, la visualizzazione di un singolo concetto sul tempo ("Visualization of a single-concept over time") che sarebbe la visualizzazione di un concetto per un gruppo di pazienti mostrata in un intervallo di tempo (Figura 13). Un altro tipo di visualizzazione è la visualizzazione di concetti multipli per un singolo gruppo all'interno di un periodo di tempo ("Visualization of multiple-concepts for a single-group within a time period"), che viene utilizzata

quando bisogna comparare un numero di concetti per uno specifico gruppo di pazienti (Figura 14). L'ultimo tipo di visualizzazione è quella a singolo gruppo, singolo concetto, record di tempo multipli, all'interno di uno specifico intervallo di tempo ("single-group, single-concept, multiple time records, within a given time range"). Questo tipo di visualizzazione deve essere usata per determinare il pattern che caratterizza un periodo di tempo tipico, quando per esempio c'è un trattamento di pazienti cronici, come i diabetici.

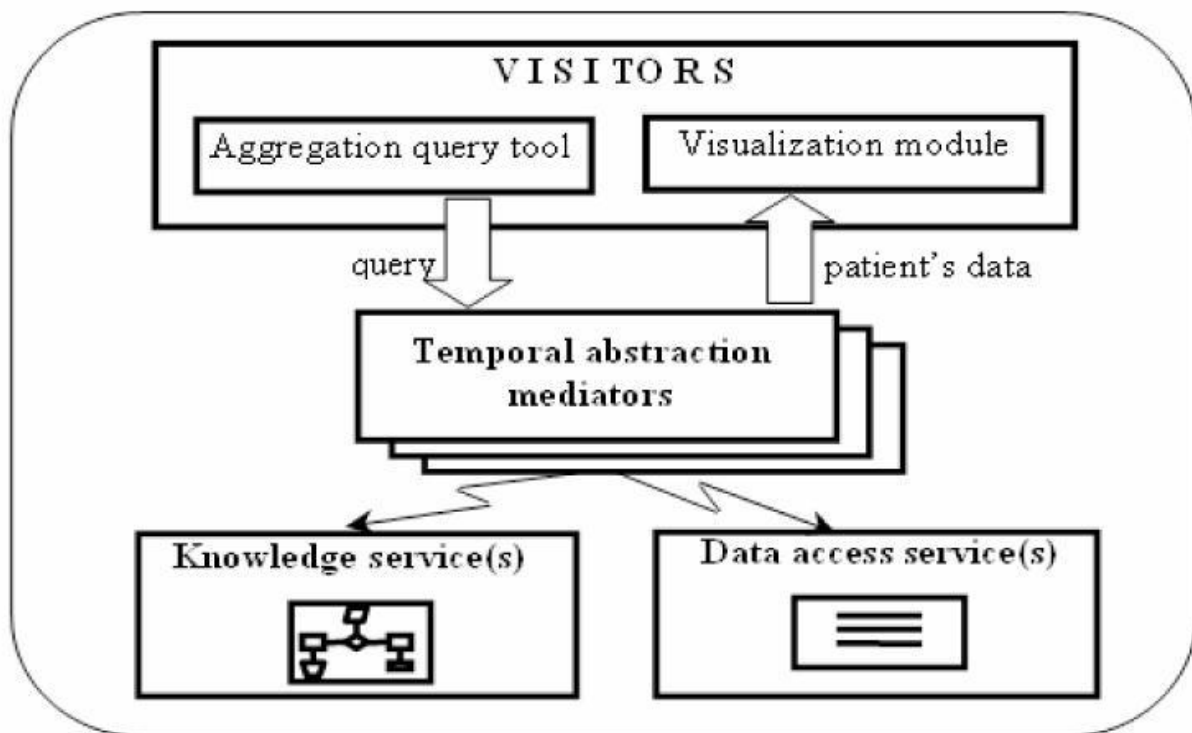


Figura 12. Framework di VISITORS [25].

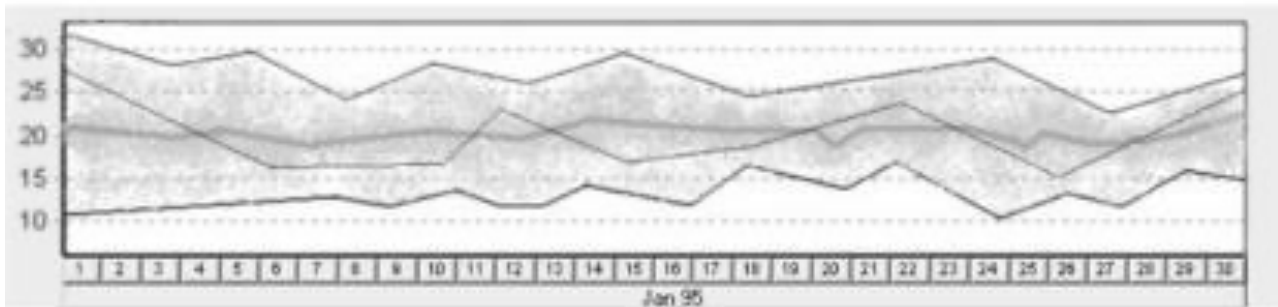
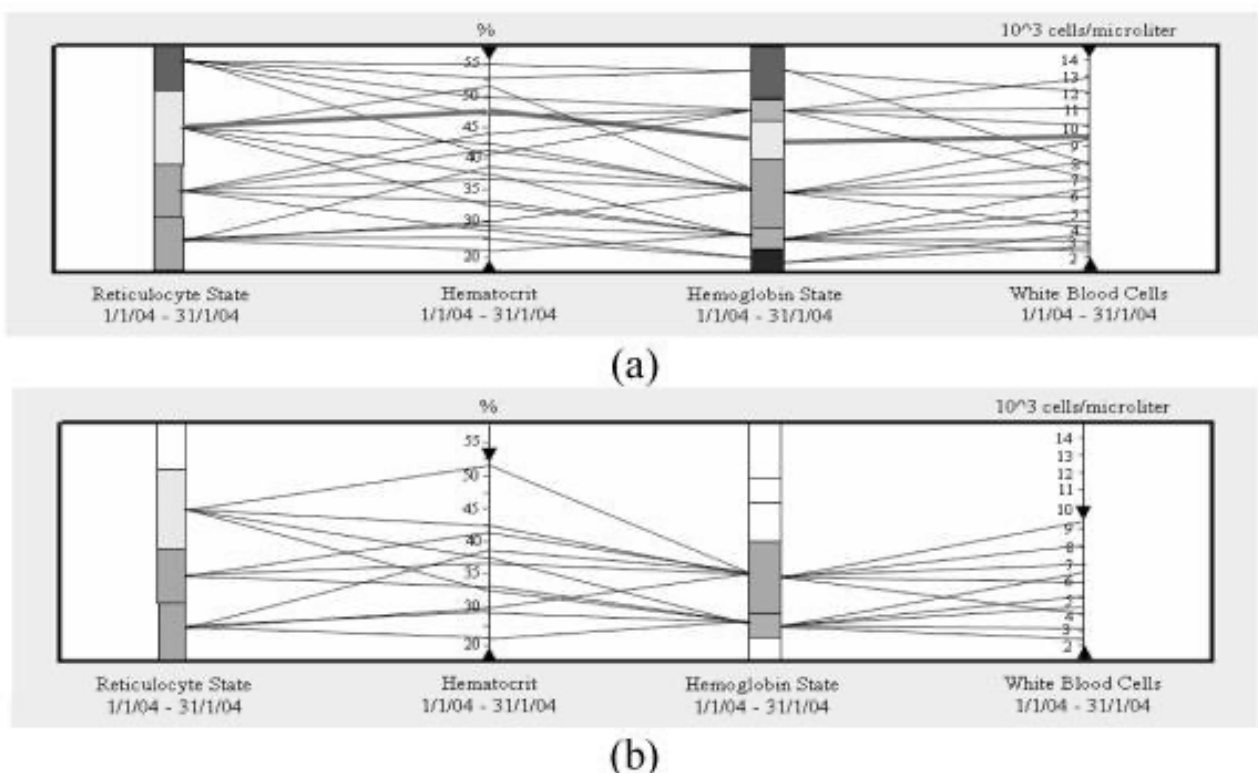


Figura 13. Visualizzazione di un singolo concetto in VISITORS [25].



Legend Value (from bottom to top): Reticulocyte State – low, normal, moderately high, high.  
 Hemoglobin state – very low, low, normal, moderately high, high, very high

Figura 14. Visualizzazione di concetti multipli per un singolo gruppo all'interno di un periodo di tempo in VISITORS [25].

Gotz *et al.* [23] hanno creato DICON (Dynamic Icons), ossia una tecnica di visualizzazione per esplorare i gruppi di pazienti simili. L'utente può visionare i gruppi rappresentati come icone su una mappa ad albero (*treemap*<sup>2</sup>). L'interfaccia permette di analizzare i dati usando le icone dinamiche.

<sup>2</sup> **Treemap**: metodo per la visualizzazione di dati gerarchici usando rettangoli innestati.

DICON sfrutta una metodologia di metafora visuale che consiste nel rappresentare l'insieme di caratteristiche di ogni paziente sotto forma di aree colorate. In Figura 15 è mostrata la codifica visuale di DICON.

Gotz and Wongsuphasawai [26,27] hanno creato Outflow, che è usato per guardare come avviene il progresso della malattia. Attraverso questo strumento, gli utenti possono verificare gli eventi multipli e le sequenze degli eventi per capire quali sono i fattori che sono correlati con specifici percorsi.

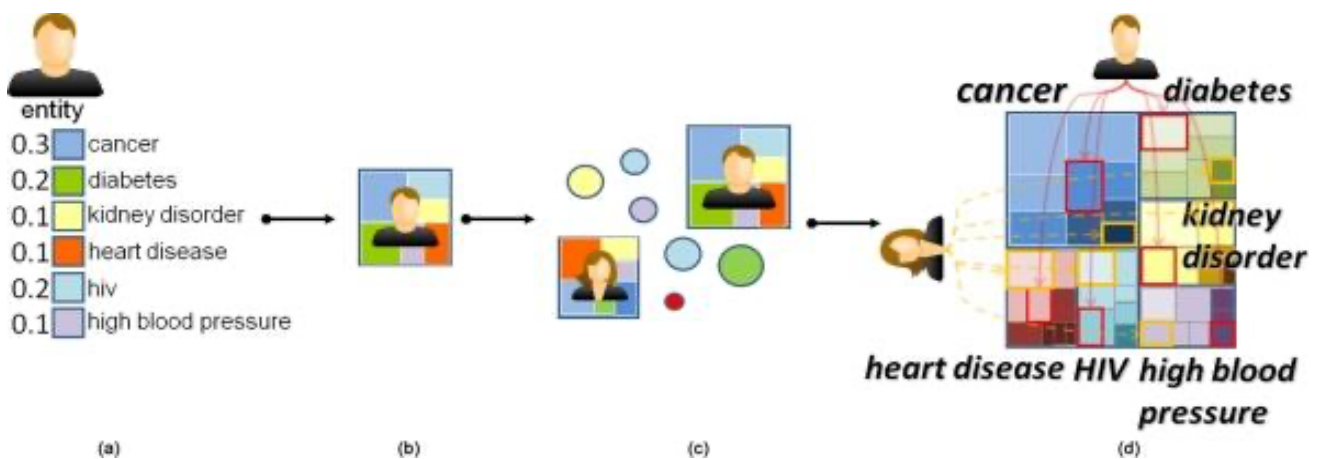


Figura 15. Codifica visuale di DICON [23].

Population Health Record (PopHR) [28] è un'applicazione web semantica per l'automatizzazione dell'integrazione ed estrazione di grandi quantità di dati eterogenei provenienti da differenti sorgenti (es., i dati amministrativi e clinici) con un'elevata risoluzione grafica. In questo modo è possibile supportare la misurazione e il monitoraggio delle prestazioni dei sistemi sanitari. Nejad *et al.* [28] creano questa piattaforma basata sui big data per la sorveglianza sanitaria della popolazione e della comunità situata in Canada. In Figura 16 viene mostrata una visualizzazione tramite ontologie usata nel PopHR.

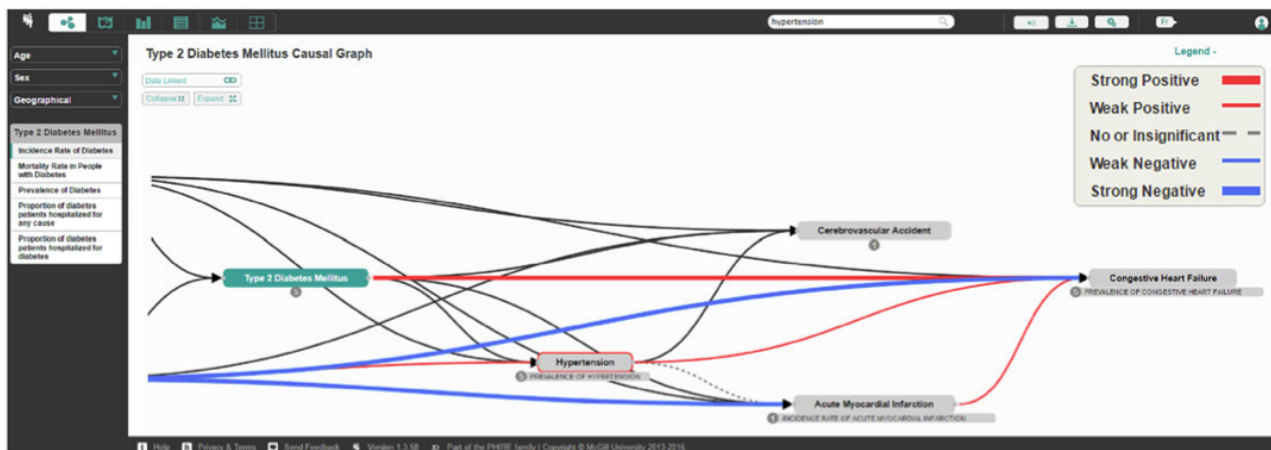


Figura 16. Rappresentazione tramite ontologie.

Meyer *et al.* [31] hanno affrontato le problematiche legate alla visualizzazione di grandi volumi di dati sanitari su dispositivi mobili, dove nonostante la limitata grandezza del display e quindi le limitate possibilità di interazione, è per la maggior parte delle persone lo strumento principale per l'accesso e il controllo dei propri dati sanitari. Gli autori hanno trovato la necessità di visualizzare lo stato di salute corrente per ogni singolo paziente delle ultime 72 ore e successivamente hanno visualizzato in modo più dettagliato i dati degli ultimi 7 giorni. Per fornire una visualizzazione veloce dello stato di salute corrente, Meyer *et al.* [31] hanno creato alcune metafore visuali nelle dashboard. Inoltre, gli autori adattavano queste visualizzazioni chiedendo una valutazione agli utenti. Un primo prototipo è mostrato in Figura 17.



Figura 17. Prototipo dello stato di salute corrente di Meyer et al. [31].

Un altro tipo di visualizzazione è stata creata per lo stato di salute degli ultimi 6 mesi (Figura 18).

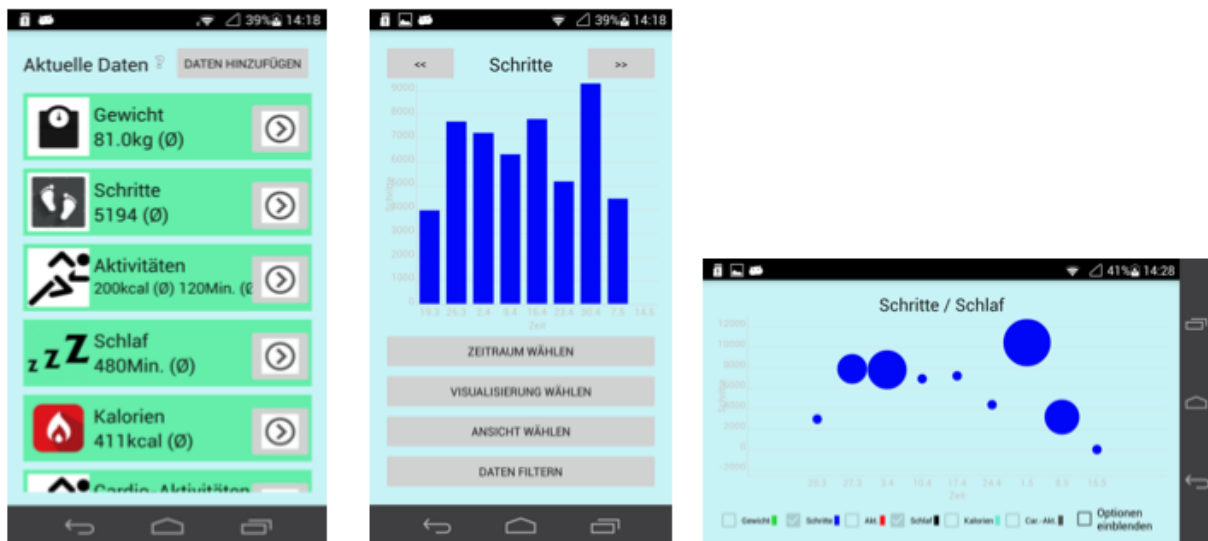


Figura 18. Prototipo dello stato di salute passato di Meyer et al. [31]

Mahmud *et al.* [35] hanno sviluppato un framework per la visualizzazione e l'analisi di dati per la predizione di health-shocks su dataset sanitari di grandi dimensioni. Con il termine health-shocks identificano gli effetti socioeconomici su un individuo, la sua famiglia e la società dovuti alla malattia critica di un familiare. Il framework è stato sviluppato su Amazon Web Services (AWS) e i sistemi GIS per raccogliere dati, memorie, indice e visualizzazione di dati attraverso dispositivi informatici (Figura 19).

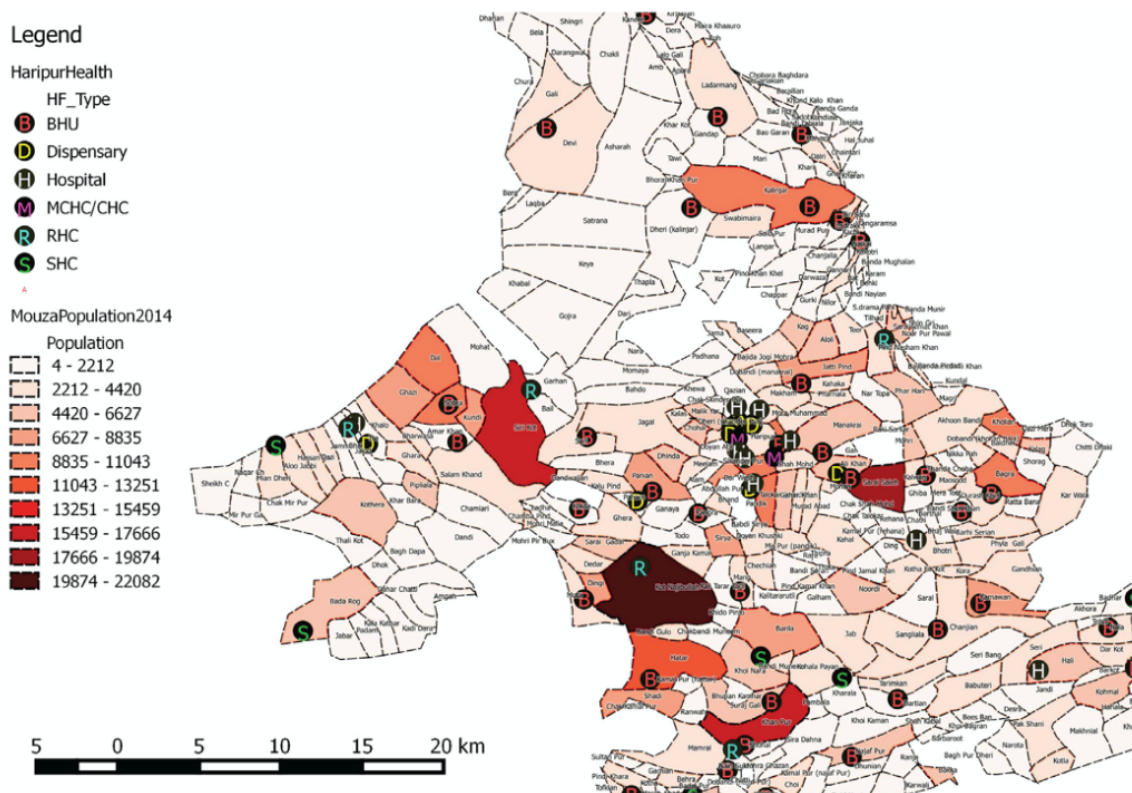


Figura 19. Mapa tematica della presenza di varie strutture sanitarie.

