

INAIL

Quaderni di ricerca

RECKON: Tecnologie abilitanti per il monitoraggio degli elementi di contesto (Operatore-Macchina-Ambiente) per la prevenzione di incidenti sul lavoro

Davide Accordini, Patrizia Agnello, Silvia M. Ansaldi, Stefano Arrigoni, Francesco Braghin, Enrico Cagno, Cinzia Frascheri, Emilia Lenzi, Davide Piantella, Manuel Roveri, Letizia Tanca, Gabriele Viscardi

23

gennaio 2024

INAIL
Quaderni di ricerca

RECKON: Tecnologie abilitanti per il monitoraggio degli elementi di contesto (Operatore-Macchina- Ambiente) per la prevenzione di incidenti sul lavoro

Davide Accordini, Patrizia Agnello, Silvia M. Ansaldi, Stefano Arrigoni,
Francesco Braghin, Enrico Cagno, Cinzia Frascheri, Emilia Lenzi,
Davide Piantella, Manuel Roveri, Letizia Tanca, Gabriele Viscardi

numero 23 - gennaio 2024

I *Quaderni di ricerca* dell'Inail sono lo strumento a disposizione dei ricercatori e dei ruoli professionali dell'Istituto per rendere pubblici i risultati più rilevanti delle loro attività; accolgono anche contributi di ricercatori esterni, che partecipano ai progetti dell'Inail.

Gli autori hanno la piena responsabilità delle opinioni espresse nei *Quaderni*, che non vanno intese come posizioni ufficiali dell'Inail.

I *Quaderni* sono disponibili online all'indirizzo www.inail.it.

Comitato di redazione: Biancamaria Pietrangeli, Fabrizio D'Ascenzo,
Benedetta Persechino, Patrizio Rossi

Segreteria di redazione: Cristina Francesca Giombini

Inail - Piazzale Giulio Pastore, 6 | 00144 Roma

ISBN 978-88-7484-833-1

Stampato presso la Tipografia Inail - Milano, gennaio 2024

RECKON: Tecnologie abilitanti per il monitoraggio degli elementi di contesto (Operatore-Macchina-Ambiente) per la prevenzione di incidenti sul lavoro

Davide Accordin¹, Patrizia Agnello², Silvia M. Ansaldi², Stefano Arrigoni¹, Francesco Braghin¹, Enrico Cagno¹, Cinzia Frascheri³, Emilia Lenzi¹, Davide Piantella¹, Manuel Roveri¹, Letizia Tanca¹, Gabriele Viscardi¹

Sommario - In questo Quaderno della Ricerca viene descritto il progetto RECKON (REAL-world, Context-aware KnOWledge-based lab), finanziato da Inail nell'ambito dei bandi di ricerca in collaborazione (BRiC 2018)⁴. Il progetto che si concentra sulle tecnologie di interconnessione e sulle tecniche di analisi dei dati sensibili al contesto per migliorare la sicurezza nei luoghi di lavoro, con l'obiettivo ultimo di identificare e prevenire situazioni pericolose prima che si verifichino incidenti. In RECKON, la prevenzione è interpretata attraverso le più recenti tecniche di monitoraggio, diagnostica e prognostica dal punto di vista della sicurezza, che consentono di raccogliere e utilizzare, anche in tempo reale, una grande quantità di dati sull'intero contesto operativo. Mediante reti di sensori, il sistema RECKON raccoglie informazioni che vengono utilizzate in due modi: (i) quando si verifica una situazione potenzialmente pericolosa, il sistema emette un allarme per prevenire un incidente e (ii) ogni volta che si verifica un incidente, o un cosiddetto "near-miss" (vale a dire, un potenziale incidente evitato per pochissimo), le relative informazioni sono memorizzate in un rapporto generato automaticamente e successivamente utilizzato per aggiornare le politiche antinfortunistiche. Questo quaderno descrive il quadro operativo di RECKON, insieme ai suoi moduli e alla loro interazione.

Parole chiave: Sicurezza sul lavoro, Industria 4.0, Sensorizzazione ambiente di lavoro, Integrazione dati, Analisi dati, Context awareness

¹ Politecnico di Milano

² Inail-Dipartimento di Innovazioni Tecnologiche

³ IAL (Istituto Addestramento Lavoratori)

⁴ Questo Quaderno riporta i contenuti del progetto RECKON risultato vincitore del bando Inail BRiC 2018 con ID09

Responsabile scientifico: Letizia Tanca del Dipartimento di Elettronica, Informazione e Bioingegneria - Politecnico di Milano