Anche Galletta *et al.* [33] hanno implementato un tool grafico per monitorare i dati sanitari dei pazienti da remoto. Le misurazioni provengono da diverse fonti, *i.e.*, sensori indossabili, dispositivi IoT per l'analisi del sangue e strumenti medici. Per rappresentare questi dati, Galletta *et al.* [33] usano una codifica visuale RGB, dove il parametro misurato assumerà tre valori: *hypo*, *optimal* e *hyper*. Il valore *Hypo* è utilizzato quando il parametro assume un valore troppo basso e ad esso è

associato al colore blu; *optimal* quando il parametro assume un valore pari a 0 ed ad esso è associato al colore verde; *hyper* quando il parametro assume un valore troppo alto ed è associato al colore rosso. Dato che non è facile avere la rappresentazione con la codifica RGB, gli autori hanno deciso di rappresentare lo stato di salute di ogni paziente attraverso cerchi multipli concentrici, come raffigurato in Figura 20.

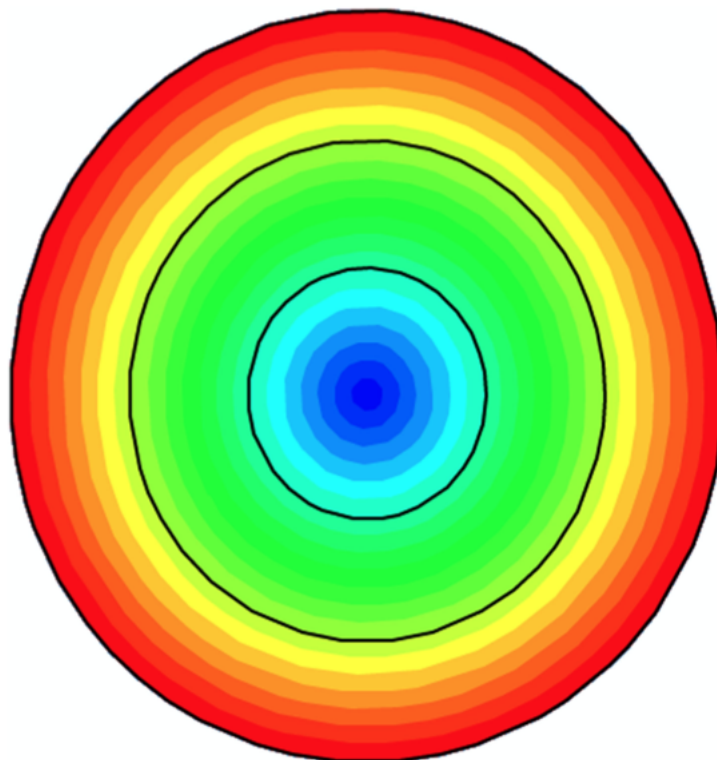


Figura 20. Scheletro della visualizzazione specifica per paziente [33].

Lo sviluppo di sistemi interattivi che combinano le tecniche di analisi ed estrazione della conoscenza dai dati e le tecniche di visualizzazione degli stessi possono risultare di grande efficacia quando si lavora con dati biomedici. Infatti, Holzinger [36] spiega come la combinazione di queste tecniche può supportare l'utente finale in maniera interattiva nell'analisi e visualizzazioni di grandi quantità di dati altamente complessi. In questo modo vengono evidenziate le parti più rilevanti necessarie all'utente e quindi avere un processo decisionale più efficiente.

Ad esempio, Jovanov *et al.* [37] hanno introdotto un nuovo modo di presentare i dati biomedici. Nello specifico, questo riguarda il monitoraggio dell'attività mentale attraverso delle visualizzazioni

3D, usando l'elettroencefalogramma (EEG) e il magnetoencefalogramma (MEG). Questo tipo di visualizzazione è basato su mappe topografiche con la testa umana in forma 3D. Per questo motivo è stato creato TEMPO per avere una sezione sperimentale da esplorare (Figura 21).

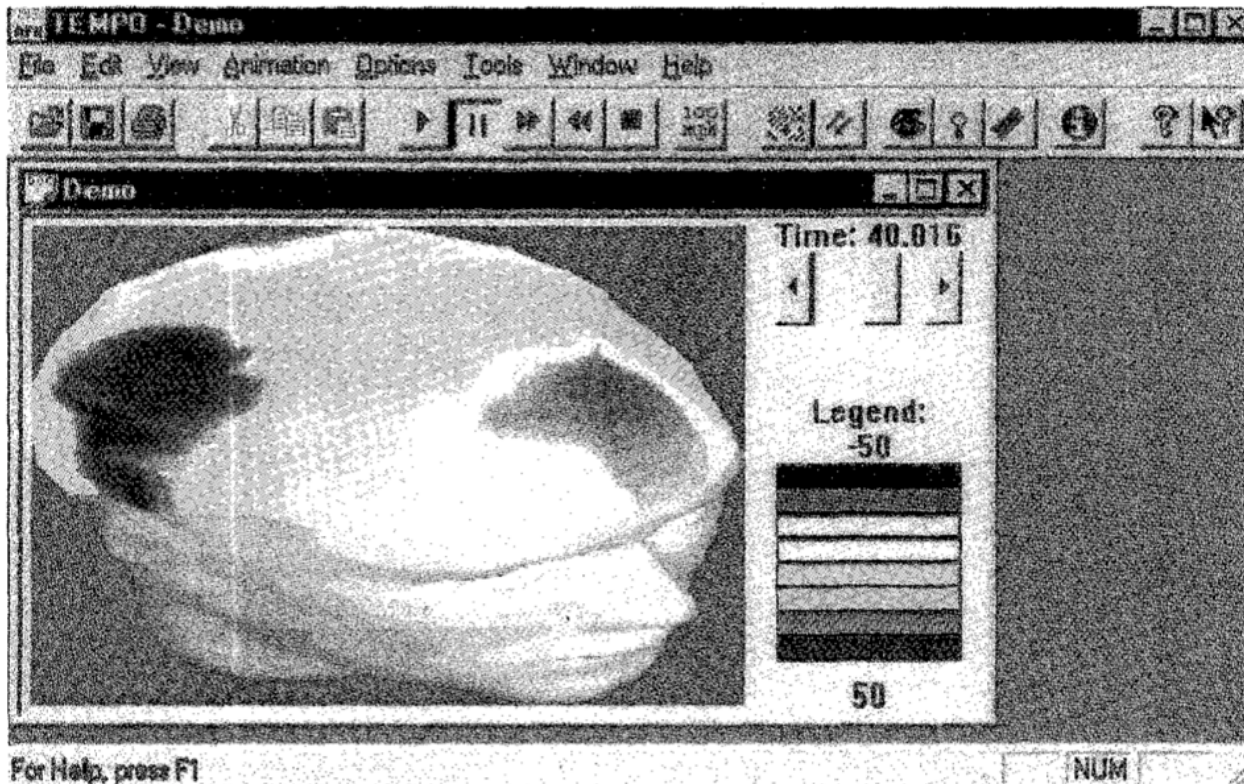


Figura 21. TEMPO.

Zhang *et al.* [39] hanno creato un nuovo framework di visualizzazione space-conscious, DBMap che consente di visualizzare i dati su più dimensioni in modo efficiente tenendo in considerazione lo spazio. Questo framework è costruito sul concetto di TreeMap, che è una rappresentazione grafica vincolata dallo spazio, ossia una matrice di rettangoli, dove ogni rettangolo è utilizzato per rappresentare le parti relative del database attraverso le dimensioni e il colore. Il DBMap permette di studiare e classificare la conoscenza nascosta integrata nei dati delle immagini mediche.

Kerfoot *et al.* [38] hanno creato la piattaforma Eidolon, che riesce ad adattare i workflow di ricerca integrando la visualizzazione con immagini e le misure dell'analisi quantitativa. In Figura 22 è

mostrata l'interfaccia grafica di Eidolon che permette di processare e visualizzare i dati, dando, inoltre, la possibilità di utilizzare i dati per effettuare degli esperimenti.

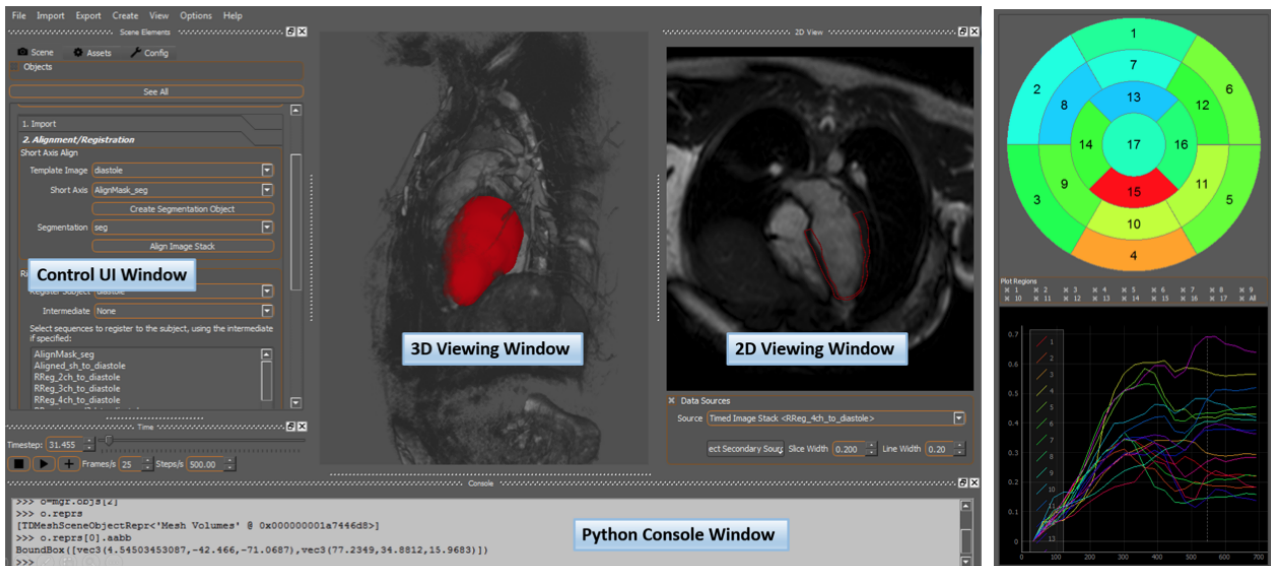


Figura 22. Interfaccia utente di Eidolon.

Con la creazione delle cartelle cliniche elettroniche, è nato il concetto di Personally Controlled Health Records (PCHR) che è una speciale cartella clinica personale dove gli utenti possono controllare i contenuti [11, 12, 13]. Adams *et al.* [10] hanno creato un'applicazione Android, chiamata MeD-Lights che si basa sulla metafora dei colori del semaforo (rosso, giallo e verde). Questo modello è utilizzato per rappresentare i livelli di sensibilità dei record e per la condivisione dei dati con il personale medico. Lo schema dei colori del semaforo serve a capire il livello di protezione di una determinata informazione della cartella clinica elettronica. Quindi, il colore rappresenta il modo di condivisione del dato di una cartella clinica elettronica. Precisamente, il verde indica che le informazioni vengono condivise a tutti, il giallo indica che solo alcune informazioni vengono condivise e il rosso indica che le informazioni non sono condivise. In Figura 23 è possibile vedere una schermata dell'applicazione.

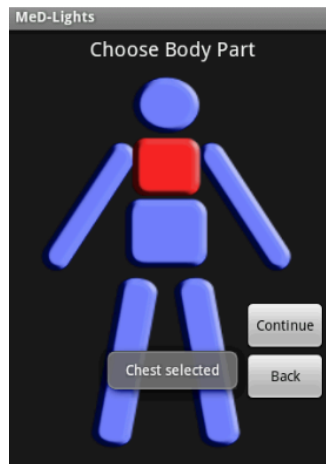


Figura 23. Schermata dell'applicazione MeD-Lights.

Adams *et al.* [10] si sono basati sul lavoro di Kapadia *et al.* [14] e hanno creato un framework con delle policy, basata sulla metafora dei muri fisici con tre livelli di trasparenza (trasparente, traslucida e opaca), dove gli utenti possono controllare in modo astratto l'accesso alle loro impronte digitali. In questo modo, gli utenti possono comprendere come proteggere le proprie impronte digitali, come farebbero nel mondo fisico. In Figura 24 è possibile vedere l'esempio che viene considerato in questo studio, ossia viene simulata la presenza di un utente in un meeting room con tutte le implicazioni che ci potrebbero essere, come i sensori audio, la temperatura della stanza e i sensori video.

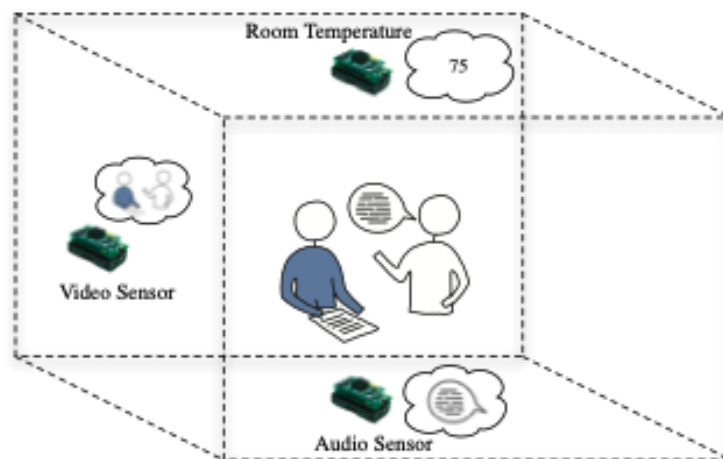


Figura 24. Esempio di stanza fisica

## 3 Dashboard per l'analisi dei dati

In base all'utente loggato, che sia esso un "Amministratore", "Caregiver", "Medico Specialista" o "Responsabile di Monitoraggio" il sistema mostrerà una particolare "dashboard" e/o "home", attraverso la quale l'utente loggato potrà effettuare delle analisi di suo interesse.

### 3.1 Amministratore

Nel caso in cui l'utente loggato è "Amministratore", e in quanto tale abilitato solo alle funzionalità di "Gestione kit di monitoraggio" e "Gestione Utenti", il sistema non mostrerà la sezione "Dashboard" e mostrerà la sola sezione "Home" composta da due grafici "Utenti" e "Kit di Monitoraggio" riguardanti rispettivamente gli utenti e i kit di monitoraggio presenti in ATTICUS (Figura 25).



Figura 25. Home - Amministratore di sistema

Attraverso il grafico "Utenti", l'amministratore potrà ottenere informazioni circa i ruoli degli utenti registrati, in particolare potrà sapere quanti utenti nel sistema hanno ruolo "Amministratore" (incluso sé stesso), quanti utenti hanno ruolo "Caregiver", quanti utenti hanno ruolo "Infermieri", quanti "Medici Specialisti" e quanti "Responsabili di Monitoraggio".

Dal grafico “Kit di Monitoraggio” l’amministratore loggato potrà recuperare le informazioni circa la distribuzione dei kit di monitoraggio presenti nel sistema. In particolare, potrà verificare quanti sono i kit “Assegnati”, quanti sono i kit “Dismessi” e quanti sono i kit “Disponibili”.

## 3.2 Caregiver

Nel caso in cui l’utente loggato ha il ruolo di “Caregiver” il sistema non mostrerà la sezione “Dashboard” e mostrerà la sezione “Home” suddivisa in tre parti. Il “Caregiver” loggato potrà visualizzare, attraverso la sezione “Home”, le informazioni riguardanti gli allarmi degli assistiti per il quale è stato associato come “Caregiver” (Figura 26).

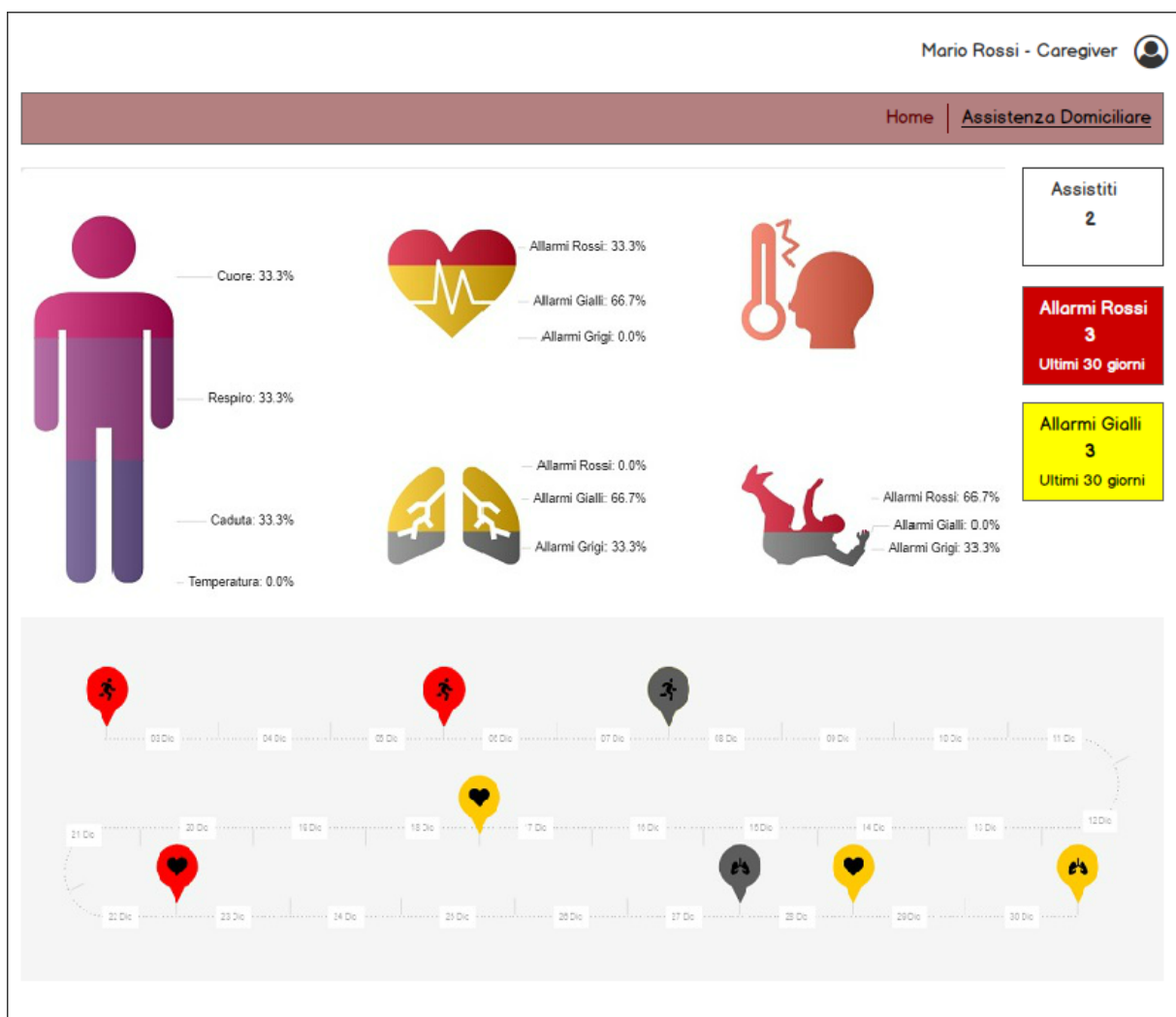


Figura 26. Home - Caregiver

In particolare, nella “Home” mostrata al “Caregiver”, in alto a destra verranno mostrati tre riquadri informativi contenenti:

- il numero di assistiti associati all’utente loggato (riquadro bianco e scritta nera);
- il numero di allarmi rossi degli assistiti associati all’utente loggato negli ultimi 30 giorni (riquadro rosso e scritta bianca);
- il numero di allarmi gialli degli assistiti associati all’utente loggato negli ultimi 30 giorni (riquadro giallo e scritta nera).

Nella sezione centrale verranno proposte 5 figure:

- la sagoma di un uomo, attraverso la quale il “Caregiver” potrà visualizzare la percentuale di allarmi riscontrati dal sistema in termini di patologie (Caduta, Cuore, Respiro e Temperatura) degli assistiti per i quali è stato associato come Caregiver;
- l’immagine di un cuore, attraverso la quale il “Caregiver” potrà visualizzare la percentuale di allarmi rossi, gialli e grigi di tipo “Cuore” degli assistiti a lui assegnati;
- l’immagine di un polmone, attraverso la quale il “Caregiver” potrà visualizzare la percentuale di allarmi con gravità rossi, gialli e grigi di tipo “Respiro” degli assistiti a lui assegnati;
- L’immagine di un uomo che cade, attraverso la quale il “Caregiver” potrà visualizzare la percentuale di allarmi con gravità rossi, gialli e grigi di tipo “Caduta” degli assistiti a lui assegnati;
- L’immagine di un termometro, attraverso la quale il “Caregiver” potrà visualizzare la percentuale di allarmi con gravità rossi, gialli e grigi di tipo “Temperatura” degli assistiti a lui assegnati.

Infine, il “Caregiver” potrà visualizzare la data, la tipologia e la gravità degli allarmi riscontrati negli ultimi 30 giorni attraverso un unico grafico riportato in basso alla sezione “Home”. Nel caso in cui, negli ultimi 30 giorni, per uno o più assistiti assegnati al “Caregiver” loggato è stato sollevato un allarme, il sistema mostrerà sul grafico uno specifico segnaposto (Figura 27).

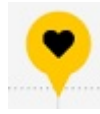


Figura 27.Home - Caregiver - Segnaposto

Ogni singolo segnaposto del grafico indicherà la gravità dell'allarme attraverso un colore. In particolare, assumerà:

- colore rosso per gli allarmi di tipo rosso considerati più gravi;
- colore giallo per gli allarmi di tipo giallo considerati di gravità media;
- colore grigio per gli allarmi di tipo grigio dovuti ad un problema tecnico.

Inoltre, attraverso il segnaposto, il "Caregiver" potrà risalire all'informazione della patologia di allarme sollevata dal sistema. In particolare, il segnaposto, oltre al colore, sarà caratterizzato da un'icona diversa in base alla patologia di allarme riscontrata dal sistema:

- (cuore): allarme di tipo Cardiaco;
- (polmone): allarme di tipo Respiratorio;
- (uomo): allarme di tipo Caduta.

Attraverso il segnaposto, oltre a risalire alla tipologia dell'allarme individuata dal colore e alla patologia dell'allarme individuata dell'icona, il "Caregiver" potrà risalire alla data in cui si è verificato l'allarme, in quanto, il segnaposto verrà posizionato sulla linea temporale degli allarmi degli ultimi 30 giorni in un preciso posto riportandone la data.

### 3.3 Medico Specialista e Responsabile di Monitoraggio

Nel caso in cui l'utente loggato ha il ruolo di "Medico Specialista" o "Responsabile di Monitoraggio", il sistema mostrerà sia la voce di menu "Home" sia la voce di menu "Dashboard", quest'ultima a sua volta divisa in due sotto-voci: "Allarmi: i miei assistiti" e "Allarmi: tutti gli assistiti" (Figura 28). La voce "Home" riporterà le stesse informazioni mostrate in "Dashboard – Allarmi – i miei assistiti"



Figura 28. Dashboard – Menù - Medico Specialista – Responsabile di Monitoraggio

### 3.3.1 Allarmi: I miei assistiti

Selezionando la sottovoce “Allarmi - i miei assistiti”, il “Medico Specialista” e/o il “Responsabile di Monitoraggio” potrà ottenere informazioni circa gli allarmi riscontrati dai suoi assistiti. In particolare, il sistema mostrerà tre sezioni (Figura 29).

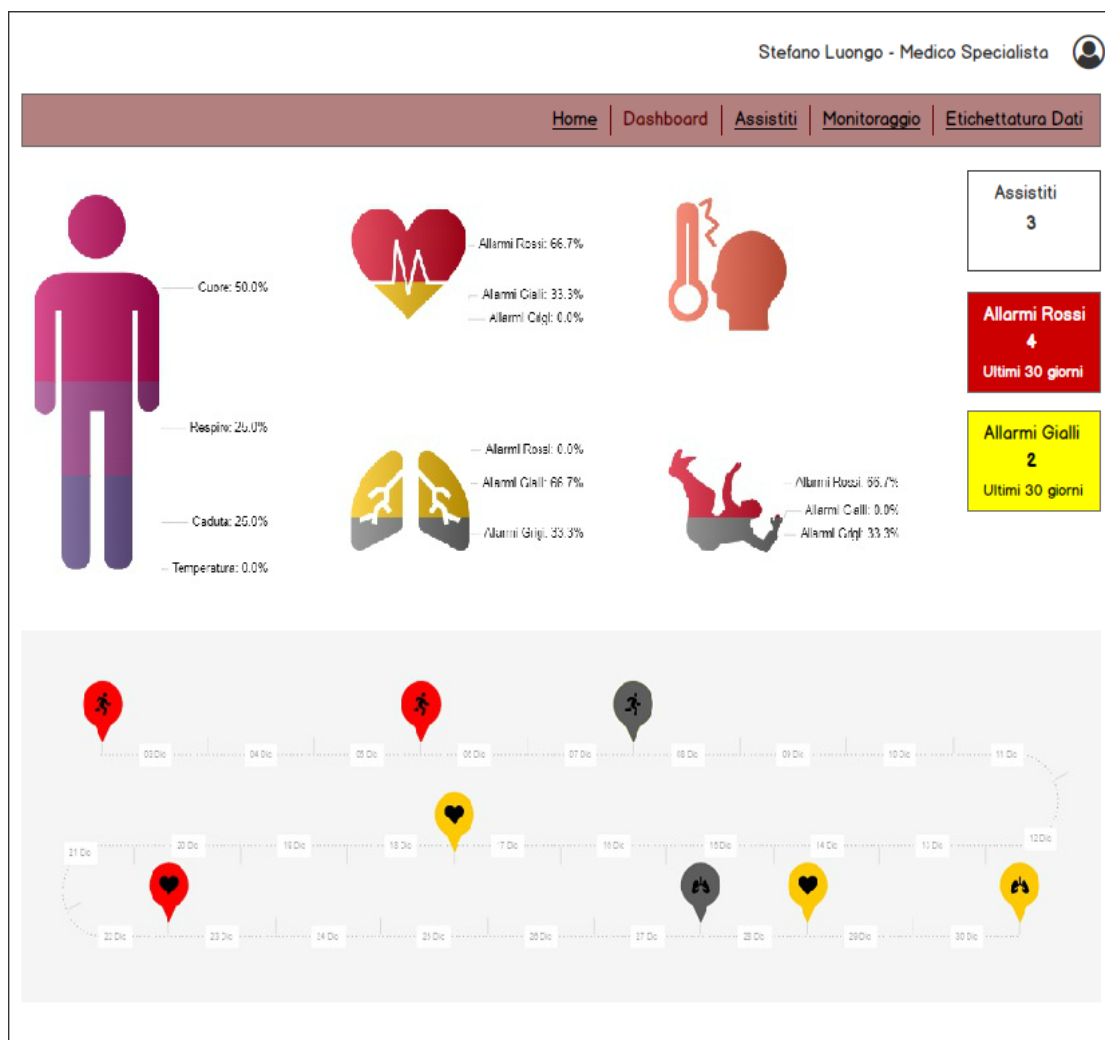


Figura 29. Dashboard - i miei assistiti - Medico Specialista – Responsabile di Monitoraggio

Nella prima sezione a sinistra saranno riportate 5 figure:

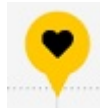
- la sagoma di un uomo, attraverso la quale il “Medico Specialista” o “Responsabile di Monitoraggio” loggato potrà visualizzare la percentuale di allarmi riscontrati dal sistema in termini di patologie (Caduta, Cuore, Respiro e Temperatura) dei soli assistiti per i quali è stato associato come “Medico Specialista” o come “Responsabile di Monitoraggio”;
- l’immagine di un cuore, attraverso la quale il “Medico Specialista” o “Responsabile di Monitoraggio” loggato potrà visualizzare la percentuale di allarmi rossi, gialli e grigi di tipo “Cuore” degli assistiti a lui assegnati;
- l’immagine di un polmone, attraverso la quale il “Medico Specialista” o “Responsabile di Monitoraggio” loggato potrà visualizzare la percentuale di allarmi di tipologia rossi, gialli e grigi di tipo “Respiro” degli assistiti a lui assegnati;
- l’immagine di un uomo che cade, attraverso la quale il “Medico Specialista” o “Responsabile di Monitoraggio” loggato potrà visualizzare la percentuale di allarmi di tipologia rossi, gialli e grigi di tipo “Caduta” degli assistiti a lui assegnati;
- l’immagine di un termometro, attraverso la quale il “Medico Specialista” o “Responsabile di Monitoraggio” loggato potrà visualizzare la percentuale di tipologia rossi, gialli e grigi di tipo “Temperatura” degli assistiti a lui assegnati.

In alto a destra, verranno mostrati tre riquadri informativi contenenti:

- il numero di assistiti associati all’utente loggato (riquadro bianco e scritta nera);
- il numero di allarmi rossi degli assistiti associati all’utente loggato negli ultimi 30 giorni (riquadro rosso e scritta bianca) riscontrati dal sistema;
- il numero di allarmi gialli degli assistiti associati all’utente loggato negli ultimi 30 giorni (riquadro giallo e scritta nera) riscontrati dal sistema.

Infine, il “Medico Specialista” o “Responsabile di Monitoraggio” potrà visualizzare la data, la tipologia e la gravità degli allarmi riscontrati negli ultimi 30 giorni attraverso un unico grafico

riportato in basso. Nel caso in cui, negli ultimi 30 giorni, per uno o più assistiti assegnati al “Medico Specialista” o al “Responsabile di Monitoraggio” loggato, è stato sollevato un allarme, il sistema mostrerà uno specifico segnaposto sul grafico temporale che riporterà tipologia e patologia nella data in cui si è verificato l’allarme (Figura 30).



*Figura 30. Dashboard – Medico Specialista – Responsabile di Monitoraggio – Segnaposto – i miei assistiti*

Ogni singolo segnaposto del grafico indicherà la gravità dell’allarme attraverso un colore. In particolare, assumerà:

- colore rosso per gli allarmi di tipo rosso considerati più gravi;
- colore giallo per gli allarmi di tipo giallo considerati di gravità media;
- colore grigio per gli allarmi di tipo grigio dovuti ad un problema tecnico.

Inoltre, attraverso il segnaposto, il “Medico Specialista” o “Responsabile di Monitoraggio” potrà risalire all’informazione della patologia di allarme sollevata dal sistema. In particolare, il segnaposto, oltre al colore, sarà caratterizzato da una particolare icona in base alla patologia di allarme riscontrata dal sistema:

- (cuore): allarme di tipo Cardiaco;
- (polmone): allarme di tipo Respiratorio;
- (uomo): allarme di tipo Caduta.

Attraverso il segnaposto, oltre a risalire alla gravità dell’allarme individuata dal colore dello stesso e alla patologia dell’allarme individuata dell’icona, il “Medico Specialista” o “Responsabile di Monitoraggio” potrà risalire alla data in cui si è verificato l’allarme, in quanto, il segnaposto verrà posizionato sulla linea temporale degli allarmi degli ultimi 30 giorni riportandone la data.

### 3.3.2 Allarmi: tutti gli assistiti

Selezionando la funzionalità “Allarmi – tutti i miei assistiti”, il “Medico Specialista” o il “Responsabile di Monitoraggio” potrà ottenere informazioni circa gli allarmi riscontati da tutti gli assistiti presenti nel sistema. In particolare, il sistema mostrerà oltre alle tre sezioni descritte nel paragrafo 3.3.1, due nuove sezioni (Figura 31). Le prime tre sezioni saranno uguali alle tre sezioni descritte nel paragrafo 3.3.1 con l’unica differenza che il sistema recupererà le informazioni di tutti gli assistiti presenti, non solo di quelli associati allo specifico utente loggato.

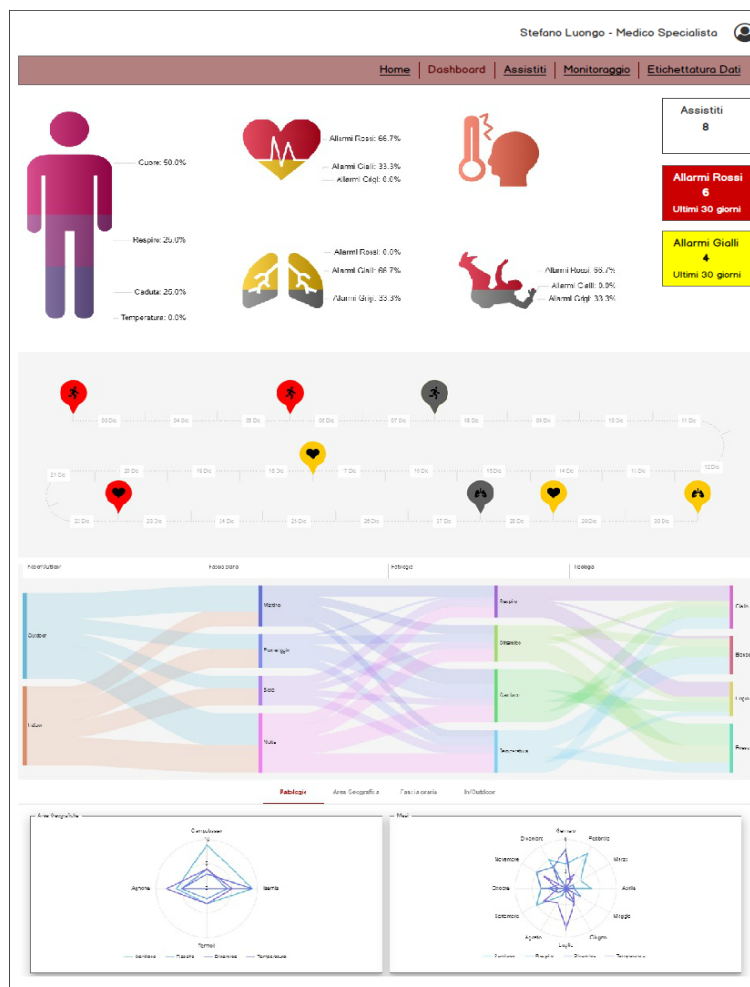


Figura 31. Dashboard - Tutti gli assistiti - Medico Specialista/Responsabile di Monitoraggio

La prima nuova sezione riporterà un grafico a flusso; la prima riga in alto riporterà le categorie: “Patologia”, “Tipologia”, “Fascia Oraria” e “Indoor/Outdoor” (Figura 32).

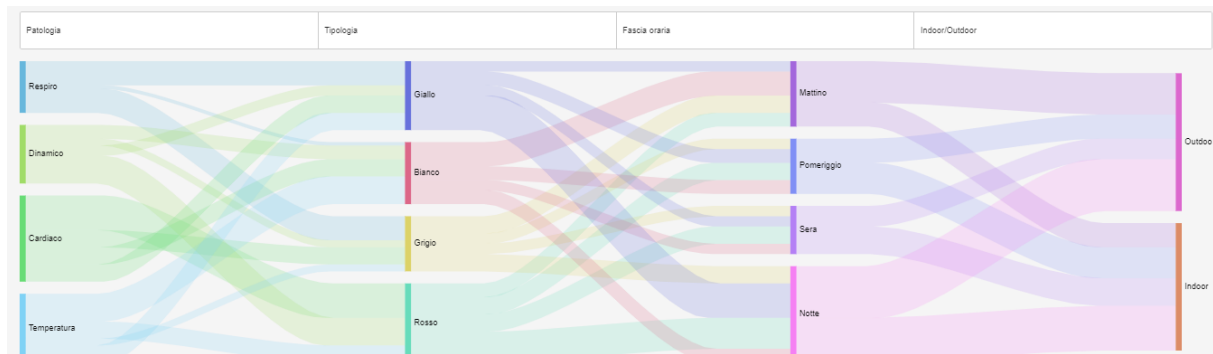


Figura 32. Dashboard - Tutti gli assistiti - Medico Specialista - Responsabile di Monitoraggio

Per ciascuna categoria il grafico riporterà le informazioni di riferimento. Per la categoria “Patologie” verranno riportate le 4 patologie trattate dal sistema: “Cardiaco”, “Dinamico”, “Respiro” e “Temperatura”, ciascuna di esse rappresentante su una singola riga della colonna “Patologie” (Figura 33).

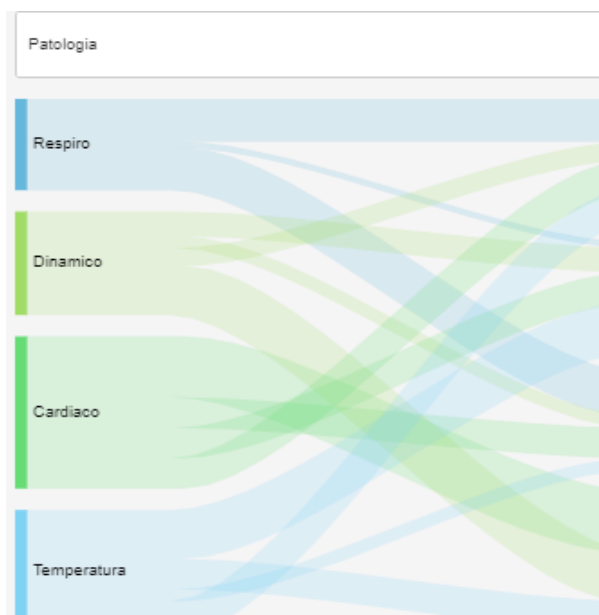


Figura 33. Dashboard - Medico Specialista - Responsabile di Monitoraggio

Per la colonna “Tipologia” verranno riportate le 4 tipologie trattate dal sistema: “Bianco”, “Giallo”, “Grigio” e “Rosso”, ognuno su una singola riga della categoria “Tipologia”.

Per la colonna “Fascia Oraria” verranno riportate le 4 fasce orarie trattate dal sistema: “Mattina”, “Pomeriggio”, “Sera” e “Notte”, ognuno su una singola riga della colonna “Fascia Oraria”.

Per la colonna “Indoor/Outdoor” verranno riportate le 2 possibilità trattate dal sistema: “Indoor” e “Outdoor”, ognuno su una singola riga della colonna “Indoor/Outdoor”.

Dal grafico il “Medico Specialista” o il “Responsabile di Monitoraggio” potrà visualizzare le informazioni degli allarmi attraverso un flusso. Le informazioni di ogni singolo elemento di ciascuna categoria verranno smistate nella categoria successiva a cascata in base alle caratteristiche di ciascun allarme. Ad esempio nella Figura 32 le informazioni dell’elemento “Respiro” della categoria “Patologie” verranno smistate negli elementi “Rosso”, “Giallo” e “Bianco” della categoria “Tipologia”, l’elemento “Rosso” della categoria “Tipologia” verrà smistato negli elementi “Mattina”, “Pomeriggio”, “Sera” e “Notte” della categoria “Fascia Oraria”, l’elemento “Mattina” della Categoria “Fascia oraria” verrà smistato negli elementi “Indoor” e “Outdoor” della categoria “Indoor/Outdoor”.

Il “Medico Specialista” o “Responsabile di Monitoraggio” loggato può invertire l’ordine delle categorie(colonne) e degli elementi di una categoria(righe) a suo piacimento, selezionando la categoria o l’elemento che desidera spostare. Non è possibile spostare elementi appartenenti a categorie diverse (es. non è possibile spostare l’elemento “Rosso” della categoria “Tipologia” nella categoria “Patologia”).

L’ultima sezione della sottovoce “Allarmi: tutti gli assistiti” è suddivisa in quattro tab “Patologie”, “Aree Geografiche”, “Fascia Oraria” e “In/outdoor”.

In fase iniziale, il sistema mostrerà le informazioni del tab “Patologie”. Il tab “Patologie” mostrerà quattro grafici: “Aree Geografiche”, “Mesi”, “Patologie” e “Confronto”, ognuno dei quali riporterà le informazioni di tutti gli allarmi sollevati dal sistema. I primi due grafici utilizzano i diagrammi di Kiviat o comunemente chiamati diagrammi a ragnatela. Nei diagrammi a ragnatela una prima variabile viene individuata da raggi di ampiezza uguale partendo dal centro, mentre una seconda variabile viene rappresentata da un punto individuato su ciascun raggio che indica la distanza

proporzionale al valore della variabile rispetto al valore massimo raggiungibile. I punti sui raggi vengono congiunti con segmenti, così che il grafico ha la forma di una stella o di una ragnatela.

Il primo grafico della sottosezione “Patologie” che utilizza i diagrammi a ragnatela è “Aree Geografiche”. Tale grafico viene suddiviso in quattro raggi, ognuno dei quali rappresenta le macroaree della regione Molise: “Agnone”, “Campobasso”, “Isernia” e “Termoli” aventi tutti la stessa ampiezza. Su ciascun raggio, indicante la macroarea, vengono individuati dei punti di riferimento che rappresentano il numero di allarme riscontrati in quell’area geografica per patologie in proporzione al numero massimo di allarmi. Una volta individuati i punti sui raggi per ciascuna patologia verranno uniti i punti appartenenti alla stessa patologia con segmenti che formeranno una ragnatela (Figura 34). I segmenti avranno un colore diverso in base alla patologia, in modo da distinguere le patologie tra di loro. L’utente loggato può scegliere di eliminare o aggiungere una patologia dal grafico selezionando o deselegionando la patologia dalla legenda riportata al di sotto del grafico.