Indice degli argomenti

Premessa	7
Acronimi e definizioni	8
1 Individuazione dell'ambito applicativo	8
1.1 Individuazione contesti critici	9
1.2 Definizione <i>framework</i> operativo	10
1.3 Architettura generale	10
2 Un approccio basato sul contesto (<i>context-aware</i>)	12
2.1 Il modello <i>Context Dimension Tree</i>	12
2.1.1 Il modello <i>CDT_Operatore</i>	13
2.1.2 Il modello <i>CDT_Operatore, Ambiente</i>	15
2.1.3 Il modello <i>CDT_Operatore, Macchine e Utensili</i>	15
2.1.4 Il modello <i>CDT_Operatore, Infortunio</i>	16
2.1.5 Il modello <i>CDT_Operatore, NearMiss</i>	16
2.2 Introduzione al linguaggio PerLa	17
2.2.1 Istruzione Query PerLa continue	18
2.2.2 PerLa Context Language	18
2.2.3 Context Dimension Tree Declaration	19
2.2.4 Context Creation	19
2.3 Ontologie per la modellazione del contesto	25
2.4 <i>Workflow</i> per l'integrazione e l'elaborazione <i>context-aware</i> dei dati	27
3 RECKONition: l'intelligenza artificiale per l'analisi della dinamica degli infortuni	31
3.1 Definizione del processo <i>decision-making</i>	31
3.1.1 Kmedoids	32
3.2 Association-Rule Mining	34
3.3 <i>Clustering</i> del linguaggio naturale	36
3.3.1 <i>Clustering</i> basato su TAG e occorrenze	36
3.3.2 <i>Clustering</i> basato su <i>Transformers</i>	38
3.4 Language model	39
4 Dall'architettura al sistema	40
4.1 Sensorizzazione del contesto	40
4.2 Sensoristica per il monitoraggio di macchinari pericolosi	41
4.3 Progettazione della sensoristica	43
4.4 Sistemi di localizzazione INDOOR	44
4.4.1 Tecnologie per sistemi di localizzazione	44
4.5 Sviluppo soluzioni real-time per situazioni a rischio	48
4.5.1 Metodologie di analisi dati per derivare informazioni sul contesto	48
4.5.2 Progettazione di sistemi di allarme real-time	50
4.6 Sensorizzazione del contesto nei Case Labs	51
4.7 Descrizione del sistema di localizzazione indoor fornito da NewEN.	53
5 Casi Studio e sperimentazione	54
5.1 L'implementazione di RECKON nelle singole aziende	55

5.2	Sensorizzazione di un Case Lab	56
5.2.1	Osservazioni	58
5.3	Implementazione delle soluzioni nei Case Labs	58
5.4	Verifica delle soluzioni real-time nei Case Labs	59
5.4.1	Osservazioni	59
5.5	Approfondimento tecnologico: WebApp e casi di studio	59
5.5.1	Esplorazione dei dati sensoristici e implementazione query PerLa	60
6	Conclusioni	65
	Ringraziamenti	66
	Riferimenti bibliografici	67

Premessa

Gli infortuni sul lavoro sono da tempo oggetto di dibattito sociale ed economico. Secondo i dati dell'Inail (Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro), solo nel 2019 si sono verificati 641.638 infortuni sul lavoro, di cui 1089 mortali [Inail-021]. Sebbene l'interesse per l'argomento sia sempre alto, l'evoluzione del contesto aziendale, in cui si verificano incidenti, rende necessario osservare e studiare continuamente questi fenomeni, in modo da essere in grado di utilizzare sistemi di prevenzione sempre più attuali ed efficaci.

Il paradigma innovativo introdotto da Industria 4.0 [Lu-017] ha portato a risultati strutturali, tecnologici, cambiamenti produttivi e organizzativi nel mondo del lavoro [LBH-015], tutti basati su un alto sviluppo tecnologico volto ad aumentare la produttività e la competitività delle imprese sul mercato [Da-012]. L'utilizzo di sensori innovativi, ad esempio, consente di analizzare grandi quantità di dati estratti direttamente dal contesto produttivo. Tuttavia, mentre c'è stato un fortissimo interesse per la digitalizzazione dell'ambiente di lavoro, questo interesse è stato quasi interamente incanalato nella dimensione delle macchine e delle loro interconnessioni con l'ambiente circostante, lasciando sullo sfondo l'aspetto più importante ovvero la presenza del lavoratore [FPT-020]. Mettere l'operatore al centro dell'interesse delle scelte tecnologiche può invece rivelarsi fondamentale, non solo in termini di profitti e competitività, ma anche per fornire supporto all'operatore in termini di prevenzione in materia di salute e sicurezza sul lavoro [Pa-014].

È quindi evidente che i punti cruciali del problema sono un'analisi accurata e aggiornata sia del contesto lavorativo sia di come l'operatore interagisce con esso, e un uso efficiente delle informazioni derivanti da questa analisi. Tale è lo scopo del sistema RECKON, pensato per supportare le aziende nel monitoraggio e nella prevenzione degli infortuni sul lavoro.