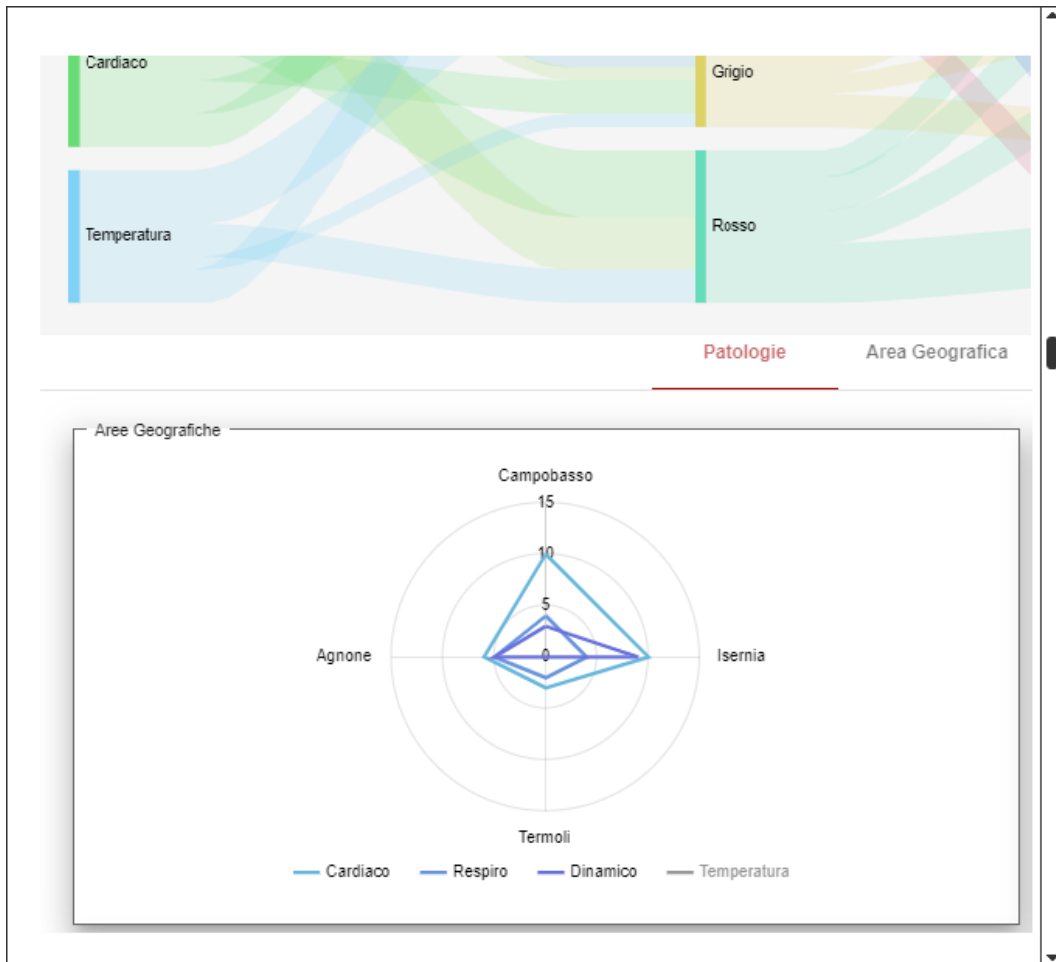


Figura 34. Dashboard - Tutti gli assistiti - Patologie - Aree Geografiche - Medico Specialista - Responsabile di Monitoraggio

Il secondo grafico della sottosezione “Patologie” che utilizza i diagrammi a ragnatela è “Mesi”. Tale grafico viene suddiviso in dodici raggi, ognuno dei quali rappresenta i dodici mesi dell’anno aventi tutti la stessa ampiezza. Su ciascun raggio, indicante il mese dell’anno, vengono individuati dei punti di riferimento che rappresentano il numero di allarme riscontrati in quel mese per patologie in proporzione al numero massimo di allarmi. Una volta individuati i punti sui raggi per ciascuna patologia essi verranno uniti con segmenti e formeranno una ragnatela (Figura 35). L’utente loggato

può scegliere di eliminare o aggiungere una patologia dal grafico selezionando o deselegionando la patologia dalla legenda riportata al di sotto del grafico.

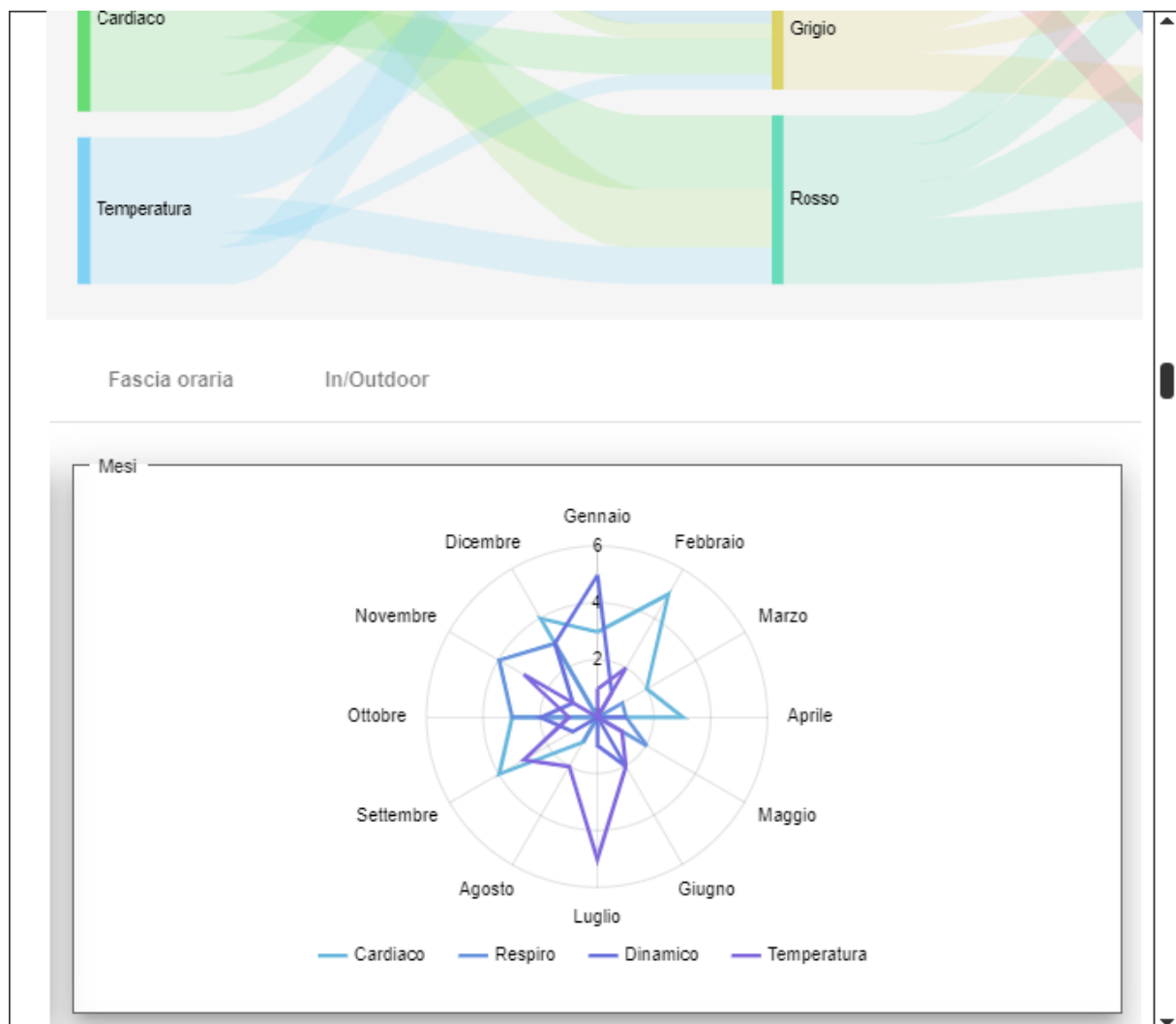


Figura 35. Dashboard - Tutti gli allarmi - Patologie - Mesi - Medico Specialista - Responsabile di Monitoraggio

Il terzo grafico “Patologie” visualizzerà attraverso un istogramma le patologie riscontrate in un determinato anno. Sull’asse dell’ascisse verranno riportati i dodici mesi dell’anno, e ogni mese dell’anno conterà un massimo di quattro rettangoli, ognuno dei quali conterrà gli allarmi di ogni singola patologia: “Cardiaco”, “Respiro”, “Dinamica” e “Temperatura”. Sull’asse delle ordinate verrà indicato il numero di allarmi (Figura 36).

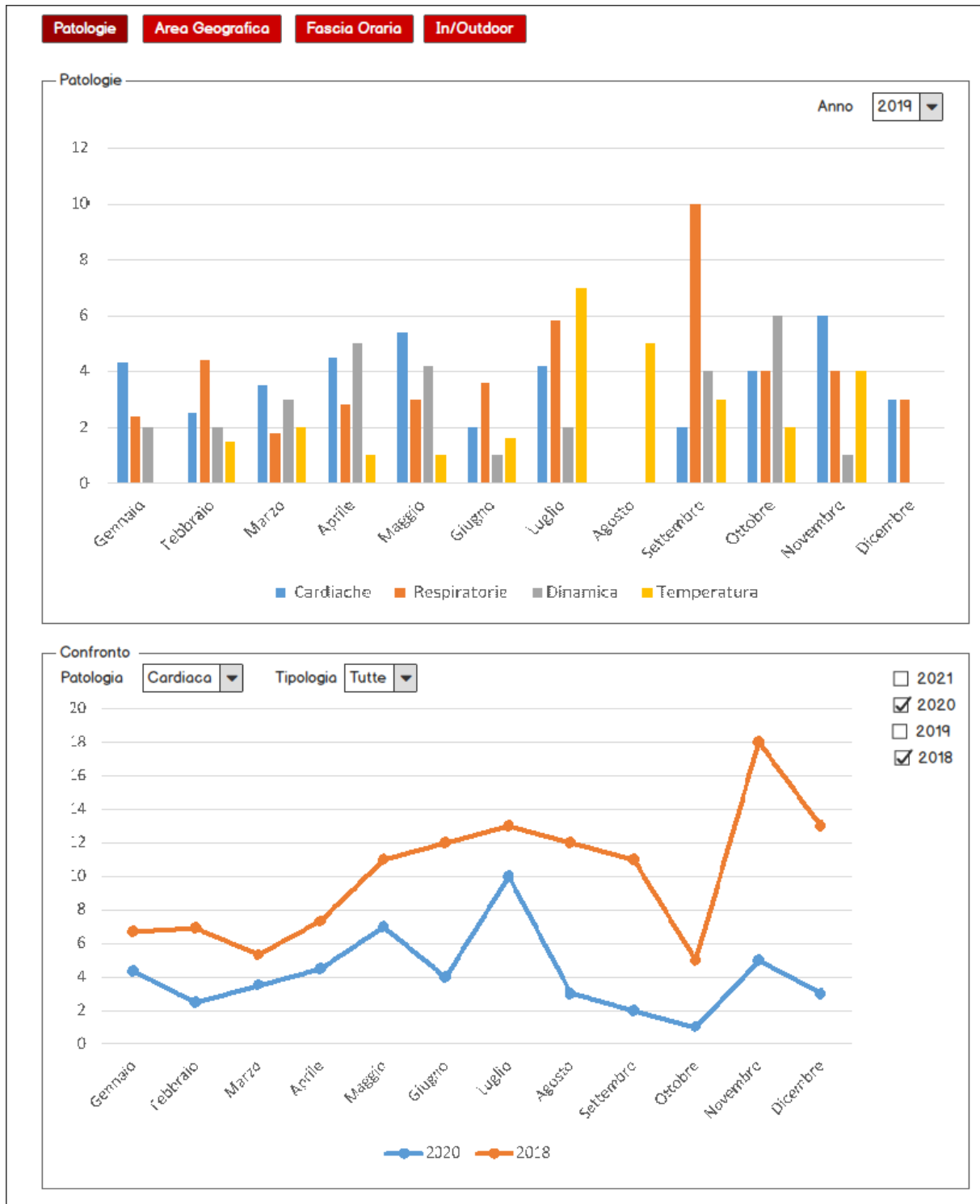


Figura 36. Dashboard - Tutti gli assistiti - Patologie - Medico Specialista - Responsabile di Monitoraggio

Nella sezione “Patologie” del tab “Patologie”, l’utente potrà scegliere l’anno per il quale desidera visualizzare l’istogramma popolando il campo “Anno”; in fase iniziale verrà mostrato l’istogramma dell’anno corrente. In questa prima sezione, il sistema mostrerà i campi riportati in Tabella 1.

Tabella 1. Dashboard – Patologie

Sezione	Campi	Descrizione	Obbligatorietà	Tipo campo
Patologie	Anno	Campo utilizzato per selezionare l’anno per il quale l’utente autenticato desidera visualizzare le informazioni. In fase iniziale è precompilato con l’anno corrente.	Si	Lista, selezionabili
Patologie	Grafico	Grafico contenente le informazioni riguardanti le patologie mostrate attraverso un istogramma in base all’anno scelto.	Si	Grafico

Se il “Medico Specialista” o il “Responsabile di Monitoraggio” loggato fosse interessato a visualizzare con maggiore dettaglio le informazioni riguardanti un particolare mese dell’anno precedentemente scelto, potrà selezionare il mese d’interesse dall’istogramma. A questo punto il sistema mostrerà un pop up composto da due grafici: un istogramma a sinistra e un grafico a torta a destra. Sull’ascisse dell’istogramma verranno riportati i giorni del mese selezionato, mentre sulle ordinate verranno riportati il numero degli allarmi confermati. Ciascun giorno dell’istogramma potrà avere un massimo di 4 rettangoli ognuno dei quali rappresenterà una patologia trattata dal sistema (Figura 37); a destra dell’istogramma verrà mostrato un grafico a torta contenente la distribuzione degli allarmi (rosso e giallo) riscontrati durante il mese indicato, anch’essi raggruppati per patologia. L’utente potrà visualizzare la distribuzione degli allarmi per tipologia di ciascuna patologia, spostandosi con le frecce situate accanto al grafico a torta. I campi mostrati in questa sezione di dettaglio sono riportati in Tabella 2.

Tabella 2. Dashboard – Patologie: dettaglio Mensile

Sezione	Campi	Descrizione	Obbligatorietà	Tipo campo
Dettaglio Patologie - Istogramma	Anno	Campo contenente l'anno per il quale si sta visualizzando il dettaglio mensile	Si	Testuale, non modificabile
Dettaglio Patologie - Istogramma	Mese	Campo contenente il mese per il quale si sta visualizzando il dettaglio mensile	Si	Testuale, non modificabile
Dettaglio Patologie - Istogramma	Grafico	Grafico contenente le informazioni attraverso un istogramma in base al mese scelto.	Si	Grafico
Dettaglio Patologie – Grafico a torta	Patologie	Campo contenente l'informazione circa la patologia per il quale si sta visualizzando la distribuzione degli allarmi riscontrati per tipologia e patologia attraverso un grafico a torta	Si	Testuale, non modificabile
Dettaglio Patologie- Grafico a torta	Grafico	Grafico contenente le informazioni attraverso un grafico a torta in base al mese scelto.	Si	Grafico

Selezionando il bottone “Esci” riportato nel pop up, l'utente loggato potrà ritornare alla pagina di provenienza.

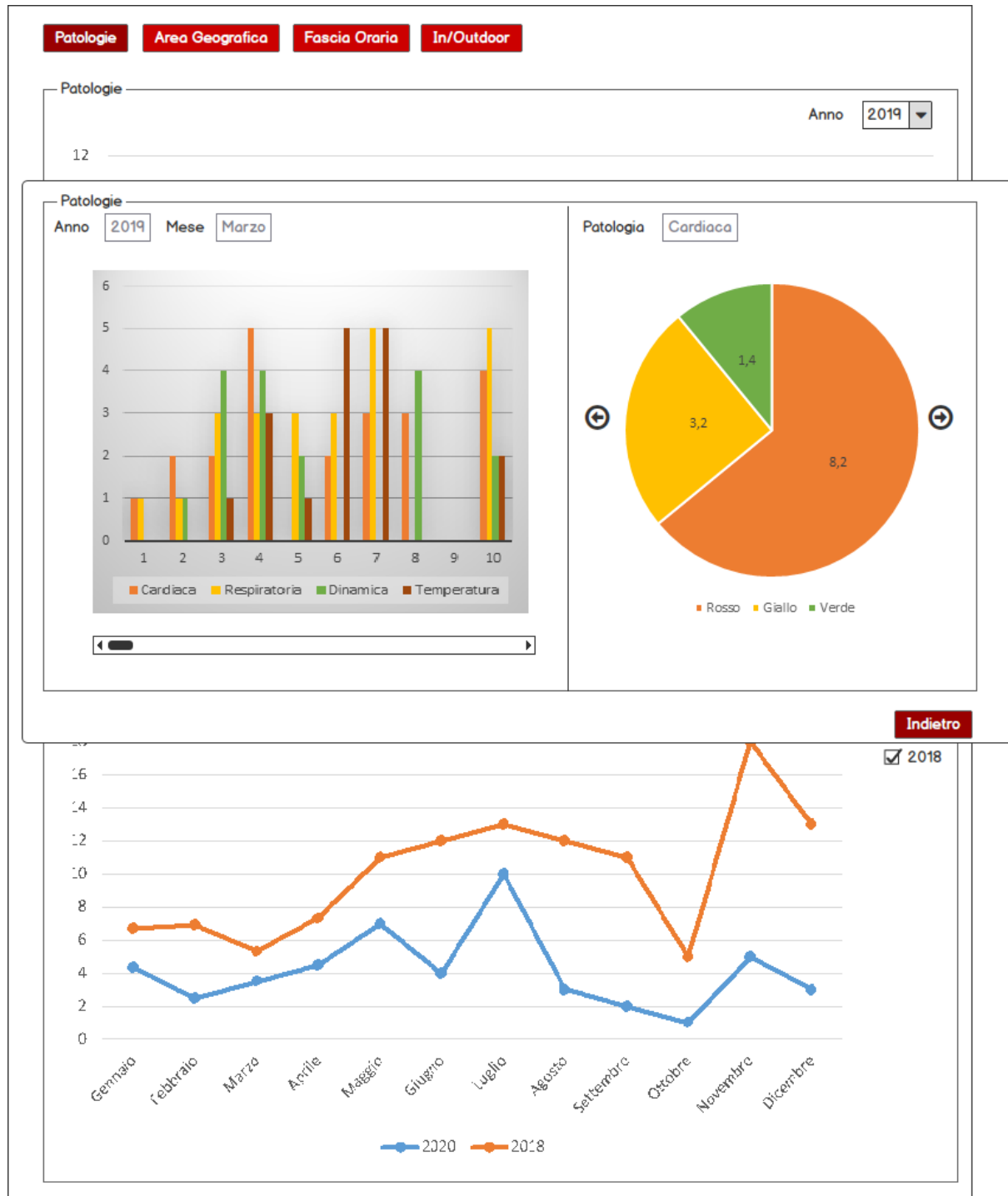


Figura 37. Dashboard – Patologie - Dettaglio Mensile

Nella sezione “Confronto” del tab “Patologie”, il “Medico Specialista” e il “Responsabile di Monitoraggio” loggato potrà confrontare le informazioni riguardanti gli allarmi confermati per una stessa patologia in anni differenti. Oltre la patologia e l’anno da confrontare, il “Medico Specialista” o “Responsabile di Monitoraggio” loggato potrà scegliere se il confronto dovrà tenere presente di tutte le tipologie di allarme oppure considerare solo una particolare tipologia, il tutto popolando il campo “Tipologia” della sezione “Confronto”. Una volta scelti gli elementi da confrontare, il sistema automaticamente mostrerà attraverso un grafico gli allarmi della patologia scelta per ciascun anno. Sull’asse dell’ascisse verranno riportati i mesi dell’anno selezionato mentre sull’asse delle ordinate il rispettivo numero di allarme se presente. Ogni anno sarà mostrato attraverso una linea di congiunzione dei punti di intersezione tra mesi e numero di allarmi e avrà un proprio colore (Figura 38).

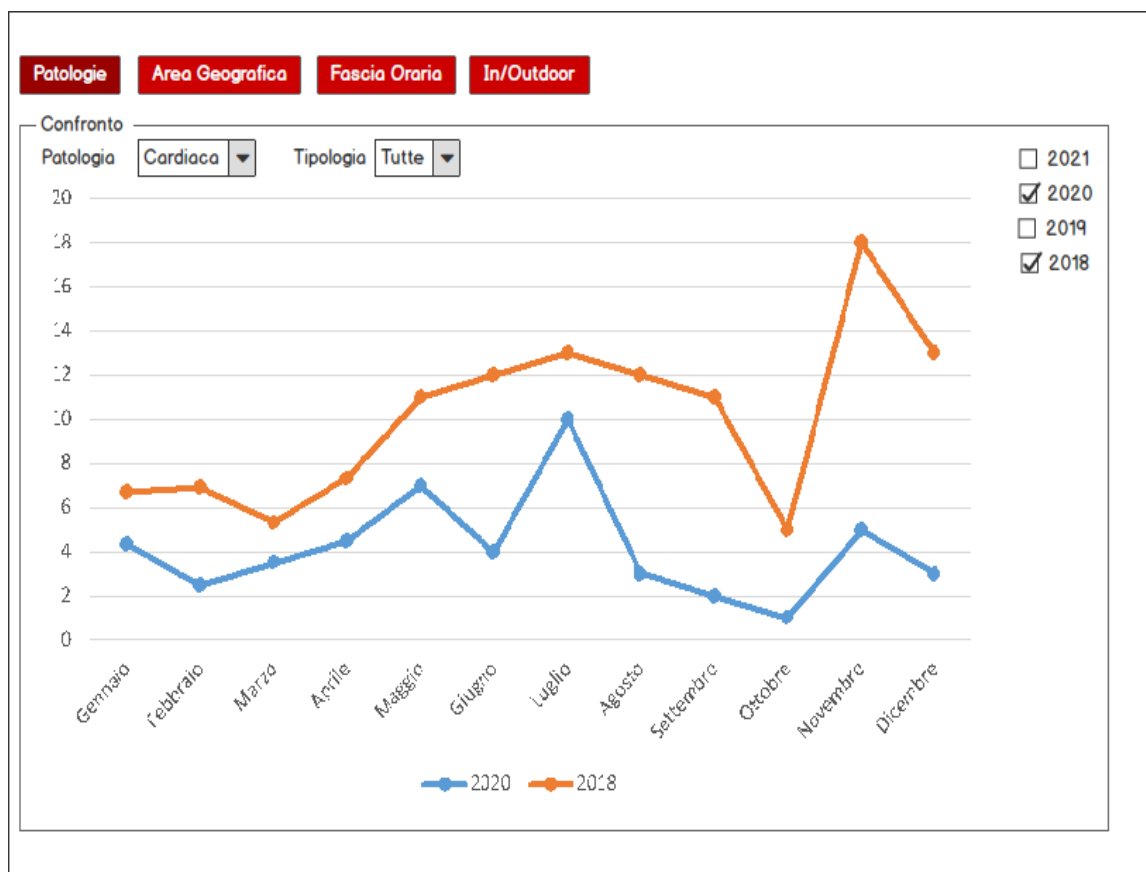


Figura 38. Dashboard - Tutti gli assistiti - Patologie - Confronto - Medico Specialista - Responsabile di Monitoraggio

Selezionando il tab “Area Geografica” il sistema mostrerà all’utente loggato tre sezioni: “HeatMap”, “Cluster” e “Confronto” (Figura 39).

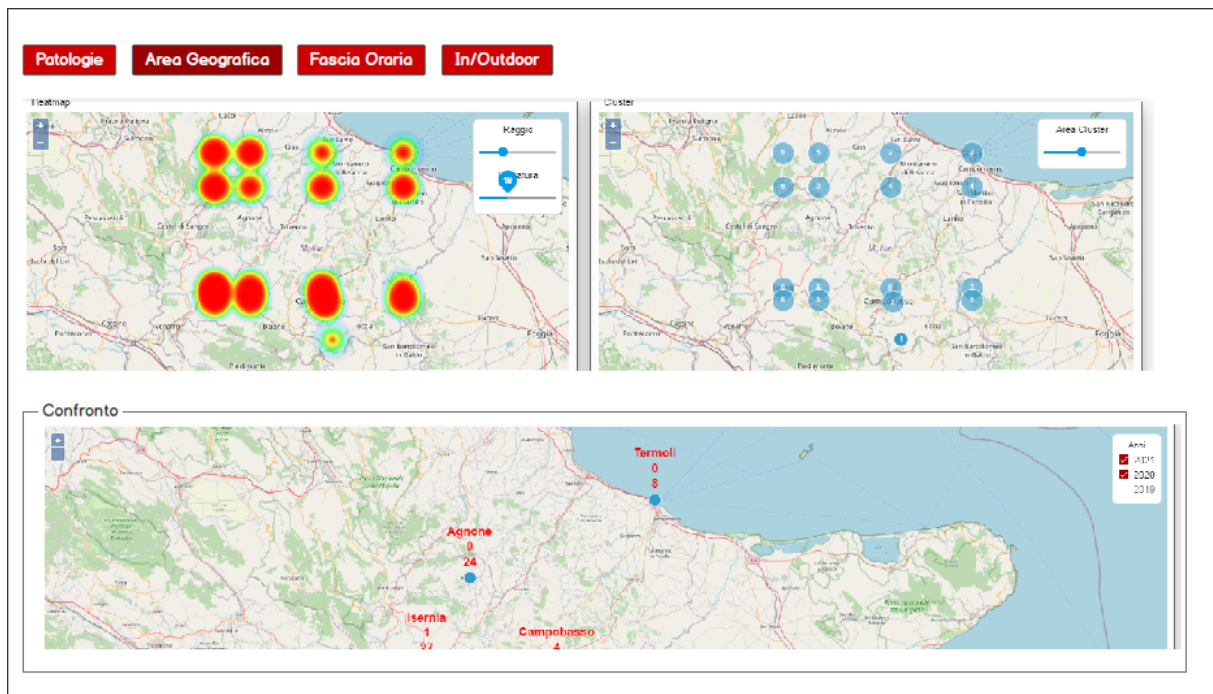


Figura 39. Dashboard – Tutti gli assistiti - Area Geografica – Medico Specialista – Responsabile di Monitoraggio

Nella prima sezione verrà mostrato un grafico “HeatMap” (mappa di calore) che permetterà attraverso una rappresentazione visuale dei dati di ottenere, a colpo d’occhio, la distribuzione degli allarmi. Nella mappa di calore utilizzata per ATTICUS, i colori rappresentano la densità dei punti, in modo da permettere di individuare più facilmente le aree con maggior numero di allarme. Inoltre, in questa sezione, sono presenti due parametri sui quali si base la visualizzazione grafica: il “Raggio” e la “Sfocatura”; entrambi i parametri possono essere aumentati o diminuiti in base al gradimento del “Medico Specialista” o “Responsabile di Monitoraggio” loggato. Muovendo il cursore del campo “Raggio” si agirà sulla grandezza dei segnaposti, più è grande il raggio maggiore è la grandezza del segnaposto, mentre muovendo il cursore del campo “sfocatura” si agirà sulla sfocatura del segnaposto, più è piccolo il valore della sfocatura richiesta minore sarà la sfocatura, più è grande il valore di sfocatura impostato dall’utente e con maggiore sfocatura apparirà il segnaposto (Figura 40).

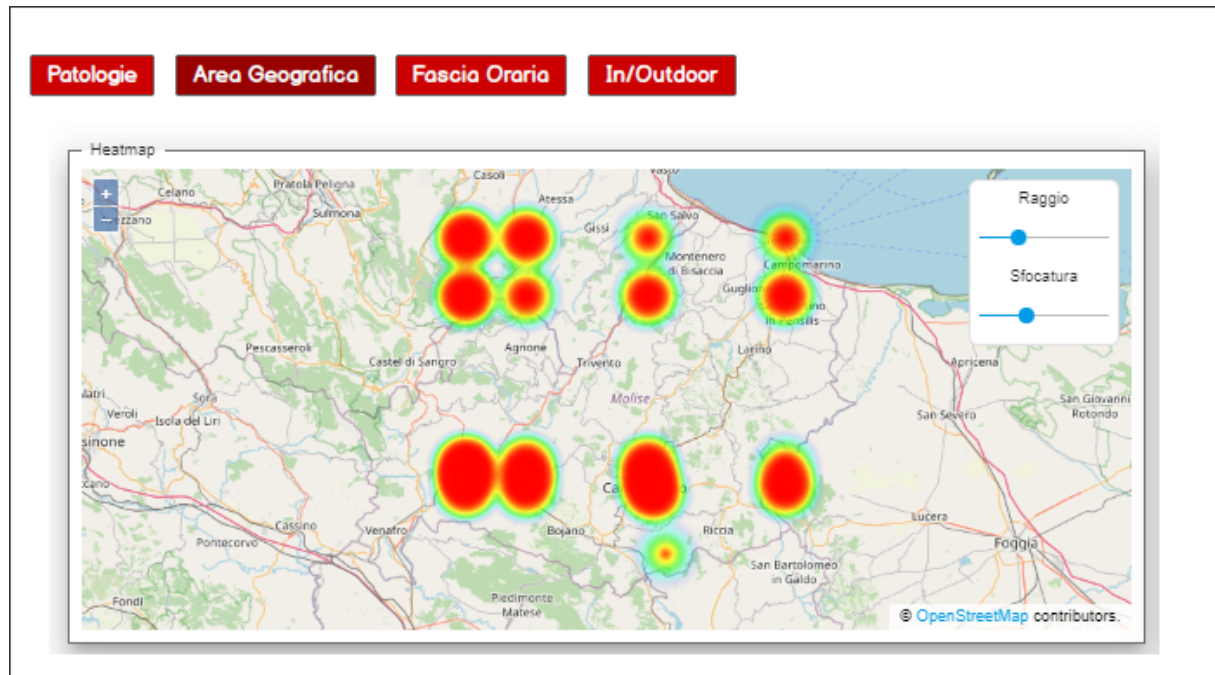


Figura 40. Dashboard - Tutti gli assistiti - Heatmap - Medico Specialista - Responsabile di Monitoraggio

Nella sezione “Cluster” gli allarmi verranno raggruppati in zone in base alla loro distanza reciproca. Tale distanza potrà essere scelta dal “Medico Specialista” o “Responsabile di Monitoraggio” loggato spostando il cursore del campo “Cluster” riportato in alto a destra della sezione “Cluster”. Maggiore sarà il valore del cursore e più grande sarà il numero degli allarmi raggruppati in una singola sfera. All’interno della sfera di raggruppamento verrà riportato il numero di allarmi individuato (Figura 41).

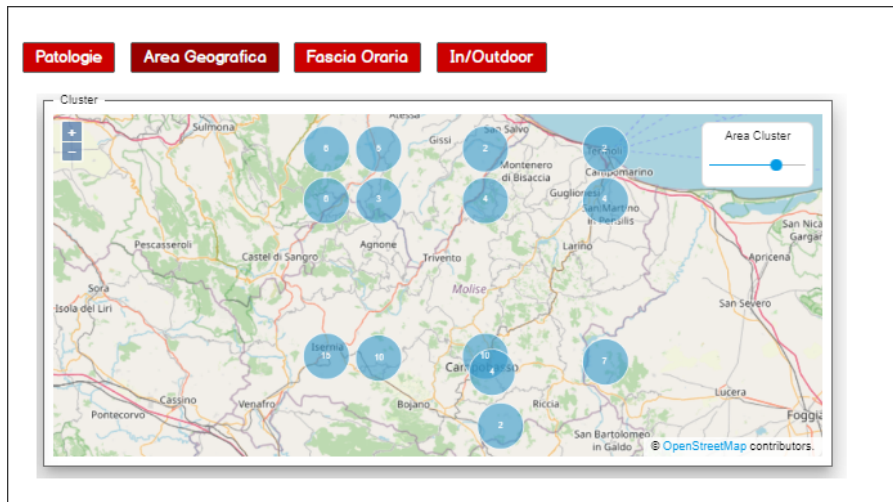


Figura 41. Dashboard - Tutti gli assistiti - Area Geografica - Cluster - Medico Specialista - Responsabile di Monitoraggio

Nella terza e ultima sezione del tab “Area Geografica”, il “Medico Specialista” o il “Responsabile di Monitoraggio” loggato potrà confrontare il numero di allarmi ottenuti nelle quattro macroaree geografiche individuate nel Molise: “Agnone”, “Isernia”, “Campobasso” e “Termoli”. Il “Medico Specialista” o il “Responsabile di Monitoraggio” loggato, per effettuare il confronto tra macroaree dovrà necessariamente scegliere due anni da confrontare riportati dal sistema nel campo “Anni”. Il confronto sarà effettuato per un massimo di due anni alla volta e in fase iniziale il sistema mostrerà le informazioni degli ultimi due anni (Figura 42).

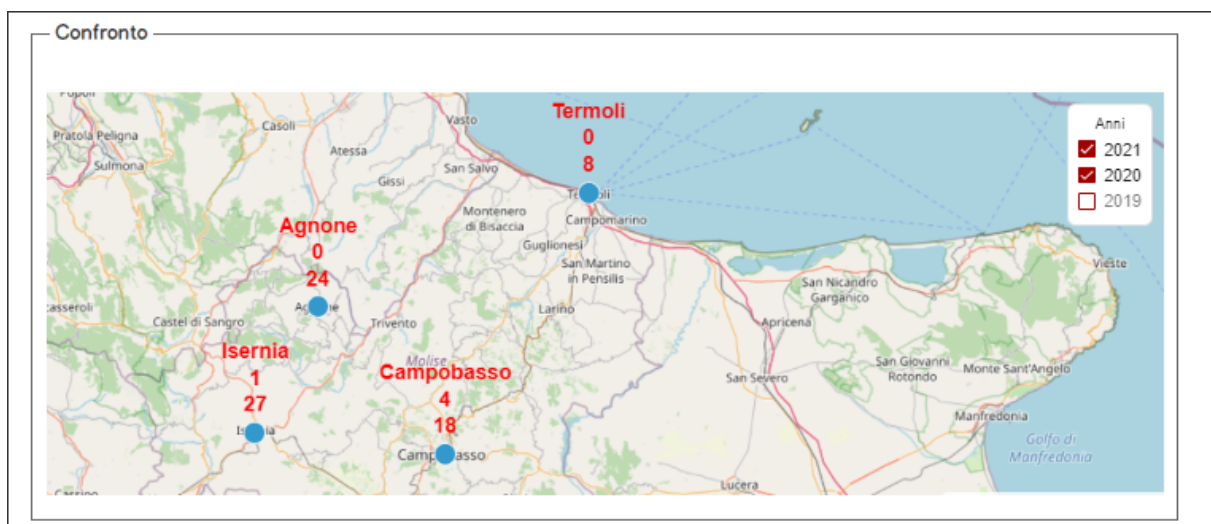


Figura 42. Dashboard-Area Geografica - Confronto - Medico Specialista - Responsabile di Monitoraggio

Nel caso in cui il “Medico Specialista” o il “Responsabile di Monitoraggio” fosse interessato a conoscere in dettaglio le informazioni su una particolare area delle macroaree elencate, potrà selezionare il segnaposto di riferimento. Alla selezione dell’area di interesse, il sistema aprirà un pop-up suddiviso in due parti: nella parte a sinistra riporterà in alto il nome della zona selezionata e al di sotto del nome la mappa geografica dell’area selezionata arricchita da marker che individuano oltre al punto in cui si è verificato l’allarme anche la tipologia dell’allarme riscontrato (Figura 43). I Marker saranno di colore rosso nel caso di allarme critico e di colore giallo nel caso di allarme potenzialmente critico. A destra della mappa sarà riportata la sezione “Anno” dal quale l’utente potrà selezionare l’anno per il quale desidera ottenere informazioni sulla zona d’interesse, in fase iniziale sarà impostato sull’anno corrente. Il “Medico Specialista” o il “Responsabile di Monitoraggio” potrà uscire dal pop up e ritornare alla maschera di provenienza selezionando il bottone “Indietro” riportato in basso a destra.

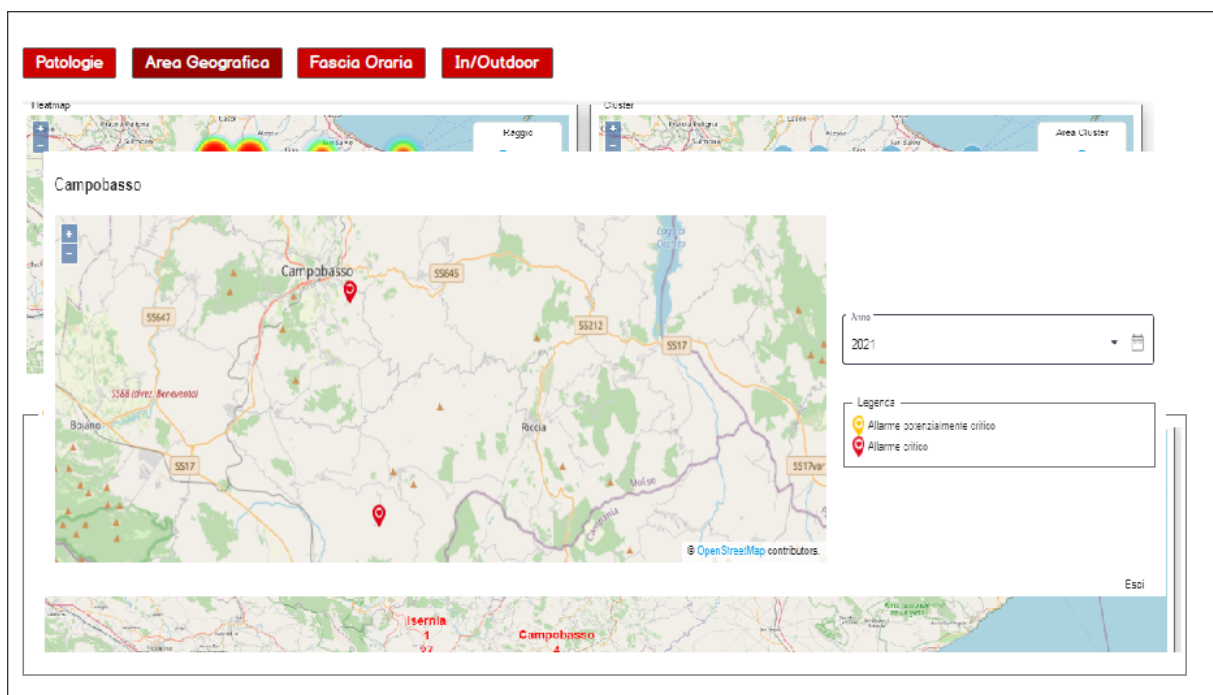


Figura 43. Dashboard – Tutti gli assistiti - Area Geografica – Confronto - Dettaglio –Medico Specialista – Responsabile di Monitoraggio

Selezionando il tab “Fascia Oraria” il sistema mostrerà al Medico Specialista e al Responsabile di Monitoraggio loggato due sezioni: “Fascia oraria per patologia” e “Fascia oraria per tipologia”; nella

prima sezione verrà mostrata attraverso un istogramma la distribuzione degli allarmi per patologia suddivisa per fascia oraria. L'asse dell'ascisse sarà composta da quattro fasce orarie "Mattino", "Pomeriggio", "Sera" e "Notte", e ogni fascia oraria avrà un massimo di 4 rettangoli, dove ciascun rettangolo rappresenta la "Patologia" riscontrata ("Cardiaco", "Respiro", "Dinamico" e "Temperatura") in quella fascia oraria. L'asse delle ordinate riporterà il numero di allarmi riscontrati.

La seconda sezione, invece, mostrerà attraverso un istogramma, la distribuzione degli allarmi per tipologia suddivisa per fascia oraria. In questo caso, l'asse dell'ascisse sarà composta da quattro fasce orarie "Mattino", "Pomeriggio", "Sera" e "Notte", e ogni fascia oraria avrà un massimo di 4 rettangoli, dove ciascun rettangolo rappresenta la "Tipologia" di allarme riscontrato ("Rosso", "Giallo", "Bianco" e "Grigio") in quella fascia oraria. L'asse delle ordinate riporterà il numero di allarmi riscontrati (Figura 44).

Per entrambe le sezioni, il "Medico Specialista" e il "Responsabile di Monitoraggio" loggato potrà scegliere l'anno per il quale desidera ottenere le informazioni popolando il campo "Anno"; in fase iniziale verranno visualizzati gli istogrammi tenendo presente l'anno corrente. I campi mostrati nella sezione sono riportati nella Tabella 3.