RECKON sfrutta l'integrazione dell'analisi storica con la concettualizzazione e la conseguente sensorizzazione del contesto lavorativo, consentendo di evidenziare la corrispondenza tra una situazione che è già stata individuata come potenzialmente pericolosa e l'attuale effettivo funzionamento del sistema. Più precisamente, il sistema può considerarsi composto da tre parti concettuali: (i) sensibilizzazione delle aziende in esame (nel caso di studio attuale, imprese metallurgiche); (ii) elaborazione e concettualizzazione di flussi di dati provenienti da sensori integrati e sensibili al contesto; (iii) integrazione dei dati storici raccolti nel database degli eventi critici con altri dataset sugli infortuni (sia interni che esterni all'Inail).

Grazie alle ultime due attività, è possibile costruire una base di conoscenza per le successive analisi, al fine di poter migliorare continuamente il sistema di monitoraggio context-aware.

Acronimi e definizioni

ADAM - Adaptive Moment Estimation
AoA - Angolo di arrivo
AoD - Angolo di partenza
API - Associazione Piccole e Medie Industrie
AR - Association Rules
BERT - Bidirectional Encoder Representations from Transformers
BLE - Bluetooth Low Energy
CDT - Context Dimension Tree
DPI - Dispositivi di Protezione Individuali
FIS - Frequent Item Set
ICD - International Classification of Diseases
IMU - Unità di Misura Inerziale
inFIS - infrequent Itemset inFIS
IoT - Internet of Things
IPCA - Incremental Principal Component Analysis
LLQ - Low Level Query
LM - Language Model
LSTM - Long Short-Term Memory
NAR - Negative Association Rule
NLP - Natural Language Process
OHS - Operational Health and Safety
PAR - Positive Association Rule
PMI - imprese di piccole e medie dimensioni
ReLU - Rectified Linear Units
RFID - Radio-FrequencyID
RSSI - Received Signal Strength Indication
RTLS - Real-Time high-precision indoor Location
SNOMED - Systematized Nomenclature of MEDicine
TDoA - Time Difference of Arrival
TOF - Time of Flight
TWR - Two Way Ranging
TF_IDF - Time Frequency - Inverse Document Frequency
UMLS - Unified Medical Language System
WLAN - Wireless Local Area Network
XML - eXtensible Markup Language

1. Individuazione dell'ambito applicativo

L'attività preliminare del progetto RECKON è stata quella di studiare ed analizzare i dati relativi agli infortuni sul lavoro in ambito industriale, con l'obiettivo di progettare e sviluppare il *framework* operativo pensato per il monitoraggio, la diagnostica e la prevenzione integrata del rischio.

1.1 Individuazione contesti critici

Contestualizzando l'analisi dei dati sugli infortuni in industria, una prima criticità emerge considerando le imprese di piccole e medie dimensioni (PMI), definite secondo i dettami della Commissione Europea [EC-003]. Come testimoniato da Fabiano et al. [FCP-004], è evidente un aumento della frequenza di incidenti fatali al diminuire delle dimensioni aziendali, indipendentemente dal settore, che rende la salute dei dipendenti delle PMI più a rischio rispetto alla controparte impiegata in aziende di grandi dimensioni. Tale fenomeno, qui definito con il termine di *safety divide*, è confermato dai risultati ottenuti da Walters e Wadsworth [WW-016] limitatamente a micro e piccole imprese. Tale criticità assume una prospettiva di primo piano se analizzata in funzione della numerosità e del ruolo che le PMI rivestono in Italia, dove rappresentano la spina dorsale dell'economia nazionale, nonché il 99,7% del totale delle aziende considerando il macro settore industriale in senso stretto, responsabili del 61,5% di valore aggiunto tramite l'impiego del 74,3% dei dipendenti [Istat-019].

Parallelamente all'alta frequenza e alla gravità degli incidenti, una seconda criticità che caratterizza le PMI è la minore capacità di adottare soluzioni innovative atte a prevenire e scongiurare situazioni di pericolo. Stando alla passata letteratura, la spiegazione di questo fenomeno va ricercata in fattori quali la posizione economica delle PMI, spesso caratterizzate da risorse limitate che vengono concentrate nella sola area di processo dell'azienda per poterne garantire l'esistenza, o nella mancanza di conoscenze e competenze internalizzate, giustificata spesso dalla mancanza di un ufficio preposto ai soli scopi di salute e sicurezza [FCP-004] [WW-016]. A tal riguardo si parla di *digital divide* basato sulle dimensioni aziendali [So-015].

Seguendo questo trend, con la penetrazione di Industria 4.0 nel panorama industriale, le aziende di grandi dimensioni riusciranno a raggiungere notevoli miglioramenti in ambito salute e sicurezza sul lavoro. Al contrario, le piccole aziende, più lente e meno inclini all'adozione di soluzioni tecnologiche innovative, non potranno usufruire a pieno di questi benefici [So-015]. Pertanto, come conseguenza diretta del *digital divide* si avrà un inasprimento del *safety divide* a discapito delle PMI. Questo avrà un effetto anche in termini di produttività, dal momento che migliorie in termini di OHS (Operational Health and Safety) migliorano indirettamente la produttività delle aziende [BBS-018].

Spostando l'analisi dei dati sugli infortuni, dalla dimensione aziendale all'area industriale di appartenenza, maggiori criticità sono evidenti nel macro settore manifatturiero responsabile di circa il 14% degli incidenti mortali, legati alla sfera lavorativa in Europa, che risulta essere il valore più alto se si escludono i settori cantieristico (20,6%) e del trasporto e immagazzinamento (17,8%) [ES-019]. Tuttavia, quando si tratta di incidenti gravi ma non mortali, il settore manifatturiero ha un triste primato, con il 19% dei casi totali [ES-019]. Contestualizzando il problema nella penisola, il bilancio si aggrava, con un numero di incidenti mortali maggiore rispetto alla media Europea. In particolare, il settore manifatturiero risulta responsabile del 15% dei casi di incidenti mortali, percentuale che lo colloca al primo posto rispetto a qualsiasi altro settore, seguito da vicino dalla cantieristica.

La vulnerabilità del manifatturiero risulta ancora più manifesta se letta dalla prospettiva del suo peso economico per il paese: nel 2018 infatti essa ha generato il 16,7% del valore aggiunto per l'economia italiana, valore che ha registrato la crescita più alta di qualsiasi altro settore nel medesimo arco temporale [CA-018]. Due milioni di persone sono impiegate nelle mi-

cro e piccole imprese manifatturiere italiane, il numero più alto tra i paesi europei [CA-018], che diventano quasi tre milioni aggiungendo nell'equazione le imprese di medie dimensioni [Istat-020].

1.2 Definizione framework operativo

Seguendo la logica della criticità in termini di sicurezza sul lavoro unita al peso economico, il progetto RECKON è diretto alle PMI Italiane che operano nel settore metalmeccanico, con l'obiettivo di sviluppare e testare soluzioni tecnologiche (e/o sistemi di soluzioni) che possano meglio controllare le situazioni di pericolo più frequenti in azienda e le loro cause, nonché comprendere gli stati e le effettive interazioni operatore-macchina-ambiente (sintetizzati nel termine "contesto") per le attività di monitoraggio, diagnostica e prevenzione gestite in modo integrato.

Una volta definiti i contesti aziendali nei quali agire (Case-Labs), si è passati alla selezione dei sensori appropriati, allo sviluppo di soluzioni tecnologiche adeguate e alla conseguente sensorizzazione del contesto operativo (macchine, operatori e ambiente). In parallelo all'implementazione della sensoristica e dell'infrastruttura hardware e software che ne sta alla base, il progetto RECKON ha sviluppato la connessione in rete dei sensori e delle altre sorgenti informative (hardware) per la condivisione e integrazione *context-aware* dei dati.

Tale sistema di *data integration context-aware* introduce il concetto di interconnessione a livello sistemico, che porta a vedere le soluzioni tecnologiche; pertanto i Case Labs all'interno dei quali sono implementate, non sono più considerate come entità a sé stanti ma come unità interagenti a livello sistemico tra di loro e con banche dati (Data-Labs) esterne o aziendali.

Lo sviluppo di questo *framework* operativo ha offerto alcuni vantaggi al progetto, quali:

- la capacità di poter sfruttare pienamente il notevole quantitativo di dati relativi agli infortuni già occorsi, comprendendone i fattori e le dinamiche determinanti, al fine di identificare, con maggior efficacia, i dati da acquisire in azienda (Case Labs) e progettare soluzioni tecnologiche, da installare in azienda, che permettano di controllare in maniera più efficace le situazioni di pericolo più frequenti;
- la capacità di poter verificare non solo la bontà delle soluzioni installate, in termini di risultati ottenuti, ma anche il loro livello di adottabilità per l'ampio ambito delle PMI metalmeccaniche in Italia;
- La capacità di poter generare informazione addizionale, essenzialmente da near-miss e/o trend di comportamenti, relativa a nuovi fattori e nuove dinamiche rilevati nelle aziende, facendola ricircolare per migliorare l'identificazione dei dati da acquisire e la progettazione delle soluzioni tecnologiche da installare, chiudendo così il virtuoso cerchio della conoscenza.

1.3 Architettura generale

L'architettura di sistema introdotta dal progetto RECKON è organizzata attraverso un approccio distribuito e gerarchico [TM-007] in grado di operare sia direttamente nelle aziende (per minimizzare il tempo di latenza prima di prendere una decisione o di attivare un allarme) sia a livello di Cloud [Di-009] (per integrare informazioni provenienti da aziende differenti, così come da altre sorgenti di dati).