Tabella 3. Dashboard – Fascia Oraria

Sezione	Campi	Descrizione	Obbligatorietà	Tipo campo
Fascia oraria – patologie	Anno	Campo utilizzato per selezionare l'anno per il quale l'utente loggato desidera visualizzare le informazioni.	Si	Lista, editabile
Fascia oraria - patologie	Grafico	Grafico contenente le informazioni riguardanti le patologie mostrate attraverso un istogramma in base all'anno scelto.	Si	Grafico, non editabile
Fascia oraria - tipologia	Anno	Campo utilizzato per selezionare l'anno per il quale l'utente loggato desidera visualizzare le informazioni.	Si	Lista, editabile
Fascia oraria - tipologia	Grafico	Grafico contenente le informazioni riguardanti le tipologie mostrate attraverso un istogramma in base all'anno scelto.	Si	Grafico, non editabile

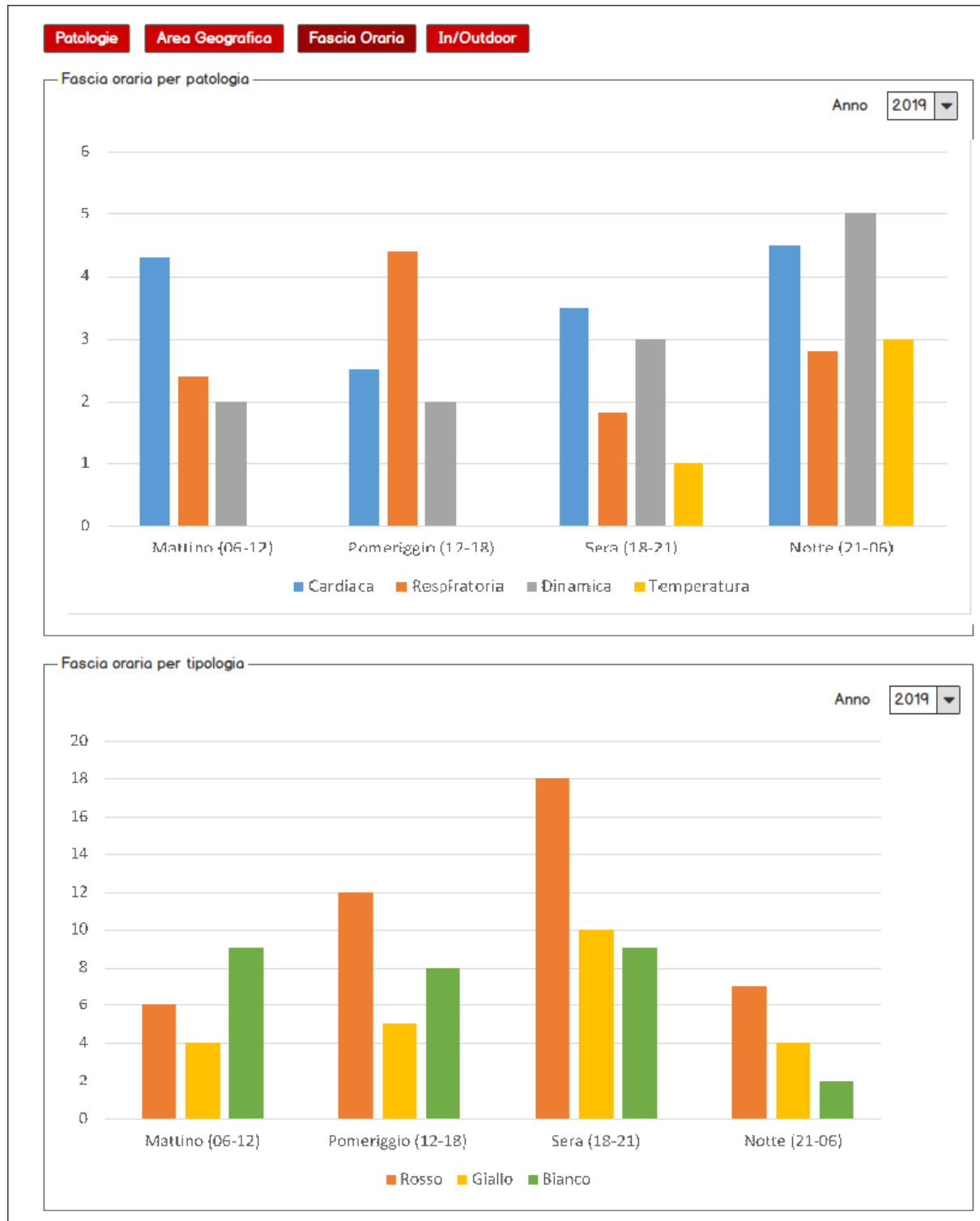


Figura 44. Dashboard – Tutti gli assistiti - Fascia Oraria – Medico Specialista – Responsabile di Monitoraggio

Selezionando il tab “In/Outdoor” il sistema mostrerà al “Medico Specialista” e al “Responsabile di Monitoraggio” loggato due sezioni: “Indoor” e “Outdoor”; nella prima sezione “Indoor” verrà mostrato attraverso un istogramma gli allarmi suddivisi per patologie riscontrati in ambito domestico durante un particolare anno specificato dall’utente loggato, nella seconda sezione “Outdoor” verrà mostrato attraverso un istogramma gli allarmi suddivisi per patologie riscontrati al di fuori dell’ambito domestico durante un particolare anno specificato dall’utente loggato (Figura 45).

Per entrambe le sezioni “Indoor” e “Outdoor”, l’asse dell’ascisse dell’istogramma verrà suddiviso in 12 mesi, e ciascun mese conterrà un massimo di 4 rettangoli ognuno dei quali rappresenterà una particolare patologia (“Cardiaco”, “Respiro”, “Dinamico” e “Temperatura”), mentre l’asse dell’ordinata riporterà il numero di allarmi sollevati. Il “Medico Specialista” e il “Responsabile di Monitoraggio” loggato potrà scegliere l’anno per il quale desidera visualizzare le informazioni popolando il campo “Anno”; in fase iniziale verranno visualizzati gli istogrammi tenendo presente l’anno corrente, e volendo l’utente loggato, potrà aggiungere un ulteriore filtro popolando il campo “Tipologia” (“tutte”, “Rosso”, “Giallo”, “Bianco” e “Grigio”). Inizialmente il sistema recupererà gli allarmi senza tener conto di una particolare tipologia ma terrà conto di tutte le tipologie.

I campi mostrati nella sezione sono riportati in Tabella 4.

Tabella 4. In/outdoor

Sezione	Campi	Descrizione	Obbligatorietà	Tipo campo
Indoor	Anno	Campo utilizzato per selezionare l'anno per il quale l'utente loggato desidera visualizzare le informazioni.	Si	Lista, editabile
Indoor	Tipologia	Campo utilizzato per selezionare la tipologia per la quale l'utente loggato desidera visualizzare le informazioni (tutte, Rosso, Giallo, Bianco)	Si	Lista, selezionabile
Indoor	Grafico	Grafico contenente le informazioni circa le patologie visualizzate attraverso un istogramma in base al mese scelto e alle tipologie avvenute in ambiente domestico.	Si	Grafico, non editabile
Outdoor	Anno	Campo utilizzato per selezionare l'anno per il quale l'utente loggato desidera visualizzare le informazioni.	Si	Lista, editabile
Outdoor	Tipologia	Campo contenente la tipologia per la quale si sta visualizzando le informazioni (tutte, Rosso, Giallo, Bianco)	Si	Lista, selezionabile
Outdoor	Grafico	Grafico contenente le informazioni circa le patologie visualizzate attraverso un istogramma in base al mese scelto e alle tipologie avvenute in ambiente non domestico.	Si	Grafico, non editabile

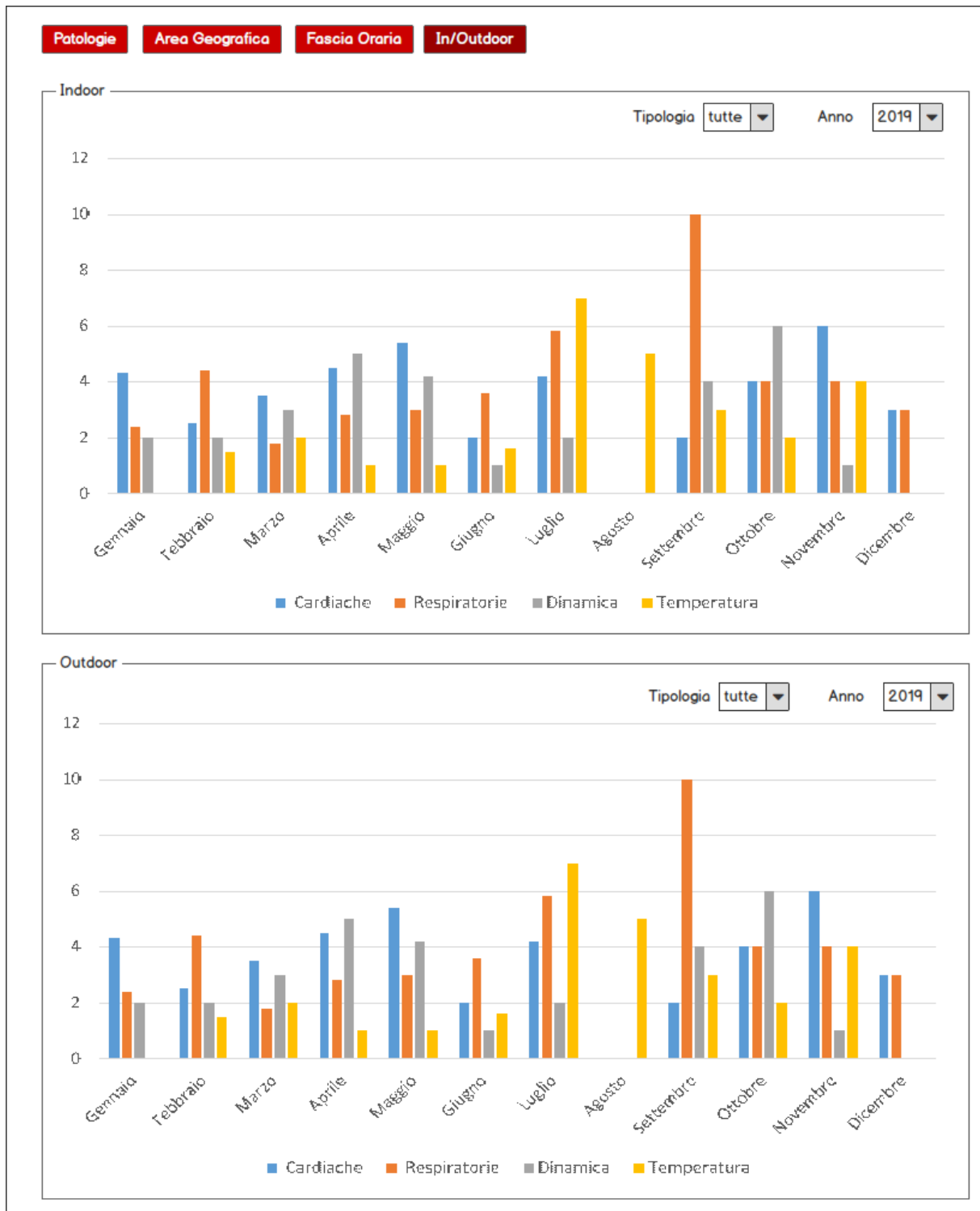


Figura 45. Dashboard – Tutti gli assistiti – In/outdoor – Medico Specialista – Responsabile di Monitoraggio

## 4 Interazione con i dati

Con la crescita delle capacità hardware, attraverso cui si possono registrare i click del mouse, i tweet e i dati relativi agli acquisti, si è costretti a gestire un'elevata quantità di dati. All'inizio la maggior parte dei dati erano strutturati in database relazionali oppure erano dati in formato testuale. Con l'avvento di nuove tecnologie in grado di produrre enormi quantità di dati al secondo (es: veicoli a guida autonoma), si è iniziata ad avere una mole eccessiva di dati. Di conseguenza, è sorto un altro problema: l'etichettatura di questi dati (data labeling). Il data labeling, definito anche come "annotazione dei dati", "tagging dei dati" o "classificazione dei dati", è il processo che permette di preparare i dataset etichettati affinché siano usati per addestrare modelli di Machine Learning, i quali imparano a riconoscere pattern nei dati etichettati. Quando un sufficiente numero di dati etichettati è stato analizzato, i modelli possono identificare gli stessi pattern nei dati che non sono stati etichettati.

I modelli di ML hanno tre processi durante il ciclo di vita:

- **Training del modello:** permette al modello di imparare a ricavare gli output dall'input. Ad esempio, l'input è un'immagine e l'output che dovrà restituire è il nome dell'animale presente nell'immagine.
- **Inferenza:** il modello fa deduzioni, con un certo livello di confidenza. Non tutti i modelli forniscono i livelli di confidenza e livelli di confidenza forniti dai modelli non sempre corrispondono alla probabilità attuale di successo dell'inferenza.
- **Apprendimento continuo:** quando il livello di confidenza per una predizione è basso, è probabile che il modello si sia sbagliato. Se la predizione del modello è utilizzata in una decisione di business, gli utenti imprese preferiscono avere un output del modello revisionato da un essere umano, in questi casi. Questo processo permette di migliorare le performance del modello attraverso l'aggiunta di nuove istanze etichettate.

Grazie al recente progresso fatto nell'hardware, sono state inventate anche le reti neurali, che possono essere usate in pratica per costruire nuove applicazioni basate sull'intelligenza artificiale (AI). La preponderanza dei big data ha aiutato l'Intelligenza Artificiale a trovare la strada giusta

perché i modelli di apprendimento supervisionati richiedono dati etichettati. Ad esempio, nel campo delle immagini viene usato Imagenet, il più grande insieme di immagini disponibili al pubblico usato dagli specialisti di computer vision in tutti i settori. Per nove anni, i suoi contributori hanno annotato manualmente più di 14 milioni di immagini.

In Figure Eight, un AWS Partner Network (APN) Advanced Technology Partner con l'AWS Machine Learning Competency, è stato notato un problema: creiamo tanti dati su tanti dispositivi, ma non ci sono abbastanza persone su questo pianeta per etichettare tutti i dati che stiamo creando. Per questo, nell'industria dell'Intelligenza Artificiale, è stato definito il concetto di human-in-the-loop: dati di alta qualità vengono etichettati manualmente dagli esperti di machine learning (ML) su tutta la linea. Al momento, si stanno sviluppando attivamente soluzioni innovative che usano l'approccio collaborativo tra umani e macchine.

Il processo adottato dai professionisti di Machine Learning è il seguente: si raccoglie un insieme iniziale di dati, li si etichettano e si addestra un modello per un'applicazione nel mondo reale. Questo approccio però non tiene in considerazione il fatto che i dati non sono tutti uguali. Ad esempio, se si volesse allenare un modello per la categorizzazione dell'abbigliamento, si potrebbe assumere che una grande quantità di immagini riguardi le maglie. In questo caso, la categorizzazione funzionerebbe bene per le maglie, ma non per le camicie.

Quando si effettua il training del modello ci sono tre approcci per influenzare il processo adottato:

1. **Scegliere le istanze da etichettare prima della formazione del modello:** usando la distribuzione dei dati nello spazio delle *feature*, è possibile determinare in anticipo quali istanze potrebbero essere più informative per allenare un modello specifico attraverso la prioritizzazione. Ad esempio, se si lavora con i tweet, è una buona idea rimuovere lo spam, annotare il quale costerebbe tempo e denaro, con un vantaggio nullo per il modello finale.
2. **Scegliere le istanze mentre il modello viene addestrato:** con l'apprendimento attivo (*active learning*) è possibile aggiungere gradualmente più istanze selezionate in modo ottimale dinamicamente (o attivamente) mentre un modello viene addestrato. Questo è

uno strumento molto potente, perché permette di controllare l'esatta quantità di dati necessari per l'apprendimento e di utilizzare dati appena sufficienti per raggiungere un'accuratezza specifica. Generalmente, si ottengono risultati migliori rispetto alla prioritizzazione, perché il learner può adattare dinamicamente la propria strategia di apprendimento invece di basarsi su ipotesi preliminari, generalmente statiche.

3. **Scegliere le istanze da correggere dopo aver allenato il modello:** l'idea alla base è quella di allenare il modello usando un training set di qualità non necessariamente alta, per poi instradare poche predizioni su cui la confidenza di predizione è limitata agli annotatori umani affinché questi le revisionino e le validino. In questo "puro" approccio human-in-the-loop, i dati corretti dagli umani non sono riportati nel modello, ma vengono usati per correggere il training set iniziale.

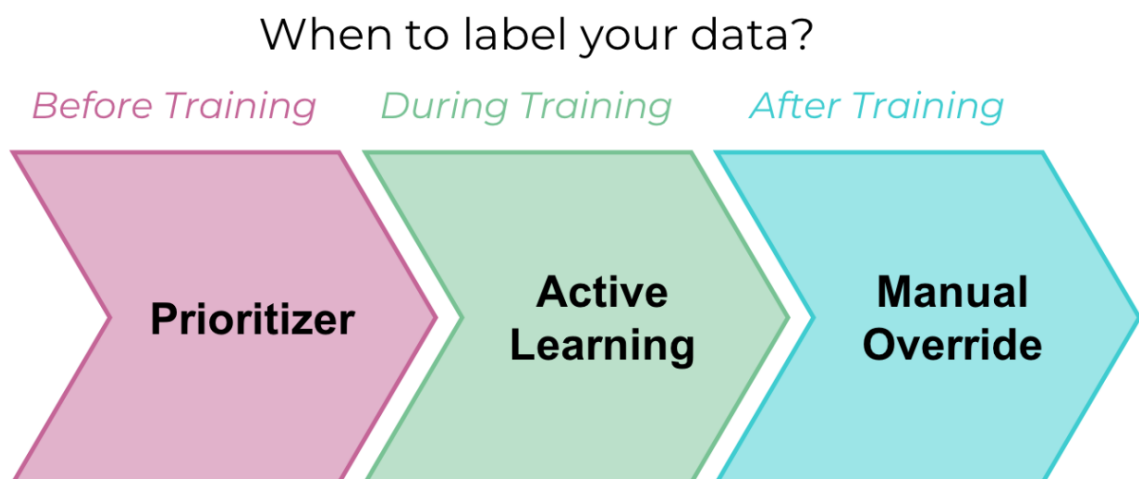


Figura 46. Tre comuni approcci per ridurre i costi di annotazione, prima, durante e dopo il processo di training

Quasi ogni progetto di Machine Learning richiede qualche etichettatura o aumento dei dati primitivi (dati raw) per creare dati che sono pronti ad alimentare modelli e algoritmi intelligenti. Ma il problema fondamentale è, come detto prima, che le organizzazioni hanno troppi dati da etichettare.

Un altro problema che sorge spesso in un gruppo di data scientist è decidere chi sarà il responsabile del data labeling, stimare il tempo necessario per effettuare una tale operazione e quali tool possono essere utilizzati. Il data labeling è una fase indispensabile del preprocessing dei dati

nell'apprendimento supervisionato (*supervised learning*). Il dato storico con attributi target predefiniti (valori) è usato per questo stile del modello di training. Un algoritmo può solo trovare attributi target se un umano li mappa. Gli etichettatori (*labelers*) devono essere estremamente attenti perché un errore o un'inaccuratezza influenza negativamente la qualità del dataset e le performance complessive di un modello predittivo.

## 4.1 Active learning

Il data labeling è richiesto sia nel training del modello iniziale sia nel miglioramento continuo. Inoltre, l'etichettatura dei dati permette alle macchine di ottenere una comprensione accurata delle condizioni del mondo reale e apre a nuove opportunità per un enorme varietà di business.

L'approccio più semplice di labeling è quello di etichettare i dati a portata di mano, creando una base per l'algoritmo di apprendimento automatico. Inoltre, è anche possibile focalizzarsi su parti di dati non etichettati per massimizzare l'apprendimento quando saranno etichettati. L'apprendimento attivo (*active learning*) mira a ridurre tempi e costi.

L'apprendimento attivo è un approccio semi-supervisionato di annotazione dei dati con l'obiettivo di rendere il data labeling più affidabile usando il più possibile le istanze poco etichettate. Nell'apprendimento attivo, l'annotatore seleziona un campione iniziale dai dati non etichettati, seleziona incrementalmente ed etichetta più dati durante il processo, finché ogni dato non è etichettato.

Gli approcci di apprendimento attivo includono:

- **Membership Query Synthesis:** l'allievo attivo (active learner) (il modello di ML che usa l'apprendimento attivo) genera un'istanza sintetica e richiede un'etichetta per esso. L'etichetta non può essere prodotta in modo semplice da un essere umano in tutti i casi. Per esempio, nel caso di un modello di ML che sta provando a risolvere un problema scientifico, l'istanza può essere un esperimento scientifico e l'etichetta può essere il risultato dell'esperimento.

- **Stream-Based Selective Sampling:** Questo metodo presuppone che l'ottenimento di un'istanza non etichettata sia gratuito. Successivamente, l'algoritmo seleziona un'istanza alla volta e decide di etichettarla o ignorarla, basandosi sulla sua informatività.
- **Pool-Based sampling:** questo metodo assume che ci sia un gran numero di istanze non etichettate. Queste vengono classificate in base a una misura di informatività e, successivamente, vengono selezionati i dati migliori da etichettare.

## 4.2 Approcci di labeling

La scelta di un approccio dipende dalla complessità di un problema, dal training set, dalla dimensione di un team di data scientist e dalle risorse finanziarie e temporali che un'azienda può allocare per implementare un progetto.