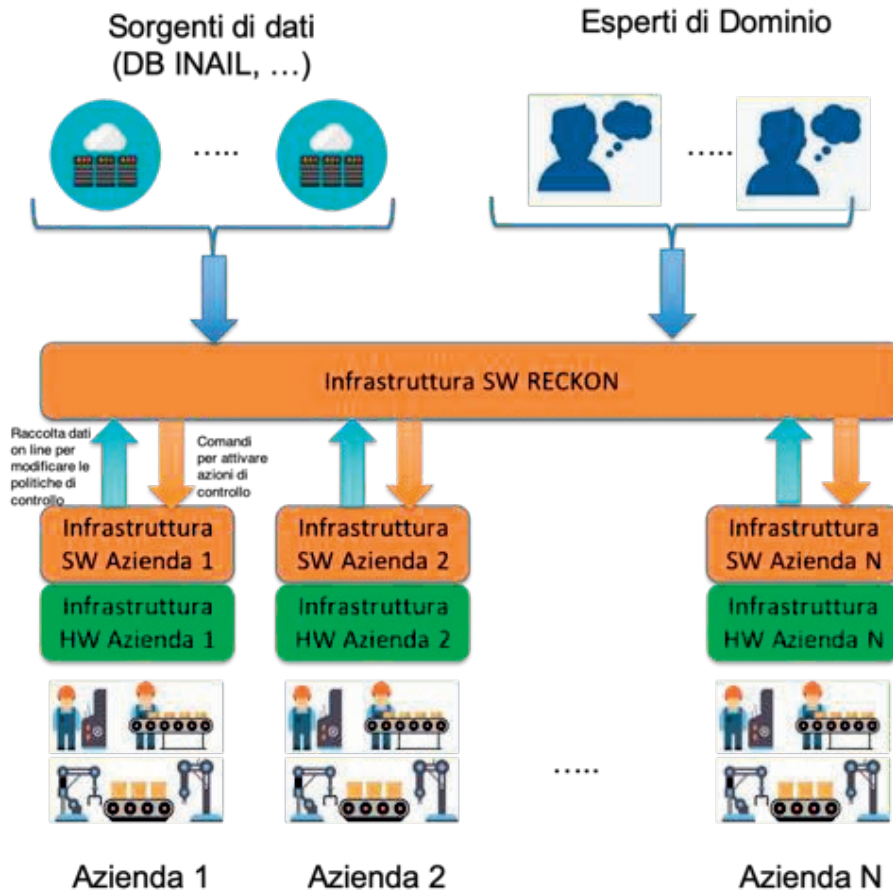


Figura 1 Architettura generale del sistema RECKON

In dettaglio, l'architettura proposta, mostrata nella Figura 1, si compone di una infrastruttura hardware e due infrastrutture software, una locale e una RECKON.

L'*infrastruttura hardware (HW) Locale* opera nell'azienda ed è composta da sensori e attuatori che sono messi in opera direttamente sul campo. I sensori possono essere specifici di una particolare azienda (ad es. un sensore per monitorare il campo magnetico in un'area critica) o possono essere generici e quindi funzionali a differenti classi di applicazioni (ad es. localizzazione di "uomo a terra" o identificazione di aree di potenziali incidenti tra lavoratori e mezzi di trasporto materiali). Una *infrastruttura software (SW) Locale* opera nell'azienda e ha lo scopo di elaborare i dati raccolti dal campo tramite l'*Infrastruttura HW Locale* per identificare prontamente situazioni (potenzialmente) critiche e prevenire infortuni al lavoratore. Aspetto critico di questa infrastruttura locale è poter elaborare i dati il più rapidamente possibile al fine di minimizzare il tempo che passa tra l'acquisizione del dato e la possibile attivazione di allerte/allarmi. La minimizzazione di tale tempo è cruciale per supportare azioni, non solo efficaci ma anche efficienti, di prevenzione di infortuni. Per questo motivo, tale elaborazione è stata pensata come collocata in locale, e non operante in *Cloud* (soluzione che non avrebbe dato garanzie sulla rapidità di risposta). Un secondo aspetto chiave dell'*Infrastruttura SW locale* è la possibilità di estrarre conoscenza ed informazioni rilevanti da quanto monitorato, facendo emergere non solo gli eventi, ma, più in generale, tutte le situazioni potenzialmente di pericolo. Questi eventi, opportunamente elaborati (eventualmente

rimuovendo informazioni per garantire la privacy di lavoratori e/o aziende), saranno poi trasmessi all'Infrastruttura SW RECKON per ulteriori analisi.

La *infrastruttura SW RECKON* ha lo scopo di integrare le informazioni provenienti dalle diverse aziende per poi definire dei piani di azione strategici, più a lungo termine rispetto alle azioni svolte dall'Infrastruttura SW Locale, in ottica di prevenzione degli infortuni. Tale infrastruttura, che può operare in Cloud dato che non ci sono vincoli operativi stringenti o real-time, sarà anche in grado di integrare ed estrarre conoscenza attraverso l'impiego di strumenti di *machine learning*, *deep learning* ed analisi semantica del testo, di sorgenti di informazioni esterne quali, ad esempio, le basi di dati Inail. Inoltre, sempre attraverso questa infrastruttura sarà possibile integrare le informazioni provenienti dagli esperti di dominio. Tale operazione sarà garantita dalla capacità di scegliere differenti configurazioni operative degli algoritmi di *learning*. L'*Infrastruttura SW RECKON* potrà quindi ricevere dati dalle aziende attraverso le *Infrastrutture SW locali* (che operano direttamente sulle aziende) e sarà in grado di inviare comandi per attivare azioni di controllo alle *Infrastrutture SW locali* per modificarne od aggiornarne le modalità di funzionamento.

2 Un approccio basato sul contesto (*context-aware*)

Il sistema RECKON si propone di gestire le situazioni pericolose in diverse piccole e medie imprese del settore metallurgico. Al variare delle aziende e delle attività lavorative che si svolgono, quindi, varia il contesto da gestire. Per questo motivo, per la progettazione del sistema e la condivisione e integrazione dei dati è stato scelto un approccio basato sul contesto. L'informazione di contesto è stata raccolta partendo dall'analisi del dominio applicativo e dei dataset degli infortuni studiati. In particolare, il focus è stato posto su ogni lavoratore che si vuole proteggere, creando così un modello di contesto operatore-centrico.

Tra i modelli disponibili all'attuale stato dell'arte, per modellare i possibili contesti è stato scelto il *Context Dimension Tree*, strumento concettuale che verrà ampiamente descritto nella sezione seguente. È stato poi utilizzato il linguaggio PerLa [Sc-017b] per esprimere le interrogazioni e i comandi a reti di sensori integrati con basi di dati. Questo linguaggio di specifica, che verrà descritto nella Sezione 2.2 permette di esprimere i contesti di interesse rendendo i comandi *context-aware*. Infine, per affrontare il problema delle diverse terminologie e ambiti di lavoro eterogenei è stata prodotta un'ontologia [DDF-003] che verrà descritta nella Sezione 2.3. L'ontologia, come vedremo, ha non solo lo scopo di rendere il *framework* il più generale possibile, ma anche di integrare il contesto nell'analisi dei dati. Al termine di questo capitolo, infatti, verrà presentato anche il sottosistema RECKON_{in}, utilizzato per l'interpretazione dei dati testuali e rilasciato a Inail al termine del progetto.

2.1 Il modello *Context Dimension Tree*

Requisito di fondamentale importanza per la progettazione, lo sviluppo e la messa in opera di sistemi *context-aware* è la capacità di “modellare e comprendere l'informazione di contesto” [Ab-999]. La letteratura propone diversi modelli e metodologie per la modellazione del contesto; nel sistema RECKON viene utilizzato il *Context Dimension Tree* (CDT) [Bo-009, SP-017] che fornisce una rappresentazione intuitiva del contesto basata su una struttura ad albero, i cui nodi hanno il seguente significato:

- *Dimensioni* (nodi neri): rappresentano le caratteristiche rilevanti del contesto;
- *Nodi concetto* (nodi bianchi): rappresentano i valori ammissibili che possono essere assunti da ogni dimensione;
- *Attributi* (nodi quadrati): sono parametri il cui valore può essere ricavato dinamicamente dall'ambiente (ad esempio mediante sensori) oppure fornito direttamente dall'utente al momento dell'esecuzione. Gli attributi sono utilizzati per rappresentare parametri con un numero elevato di valori possibili (es. le coordinate GPS di un dispositivo mobile, nel caso la dimensione sia la posizione).

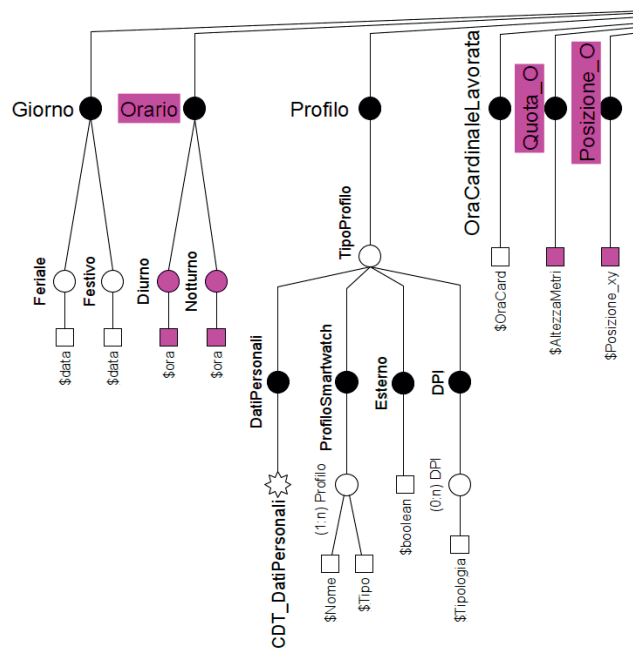


Figura 2 - CDT declaration

Un contesto, in questo modello, è formato da un insieme di nodi bianchi ed eventuali attributi: per esempio la Figura 2 rappresenta, una volta che i nodi attributo siano valorizzati, la posizione di un lavoratore in termini delle dimensioni *Orario*, *Quota* (*z*) e *Posizione* (*x,y*).