### 4.2.1 In-house labeling

Per garantire la massima precisione nel data labeling e monitorare il processo, bisogna assegnare questo compito ad un team. Generalmente, l'in-house labeling è molto più lenta degli approcci descritti successivamente, però è la soluzione migliore se l'azienda ha abbastanza risorse umane, temporali e finanziarie.

Una possibile automazione e semplificazione del processo di data labeling può avvenire sfruttando tecniche di apprendimento semi-supervisionato (*semi-supervised learning*). In dettaglio, si prende una porzione di dati da dataset non etichettato e si esegue su di esso la procedura di data labeling, per potervi costruire un classificatore multi-classe. Successivamente, il modello che ne è risultato può essere utilizzato per etichettare la restante parte del dataset.

Lo sviluppo di progetti per industrie specializzate in ambiti differenti, come finanza, medicina, o fonti energetiche richiede una valutazione dei dati svolta da esperti in quel settore. È necessaria quindi la consulenza di un esperto del domino oppure che gli stessi esperti, quando disponibili, si occupino dell'etichettatura dei dati. Il vantaggio dell'in-house labeling è quello di ottenere dati di ottima qualità, risultati predicibili e la possibilità di monitorare il processo di etichettatura. Allo stesso tempo però tutto il processo risulterà più dispendioso.

Un esempio di applicazione è “*DolGring*”, sviluppata da Altexsoft, che mira a diagnosticare e monitorare il bruxismo, ossia digrignare eccessivamente i denti o la mascella. L’applicazione si basa su un algoritmo di classificazione del rumore, che è stato addestrato con un set di dati composto da più di 6.000 campioni audio. Per individuare le registrazioni che contengono il suono del digrigno dei denti, un operatore ha ascoltato i campioni e li ha mappati con attributi. Il riconoscimento di questi suoni specifici è necessario per l’estrazione delle feature.

## 4.2.2 Crowdsourcing

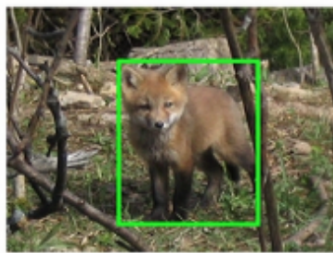
Un’alternativa è usare una piattaforma di crowdsourcing per inviare i task di etichettatura dei dati a un gran numero di etichettatori. Un esempio è la piattaforma creata da Amazon, chiamata Amazon Mechanical Turk (MTurk). Questa è una delle piattaforme leader, che offre forza lavoro on-demand. I clienti si registrano come richiedenti e loro possono creare e gestire i loro progetti con uno o più HITs (Human Intelligence Tasks) sul sito web Mechanical Turk Requester. Il sito web fornisce agli utenti un’interfaccia facile da usare per creare le attività di etichettatura. I rappresentanti di MTurk affermano con un’ampia comunità di lavoratori, il data labeling può richiedere poche ore invece di giorni o settimane.

Clickworker, altra piattaforma disponibile a livello globale, ha più di un milione di professionisti pronti per essere assegnati alla procedura di labeling su immagini, video e attività di sentiment analysis. Le prime fasi del workflow sono simili a quelle utilizzate da MTurk, ma differiscono le fasi dell’elaborazione e dell’allocazione dei task. Gli utenti registrati posizionano i loro ordini con delle richieste specifiche. In seguito, il team della piattaforma elabora e pubblica diversi task che possono essere presi in carico dai freelance registrati alla piattaforma.

I vantaggi di avere una piattaforma crowdsourcing sono la possibilità di avere dei risultati veloci e l’accessibilità della piattaforma. Infatti, questa tecnica può essere usata per progetti con scadenze strette e dataset estesi che richiedono quindi strumenti avanzati di data labeling. Inoltre, l’assegnazione delle attività di etichettature non avrà costi eccessivi.

Tra gli svantaggi c'è il rischio di ottenere un dataset di bassa qualità perché il dato etichettato potrebbe essere inconsistente. La qualità del dato dipende dagli errori che possono sorgere nell'attività di data labeling. Per esempio, la barriera linguistica o anche l'assegnazione di particolari task a professionisti non adatti a quella specifica tipologia. Le piattaforme di crowdsourcing usano diverse misure di garanzia di qualità per far fronte a questo problema, come garantire ai propri lavoratori una ricompensa solo se determinati obiettivi sono raggiunti. Inoltre, si possono definire requisiti relativi al formato, agli strumenti e/o alle procedure da far adottare ai professionisti durante il processo di etichettatura. Un'altra strategia che le piattaforme usano per garantire la qualità è la somministrazione di un test di valutazione, che gli etichettatori devono superare per poter iniziare a lavorare tramite la piattaforma.

**Rule 1: Include all visible part and draw as tightly as possible.**



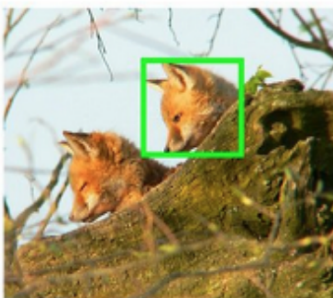
**CORRECT**



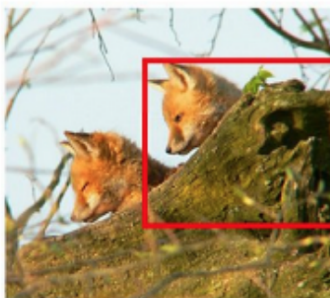
**WRONG: must be as tight as possible!**



**WRONG: must include all visible parts!**



**CORRECT**



**WRONG: occluded parts do not matter as long as all visible parts are included.**

Source: cs.stanford.edu

Figura 47. Illustrazione dell'etichettatura delle immagini

### 4.2.3 Outsourcing a privati

Un altro modo per velocizzare il processo di etichettatura è quello di cercare professionisti su siti di social networking, freelancer e reclutamento.

Il vantaggio è che è possibile verificare le competenze dei candidati con dei test mirati alla tipologia di task che dovranno svolgere. L'outsourcing comporta l'assunzione di una squadra di piccole o medie dimensioni, con la possibilità di seguire da vicino le modalità di svolgimento di tutto il lavoro. Lo svantaggio è che diventa necessario gestire il workflow, dando istruzioni chiare e dettagliate, per coordinare tutti i lavoratori esternalizzati. Inoltre, c'è il bisogno di tempo extra per controllare e revisionare le attività completate.

### 4.2.4 Outsourcing a società

È possibile anche contattare le aziende specializzate di outsourcing nella preparazione dei dati di addestramento, che si posizionano come alternativa alle piattaforme di crowdsourcing. Tali organizzazioni combinano etichettatori di dati controllati, addestrati e gestiti attivamente, localizzati in remoto a un team interno, in modo da garantire la qualità dei dati etichettati. In questo tipo di organizzazione le attività possono cambiare mentre i team di sviluppo addestrano e fanno *tuning* dei loro modelli; i team di etichettatura, quindi, devono essere in grado di adattarsi rapidamente e apportare modifiche al flusso di lavoro.

Il vantaggio è avere risultati di alta qualità, ma ad un costo maggiore rispetto al crowdsourcing.

### 4.2.5 Synthetic labeling

Questo approccio comporta la generazione di dati sintetici che simulano dati reali, sulla base di parametri e vincoli definiti dall'utente. I dati sintetici sono prodotti da un modello generativo che viene addestrato e validato su un dataset originale.

I tre tipi di modelli generativi sono i seguenti:

- **Generative Adversarial Networks** (o GAN): i modelli GAN sono basati su un metodo di apprendimento semi-supervisionato e sono separati in due sotto-modelli: una rete

generatrice e una rete discriminatoria. Esse sono usate in un framework di gioco a somma zero, ossia viene effettuata una gara in cui una rete generatrice produce campioni di dati e una rete discriminatoria, addestrata sui dati reali, cerca di definire se sono reali, cioè se sono provenienti dalla distribuzione di dati vera, o generati, cioè che sono venuti dalla distribuzione di modello. Il gioco continua fino a quando un modello generativo riceve abbastanza feedback per essere in grado di riprodurre immagini che sono indistinguibili da quelle reali;

- **Autoregressive models (AR):** i modelli AR generano variabili basate sulla combinazione lineare di valori precedenti di variabili. Nel caso di immagini generate, gli AR creano pixel individuali basati su pixel precedenti posizionati in un pixel in alto a sinistra;
- **Variational Autoencoders (VAE):** i VAE producono nuovi campioni di dati da input attraverso metodi di codifica e decodifica.

I dati sintetici hanno applicazioni multiple e possono essere usati per il training delle reti neurali, ossia dei modelli usati per le attività di riconoscimento dell'oggetto. In questo modo è possibile ottenere dataset di grandi dimensioni che contengono ad esempio file video, oppure file di testo, immagini o file audio. La grandezza della rete e del training data sono fortemente correlati alla complessità del task.

Con questa tecnica si ottiene un labeling rapido ed economico. In questo modo il dato sintetico può essere generato velocemente, personalizzato per una specifica attività e modificato per migliorare un modello e il training stesso. Inoltre, questa tecnica permettere di usare il dato non sensibile senza bisogno di permessi per il data scientist.

Gli svantaggi sono la necessità di eseguire una computazione ad alte prestazioni e di avere problemi con la qualità dei dati. Infatti, questo approccio richiede alta potenza computazionale per fare rendering dei dati e ulteriori training del modello. La qualità dei dati, invece, è legata all'uso di dati sintetici che possono non assomigliare completamente ai dati storici reali; quindi, il modello dovrà essere riallenato appena i dati reali saranno disponibili.

## 4.2.6 Data programming

I data scientist dal progetto Snorkel hanno sviluppato il data programming, un approccio per la creazione del training data e la gestione che elimina la necessità del labeling manuale. Questo approccio comporta la scrittura delle funzioni di etichettatura, ossia di script che etichettano automaticamente i dati. Le etichette risultano essere più accurate di quelle create per il labeling manuale. Un dataset rumoroso generato da un programma può essere utilizzato per una supervisione debole di modelli finali di alta qualità, come i modelli di Tensorflow.

Le funzioni di etichettatura possono essere combinate con le Generative Adversarial Networks (GANs) nel processo di labeling dei dati.

## 4.3 Tool di data labeling

I tool di data labeling sono strumenti essenziali nella creazione di un modello di ML efficace. Per poter scegliere il tool migliore è necessario determinare la tipologia di data labeling necessaria e soprattutto se c'è bisogno di sicurezza per i dati (astrazione e anonimizzazione).

Un altro importante aspetto è l'accuratezza del labeling, perché dei dati di alta qualità generano migliori performance del modello. Quando la qualità è bassa il modello di Machine Learning ha bisogno di più tempo per apprendere. Per esempio, alcuni tool di data labeling effettuano il pre-processing dei dati non strutturati con modelli di ML che etichettano parzialmente il dato. In questo modo, il personale trova più semplice etichettare i dati e aumentare la precisione di etichettatura.

I tool di data labeling possono essere divisi in differenti categorie.

- Alcuni tool sono progettati per essere eseguiti su un singolo tipo di dati (es: testo, immagini, audio o video), mentre pochi di loro possono funzionare in domini multipli simultaneamente.
- I tool di annotazione possono anche differire dal tipo di approccio utilizzato per l'etichettatura. Alcuni tool, come MedTagger e imagetagger, sono adattabili per il labeling collaborativo dei dati. Le organizzazioni che preferiscono usare il crowdsourcing possono usare questo tipo di tool per il labeling.

- I tool differiscono anche in base alla tipologia di etichettatura testuale, come Box di delimitazione, annotazione poligonale e segmentazione semantica.
- Oltre ai tool commerciali, esistono anche tool di labeling con licenze Open Source oppure Freeware. Le aziende possono preferire diverse licenze in base alle necessità dell'organizzazione.

I tool di data labeling si possono differenziare in tool basati su browser (browser-based) e basati su desktop (desktop-based).

Tra i tool di labeling per immagini e video possiamo menzionare:

- **Annotorious**, un tool di labeling per le immagini con licenza MIT. Esso permette di aggiungere commenti testuali e disegni sulle immagini di un sito. Il tool può essere facilmente integrato con solo due linee di codice aggiuntive. Gli utenti possono imparare le caratteristiche del tool e le attività di annotazione presenti nella sezione Demo. Inoltre, Annotorious può essere modificato ed esteso attraverso dei plugin che li rendono adattabili per un bisogno del progetto. Gli sviluppatori incoraggiano gli utenti nel valutare e migliorare Annotorious, condividendo le loro esperienze con la comunità.

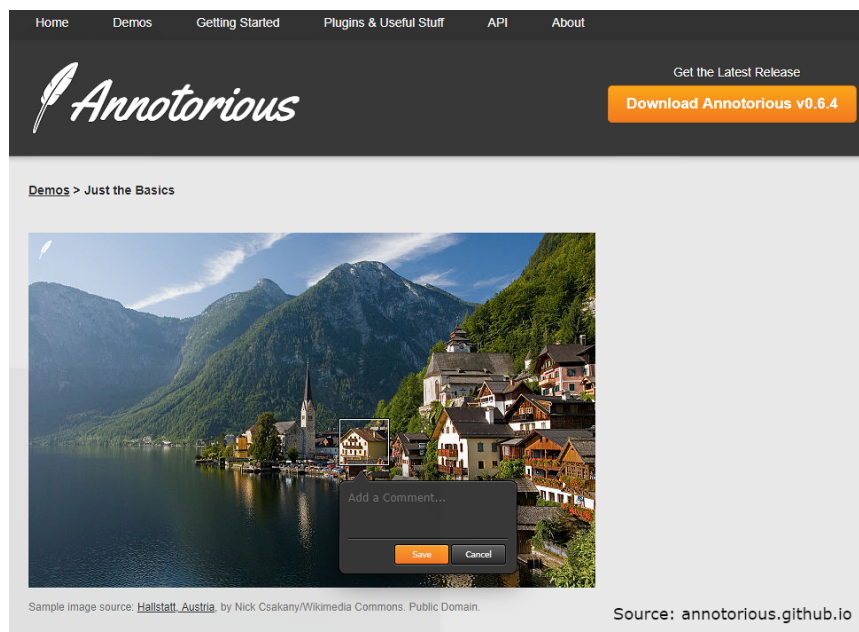


Figura 48. Annotorious



di annotazione, gli utenti possono anche scegliere se le caselle di delimitazione devono essere quadrate o rettangolari.

Tra i tool di data labeling sono presenti anche i tool specifici per text labeling, che semplificano il flusso di lavoro di etichettatura per le attività relative alla NLP, come sentiment analysis. Alcuni esempi sono:

- **Stanford CoreNLP**, il quale consente di completare vari compiti di elaborazione e analisi dei dati di testo;
- **Bella**, strumento disponibile liberamente che permette di accelerare l'etichettatura dei dati di testo. Un insieme di dati etichettato con un file CSV o un foglio di calcolo Google deve essere convertito in un formato appropriato prima della formazione del modello. Le principali caratteristiche di Bella sono quelle di avere un'interfaccia grafica e un database che permette di gestire i dati etichettati. Un utente crea e configura un progetto per ogni labeling del dataset e lui può visualizzare l'elemento, i tipi di etichetta e i tag supportati dal tool;
- **Tagtog** è uno strumento web per l'annotazione e la categorizzazione di testo automatizzato. I tre approcci proposti all'utente sono i seguenti: annotare il testo manualmente, consentire ad un team di etichettare i propri dati o usare dei modelli di machine learning per annotazioni automatizzate;

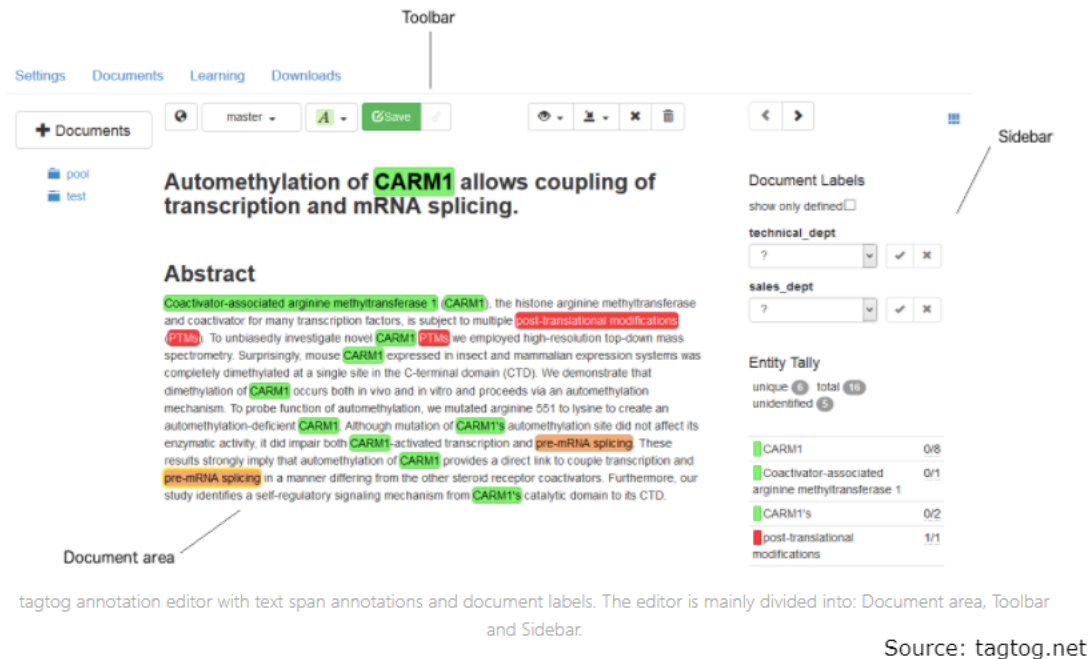


Figura 50. Tagtog

- **Daturks** fornisce i tool per la preparazione del training data. Esso permette di caricare i dati, invitare collaboratori e iniziare il tagging. Inoltre, permette di etichettare anche le immagini, i file audio e i file video.

I tool di data labeling sono necessari anche per allenare le reti neurali nel riconoscimento del suono e nelle attività di classificazione della musica:

- **Praat** è un software libero per il labeling dei file audio, in cui si possono annotare gli eventi con etichette testuali associate a posizioni temporali del file audio, ottenendo in output un file testuale di tipo TextGrid. Quindi questo tool consente di lavorare con le annotazioni di testo taggate all'interno dei file audio.
- **Speechalyzer**: tool per l'elaborazione manuale di grandi dataset vocali.
- **EchoML**: tool di annotazione dei file audio, che permette all'utente la visualizzazione dei dati.

## 4.4 Etichettatura dei dati

In ATTICUS, l'etichettatura dei dati è un processo che si verifica nel momento in cui qualcosa non ha funzionato nella predizione per allarmi di tipo cardiaco in un contesto di continuous learning. In particolare, il DSS di qualsiasi livello genera un allarme di tipo cardiaco che successivamente verrà gestito da un "Responsabile di monitoraggio" o da un "Medico Specialista", associato in precedenza all'assistito, per il quale è stato generato l'allarme. Nel caso in cui, l'allarme di tipo cardiaco, sollevato dal DSS viene successivamente rifiutato dal "Medico Specialista" o dal "Responsabile di Monitoraggio", significa che occorre perfezionare il processo di predizione del DSS. Per poter migliorare il processo di predizione del DSS, la traccia elettrocardiografica dell'allarme rifiutato viene analizzata e da essa viene estratto un campione di battiti che caratterizzano la traccia elettrocardiografica attraverso una procedura di *sampling*.

La scelta dei campioni all'interno di una traccia elettrocardiografica è guidata dall'obiettivo di selezionare un sottoinsieme di battiti che siano il più diversi possibile l'uno dall'altro. In altre parole, si vuole ridurre al minimo la probabilità che lo specialista debba etichettare più volte battiti simili. Da un lato, il classificatore trarrebbe un vantaggio limitato dall'osservare, in fase di addestramento, dati etichettati relativi a più battiti simili; dall'altro, si utilizzerebbe male il prezioso tempo dedicato dai medici specialisti alla procedura di etichettatura. Per raggiungere tale obiettivo, quindi, si è optato per l'utilizzo di un algoritmo *greedy* per la selezione dei centri (*center selection*) [42]. In primo luogo, si divide la traccia elettrocardiografica in sotto-tracce, ognuna delle quali rappresenta un singolo battito, utilizzando l'algoritmo di Pan e Tompkins [41]. Le tracce elettrocardiografiche relative a tali battiti vengono inserite in un insieme  $D$  di battiti disponibili. Come primo passo, si seleziona casualmente uno dei battiti in  $D$  e lo si sposta in un nuovo insieme  $C$ , il quale contiene tutti i battiti da etichettare. Successivamente, l'algoritmo seleziona, volta per volta, il battito  $b$  appartenente a  $D$  che ha distanza massima rispetto a  $C$  (ovvero, il più diverso rispetto ai battiti già selezionati). Per calcolare la distanza tra l'insieme  $C$  e ogni battito candidato  $b'$  appartenente a  $D$ , si calcola la distanza minima tra  $b'$  e ogni  $b$  appartenente a  $C$ , attraverso la formula  $dist(b', C) = \min_{b \in C} dist(b', b)$ . Tale operazione viene ripetuta  $k-1$  volte, dove  $k$  indica il numero di battiti che si

vogliono selezionare. Per calcolare la distanza tra battiti, si utilizza la distanza euclidea: dati due vettori,  $v$  e  $w$ , che rappresentano i segnali delle tracce elettrocardiografiche dei due battiti, si rendono tali vettori omogenei in termini di dimensioni facendo un sub-sampling del vettore di dimensione maggiore e, successivamente, si usa la formula  $\sum_{i=1}^n (v_i - w_i)^2$ . La scelta del numero di battiti da selezionare,  $k$ , può essere personalizzata in fase di installazione del software di monitoraggio.

Il campione di battiti selezionato viene trasformato in quello che si chiama task di etichettatura. Il task, creato automaticamente dal sistema, viene assegnato ai Medici Specialisti o ai Responsabili di Monitoraggio del sistema ATTICUS e prevede lo svolgimento di  $n$  attività di etichettatura, dove  $n$  è il numero di battiti estratti dalla procedura di sampling. Tale task viene creato e assegnato a tutti i “Medici Specialisti” e “Responsabili di Monitoraggio” del sistema. I “Responsabili di Monitoraggio” e i “Medici specialisti” del sistema ATTICUS potranno accedere al task assegnato attraverso la voce di menu “Etichettatura Dati”. Tra le sotto-voci riportate in “Etichettatura Dati”, il sistema mostrerà la sottovoce “Battiti” (Figura 51).

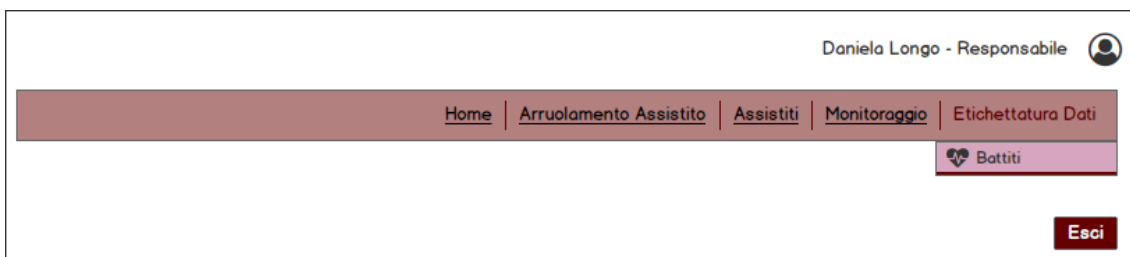


Figura 51. Etichettatura dei dati – Battiti

Alla selezione della voce “Battiti” il sistema mostrerà una pagina suddivisa in due sezioni: “Task di Etichettatura” e “Classifica” attraverso le quali l’utente loggato potrà visualizzare rispettivamente i task di etichettatura a lui assegnati e la classifica di tutti i medici specialisti o responsabile di monitoraggio che hanno eseguito le etichettature dei task.

La sezione “Task di Etichettatura” riporterà i task di etichettatura assegnati al “Medico Specialista” o “Responsabile di Monitoraggio” loggato, indicando in forma tabellare la data in cui è stato

assegnato il task di etichettatura, la data entro la quale deve essere consegnata l'etichettatura e lo stato indicante l'avanzamento dell'etichettatura.

La sezione "Classifica" riporta l'elenco di tutti gli utenti abilitati alla funzionalità di etichettatura dei dati che hanno effettuato una o più etichettature, riportando per ciascuno di essi il cognome, il nome e il numero di tag etichettati.

I campi delle sezioni sono riportati in Tabella 5.

Tabella 5. Campi della maschera Etichettatura dei Dati

Sezione	Campi	Descrizione	Obbligatorietà	Tipo campo
Task di Etichettatura	Data	Data in cui è stato assegnato il task	Si	Data, non editabile
Task di Etichettatura	Data consegna	Data in cui deve essere consegnato il task di etichettatura assegnato	Si	Data, non editabile
Task di Etichettatura	Stato	Stato del task di etichettatura	Si	Testuale, non editabile
Classifica	Cognome	Cognome del medico specialista o responsabile di monitoraggio che ha eseguito uno o più task di etichettatura	Si	Testuale, non editabile
Classifica	Nome	Nome del medico specialista o responsabile di monitoraggio che ha eseguito uno o più task di etichettatura	Si	Testuale, non editabile
Classifica	Punteggio	Punteggio ottenuto dall'utente che ha eseguito i task di etichettatura. Calcolato in base al numero di task di etichettatura eseguito.	Si	Testuale, non editabile

Daniela Longo - Responsabile

Home | Dashboard | Assistiti | Arruolamento Assistito | **Monitoraggio** | Etichettatura Dati

Battiti

Task di Etichettatura

Data	Data Consegna	Stato	Azioni
13/06/2020	18/06/2020	5/20	
19/05/2020	22/05/2020	15/20	
22/04/2020	27/04/2020	12/20	
09/03/2020	12/03/2020	8/20	

Classifica

Cognome	Nome	Punteggio
Rossi	Mario	350
Liberi	Salerno	100
Izzo	Roberta	230
Giallo	Simone	189

Figura 52. Etichettatura Dati – Battiti

Dalla sezione “Task di Etichettatura” l’utente loggato, sia “Medico Specialista” sia “Responsabile di Monitoraggio”, potrà effettuare un’etichettatura dei dati selezionando l’azione riportata in ciascun record della tabella dei task da etichettare.

Alla selezione dell’azione il sistema mostrerà un pop-up Figura 53 contenente i campi riportati in Tabella 6.

Tabella 6. Campi del pop-up del task da etichettare

Campi	Descrizione	Obbligatorietà	Tipo campo
-	Grafico del battito	Si	Grafico
Classificazione	Classificazione del battito da scegliere attraverso una lista precompilata	Si	Lista, editabile
Commento	Campo in cui inserire eventualmente un commento.	No	Testuale, non editabile

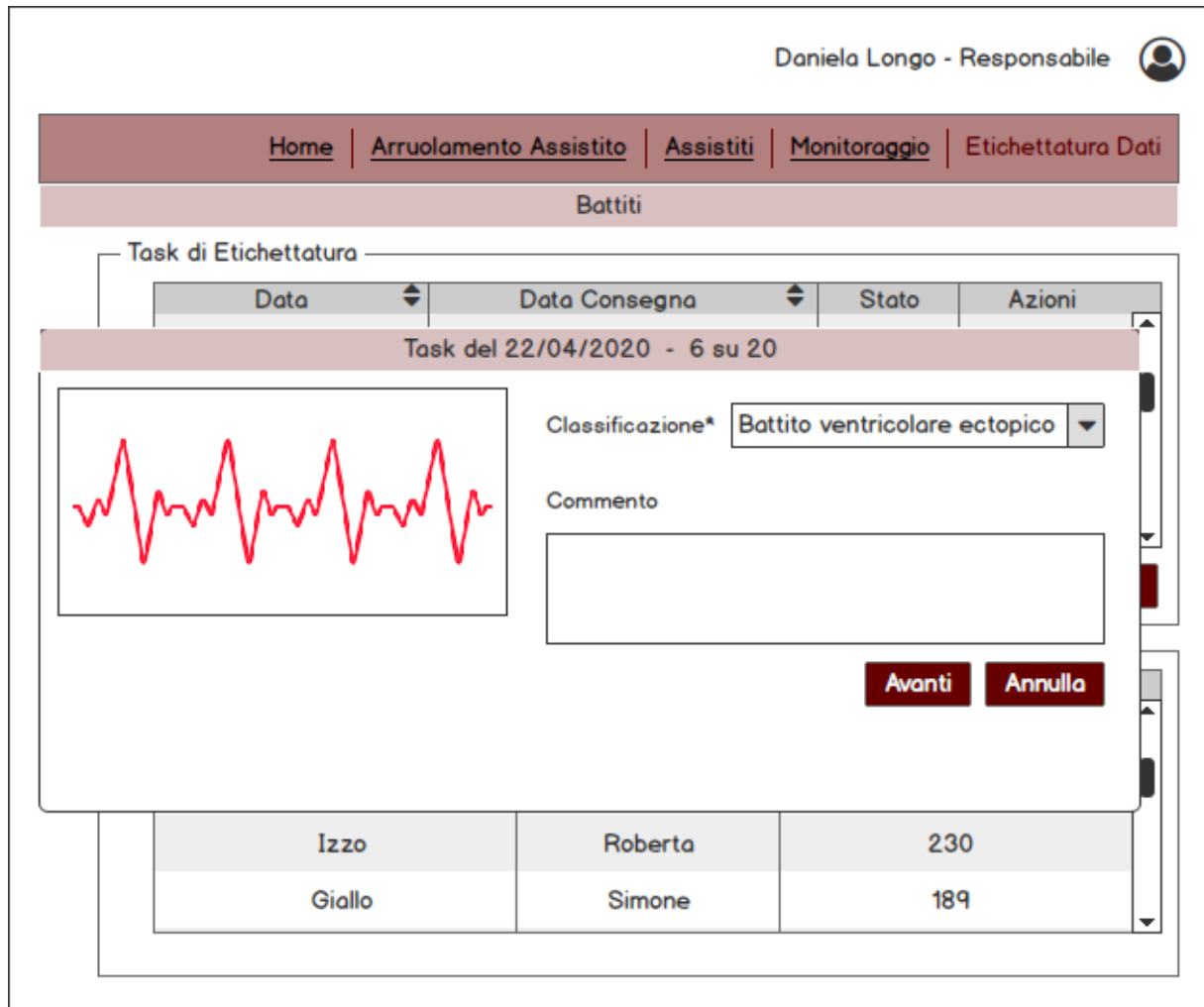


Figura 53. Pop up etichettatura dati

Il sistema all'apertura del pop up mostrerà a sinistra un battito secondo lo standard picco R picco R che il "Medico Specialista" loggato o il "Responsabile di Monitoraggio" dovrà etichettare scegliendo uno dei valori proposti nel campo "Classificazione".

Il campo "Classificazione" sarà precompilato automaticamente dal sistema con i seguenti valori: Anomalia ST, Blocco di Branca Sinistra, Blocco di Branca Destra, Extrasistoli Ventricolari, Fibrillazione Atriale e Battito Atriale Prematuro.

Oltre ai campi da compilare saranno mostrati due bottoni:

- **Avanti:** salva l'etichettatura appena esaminata dopo aver effettuato i controlli sulla validità e sull'obbligatorietà dei campi proposti nel pop up e nel caso di superamento dei controlli mostra il prossimo task da etichettare.
  - in caso di campi obbligatori non popolati il sistema mostrerà il messaggio *"Attenzione! Campi obbligatori non compilati"* e verranno evidenziati in rosso i campi obbligatori non compilati che hanno generato l'errore;
  - in caso di campi compilati in modo non valido verrà mostrato il messaggio *"Attenzione! Campo/i non popolati correttamente"* e verranno evidenziati in rosso i campi non popolati correttamente;
- **Annulla:** svuota i campi precedentemente popolati senza salvare l'etichettatura e riporta l'utente loggato alla maschera di provenienza.

## 4.5 Esportazione programmata

Nel processo di etichettatura dei dati precedentemente descritto, ogni battito appartenente ad uno specifico task riceverà una serie di etichette e per esso verrà memorizzato il "Medico Specialista" o "Responsabile di Monitoraggio" che ha eseguito il task e la classificazione data. Alla fine del processo di etichettatura del battito, quindi, ciascun battito avrà un elenco di etichette assegnate da tutti i medici che hanno eseguito il task e una sola etichetta che è quella che è stata selezionata più volte. In tal modo, periodicamente, il DSS accede all'etichettature dei battiti e nel momento in cui riscontra che un battito ha ricevuto più di  $m$  etichettature ed un'etichetta è stata selezionata per quel battito un numero di volte  $t\%$ , dove  $m$  e  $t$  sono parametrici, quel battito viene integrato in un nuovo dataset di training e viene utilizzato per riaddestrare il modello.

## Riferimenti bibliografici

- [1] J. Höller, V. Tsiatsis, C. Mulligan, S. Karnouskos, S. Avesand, and D. Boyle. *From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence*. Amsterdam, The Netherlands. Elsevier, 2014.
- [2] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the Internet of Things," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2010.
- [3] K. Romer, B. Ostermaier, F. Mattern, M. Fahrmaier, and W. Kellerer, "Real-time search for real-world entities: A survey," *Proceedings of the IEEE International Conference on*, vol. 98, no. 11, pp. 1887–1902, 2010.
- [4] Dukes, L. C., Bertrand, J., Gupta, M., Armstrong, R., Fasolino, T., Babu, S., & Hodges, L. F. (2013, September). Empirical Evaluation of Traditional vs. Hybrid Interaction Metaphors in a Multitask Healthcare Simulation. In 2013 IEEE International Conference on Healthcare Informatics (pp. 89-98). IEEE
- [5] Althoff, F., McGlaun, G., Lang, M., & Rigoll, G. (2003, April). Evaluating multimodal interaction patterns in various application scenarios. In International Gesture Workshop (pp. 421-435). Springer, Berlin, Heidelberg.
- [6] Fukazawa, R., Takashima, K., Shoemaker, G., Kitamura, Y., Itoh, Y., & Kishino, F. (2010, March). Comparison of multimodal interactions in perspective-corrected multi-display environment. In 2010 IEEE Symposium on 3D User Interfaces (3DUI) (pp. 103-110). IEEE.
- [7] Bowman, D. A., & Hodges, L. F. (1999). Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing*, 10(1), 37-53.
- [8] Steinicke, F., Ropinski, T., & Hinrichs, K. (2006). Object selection in virtual environments using an improved virtual pointer metaphor. In *Computer vision and graphics* (pp. 320-326). Springer, Dordrecht.
- [9] Martinet, A., Casiez, G., & Grisoni, L. (2010, March). The design and evaluation of 3d positioning techniques for multi-touch displays. In 2010 IEEE symposium on 3D user interfaces (3DUI) (pp. 115-118). IEEE.
- [10] Adams, E. K., Intwala, M., & Kapadia, A. (2010, November). MeD-Lights: a usable metaphor for patient controlled access to electronic health records. In *Proceedings of the 1st ACM International Health Informatics Symposium* (pp. 800-808).
- [11] Mandl, K. D., Markwell, D., MacDonald, R., Szolovits, P., & Kohane, I. S. (2001). Public standards and patients' control: how to keep electronic medical records accessible but private. *Medical information: access and privacy*. Doctrines for developing electronic medical records. Desirable characteristics of electronic medical records. Challenges and limitations for electronic medical records. Conclusions. Commentary: Open approaches to electronic patient records. Commentary: A patient's viewpoint. *Bmj*, 322(7281), 283-287.
- [12] Røstad, L. (2008, March). An initial model and a discussion of access control in patient controlled health records. In 2008 Third International Conference on Availability, Reliability and Security (pp. 935-942). IEEE.
- [13] Weitzman, E. R., Kaci, L., & Mandl, K. D. (2009). Acceptability of a personally controlled health record in a community-based setting: implications for policy and design. *Journal of medical Internet research*, 11(2), e14.

- [14] Kapadia, A., Henderson, T., Fielding, J. J., & Kotz, D. (2007, May). Virtual walls: Protecting digital privacy in pervasive environments. In International Conference on Pervasive Computing (pp. 162-179). Springer, Berlin, Heidelberg.
- [15] O'Donoghue, S. I., Baldi, B. F., Clark, S. J., Darling, A. E., Hogan, J. M., Kaur, S., ... & Swedlow, J. R. (2018). Visualization of biomedical data. *Annual Review of Biomedical Data Science*.
- [16] Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of internal medicine*, 165(13), 1493-1499.
- [17] Lewandowsky, S., & Spence, I. (1989). The perception of statistical graphs. *Sociological Methods & Research*, 18(2-3), 200-242.
- [18] Wong, B. (2010). Color coding.
- [19] Schneiderman, B., Plaisant, C., & Hesse, B. W. (2013). Improving health and healthcare with interactive visualization methods. *HCIL Technical Report*, 1, 1-13.
- [20] West, V. L., Borland, D., & Hammond, W. E. (2015). Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association*, 22(2), 330-339.
- [21] Powsner, S. M., & Tufte, E. R. (1994). Graphical summary of patient status. *Lancet*, 344(8919), 386-389.
- [22] Plaisant, C., Milash, B., Rose, A., Widoff, S., & Shneiderman, B. (1996, April). LifeLines: visualizing personal histories. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 221-227).
- [23] Gotz, D., Sun, J., Cao, N., & Ebadollahi, S. (2011). Visual cluster analysis in support of clinical decision intelligence. In *AMIA Annual Symposium Proceedings (Vol. 2011, p. 481)*. American Medical Informatics Association.
- [24] Wang, T. D., Wongsuphasawat, K., Plaisant, C., & Shneiderman, B. (2010, November). Visual information seeking in multiple electronic health records: design recommendations and a process model. In Proceedings of the 1st ACM international health informatics symposium (pp. 46-55).
- [25] Klimov, D., & Shahar, Y. (2005). A framework for intelligent visualization of multiple time-oriented medical records. In *AMIA Annual Symposium Proceedings (Vol. 2005, p. 405)*. American Medical Informatics Association.
- [26] Gotz, D., & Wongsuphasawat, K. (2012). Interactive intervention analysis. In *AMIA annual symposium proceedings (Vol. 2012, p. 274)*. American Medical Informatics Association.
- [27] Wongsuphasawat, K., & Gotz, D. (2012). Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2659-2668.
- [28] Shaban-Nejad, A., Lavigne, M., Okhmatovskaia, A., & Buckeridge, D. L. (2017). PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Ann NY Acad Sci*, 1387(1), 44-53.
- [29] Lavrač, N., Bohanec, M., Pur, A., Cestnik, B., Debeljak, M., & Kobler, A. (2007). Data mining and visualization for decision support and modeling of public health-care resources. *Journal of Biomedical Informatics*, 40(4), 438-447.
- [30] Tilahun, B., Kauppinen, T., Keßler, C., & Fritz, F. (2014). Design and development of a linked open data-based health information representation and visualization system: potentials and preliminary evaluation. *JMIR medical informatics*, 2(2), e31.

- [31] Meyer, J., Kazakova, A., Büsing, M., & Boll, S. (2016, October). Visualization of complex health data on mobile devices. In Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care (pp. 31-34).
- [32] Sopan, A., Noh, A. S. I., Karol, S., Rosenfeld, P., Lee, G., & Shneiderman, B. (2012). Community Health Map: A geospatial and multivariate data visualization tool for public health datasets. *Government Information Quarterly*, 29(2), 223-234.
- [33] Galletta, A., Carnevale, L., Bramanti, A., & Fazio, M. (2018). An innovative methodology for big data visualization for telemedicine. *IEEE Transactions on Industrial Informatics*, 15(1), 490-497.
- [34] Tominski, C., Schulze-Wollgast, P., & Schumann, H. (2005, July). 3d information visualization for time dependent data on maps. In Ninth International Conference on Information Visualisation (IV'05) (pp. 175-181). IEEE.
- [35] Mahmud, S., Iqbal, R., & Doctor, F. (2016). Cloud enabled data analytics and visualization framework for health-shocks prediction. *Future Generation Computer Systems*, 65, 169-181.
- [36] Holzinger, A. (2012). On knowledge discovery and interactive intelligent visualization of biomedical data. In Proceedings of the Int. Conf. on Data Technologies and Applications DATA (pp. 5-16).
- [37] Jovanov, E., Starcevic, D., & Radivojevic, V. (2001). Perceptualization of biomedical data. *IN MEDICINE*, 189.
- [38] Kerfoot, E., Fovargue, L., Rivolo, S., Shi, W., Rueckert, D., Nordsletten, D., ... & Razavi, R. (2016, August). Eidolon: visualization and computational framework for multi-modal biomedical data analysis. In International Conference on Medical Imaging and Augmented Reality (pp. 425-437). Springer, Cham.
- [39] Zhang, M., Zhang, H., Tjandra, D., & Wong, S. T. (2004). DBMap: a space-conscious data visualization and knowledge discovery framework for biomedical data warehouse. *IEEE Transactions on Information Technology in Biomedicine*, 8(3), 343-353.
- [40] Institute of Medicine, US. (2012) Committee on Patient Safety and Health Information Technology. Health IT and patient safety: building safer systems for better care. National Academies Press.
- [41] Pan, J., & Tompkins, W.J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 3, (pp. 230-236).
- [42] Kleinberg J., & Tardos E. (2006), *Algorithm design*. Pearson Education India.