Questo processo viene definito nella fase di progettazione del sistema, che comprende due attività: la creazione del modello del contesto, che descrive le situazioni possibili nell'ambiente da monitorare - e cioè l'insieme di tutte le strutture come quella di Figura 3, che invece presenta un contesto specifico - e la mappatura tra tali situazioni e (i) le eventuali azioni da attivare e (ii) i dati finali che devono essere forniti all'utente.

2.1.1 Il modello CDT_Operatore

Il contesto applicativo operatore-macchina-ambiente viene rappresentato dall'albero di contesto *CDT_Operatore* (riportato in Figura 3) e ha l'obiettivo di fornire una visione d'insieme delle componenti del contesto e definire, per ogni contesto, le caratteristiche che determinano l'esposizione al rischio d'infortunio di un lavoratore.

Il *CDT_Operatore* è un modello operatore-centrico: ad ogni lavoratore è associato un *CDT_Operatore* che rappresenta il contesto lavorativo in cui il lavoratore opera. Il contesto è inteso sia come ambiente fisico: l'ambiente di lavoro e le sue caratteristiche, gli oggetti con cui ogni lavoratore interagisce (macchinari, utensili, ecc); sia, più in generale, come l'insieme delle informazioni che caratterizzano la situazione del lavoratore quali, ad esempio, la sua mansione o alcuni suoi dati personali (età, sesso, cittadinanza, ecc.) utili anche per analizzare i dati raccolti.

In accordo con la letteratura, alcune dimensioni dell'albero di contesto sono comuni alla maggior parte delle applicazioni, ad esempio: il tempo, il luogo, il ruolo del soggetto che sta utilizzando il sistema, la specifica fase operativa del sistema, e l'interfaccia, cioè il mezzo attraverso il quale le informazioni vengono fornite all'utente. Il *CDT_Operatore* presenta queste dimensioni generali declinate nel contesto operatore-macchina-ambiente e nel dominio applicativo delle PMI metalmeccaniche.

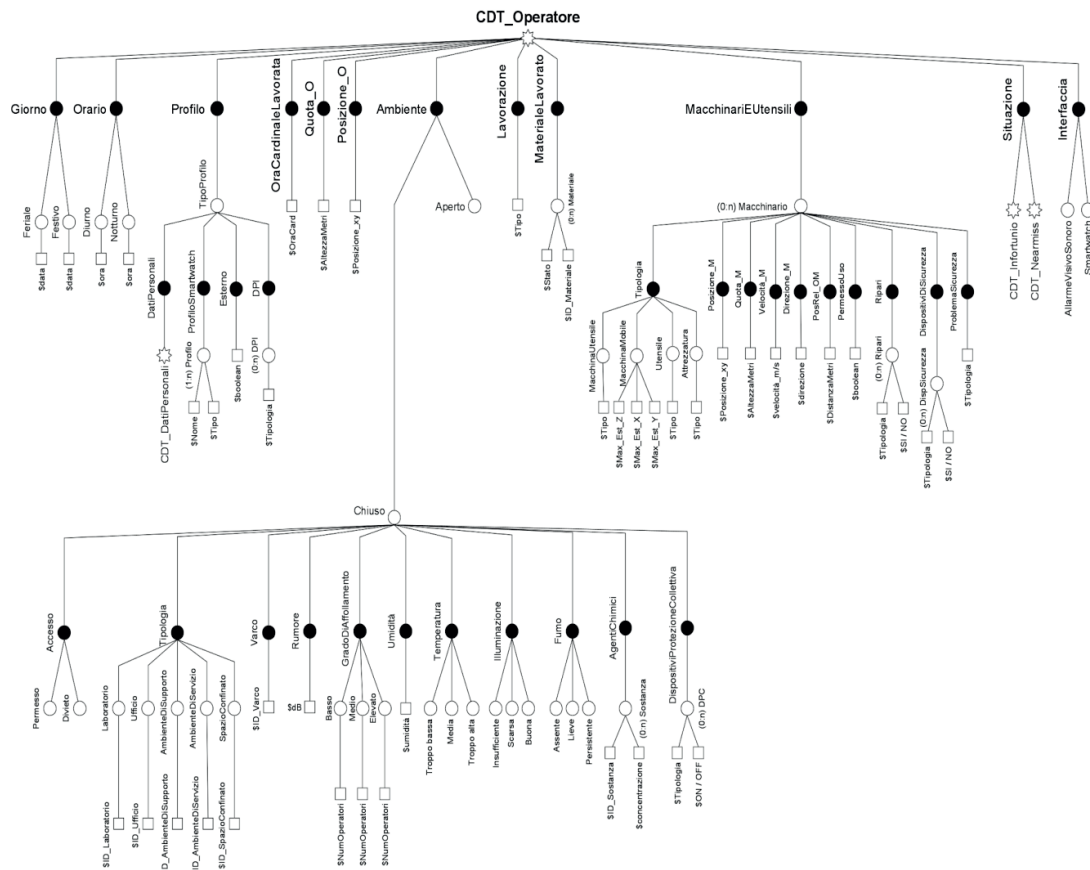


Figura 3 - CDT operatore per la rappresentazione dei contesti

Si noti che le variabili del contesto misurabili dal sistema proposto per l'installazione saranno un sottoinsieme delle dimensioni dell'albero presentato, che è concepito per modellare una situazione molto generale. I valori per le altre dimensioni potranno essere desunti da informazioni già note, come dettagli sulle macchine o sullo spazio - specifiche dell'azienda e della sua configurazione - oppure, se necessario, da altri dispositivi - come il termometro o l'igrometro - già presenti in azienda.

La specifica del modello di contesto permette di rilevare, come sua istanza, ogni contesto che si potrebbe verificare durante l'esecuzione, guidando le azioni del sistema, la raccolta e l'interpretazione dei dati in accordo con esso.

2.1.2 Il modello CDT_Operatore, Ambiente

La Figura 4 illustra il modello CDT relativo ad un ambiente chiuso. La dimensione Accesso indica se si tratta di un ambiente ad accesso limitato e definisce, per ogni tipo di profilo dell'operatore il permesso o il divieto di accedere nell'area. La dimensione Tipologia rappresenta una classificazione di ambienti comunemente usata in ambito industriale (laboratorio, ufficio, ambiente di supporto, di servizio, spazio confinato). La dimensione Varco indica la presenza di un varco virtuale, all'ingresso di un'area ad accesso limitato, tramite il quale è possibile rilevare l'accesso degli operatori nell'ambiente. Le altre dimensioni corrispondono a parametri ambientali (quali: Rumore, Umidità, Temperatura,...) utili per monitorare la salubrità dell'ambiente per la salute lavoratori. Infine, la dimensione Dispositivi di Protezione Collettiva (DPC) rappresenta l'insieme di quei dispositivi di protezione collettiva che, da normativa, devono essere presenti all'interno dello specifico ambiente chiuso.

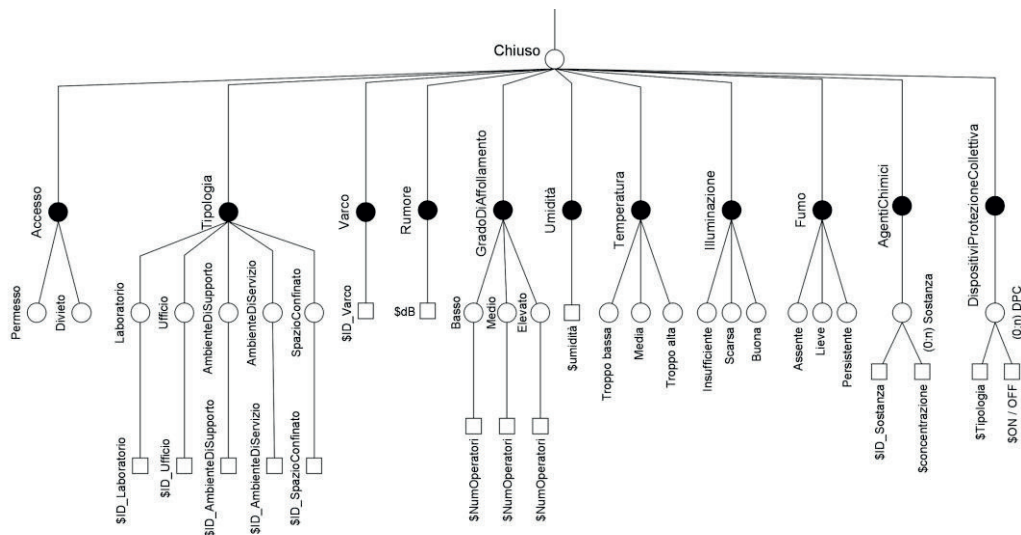


Figura 4 - CDT operatore per la rappresentazione di un ambiente chiuso

2.1.3 Il modello CDT_Operatore, Macchine e Utensili

La dimensione Macchine e Utensili del CDT_Operatore, illustrata in Figura 5, rappresenta l'insieme delle macchine, attrezzature e utensili presenti all'interno dell'ambiente lavorativo e con cui l'operatore interagisce. La dimensione Tipologia individua il tipo di oggetto rilevato in prossimità dell'operatore, quali: macchine utensili o fisse, macchine mobili (es: muletto), attrezzature di lavoro (scale, trabattelli,...), utensili (martello, chiave inglese,...). Seguono alcuni parametri di localizzazione (posizione, quota) e di dinamica (velocità, direzione) del macchinario, di posizione relativa lavoratore-macchina, di sicurezza (ripari, dispositivi specifici, problemi di sicurezza).

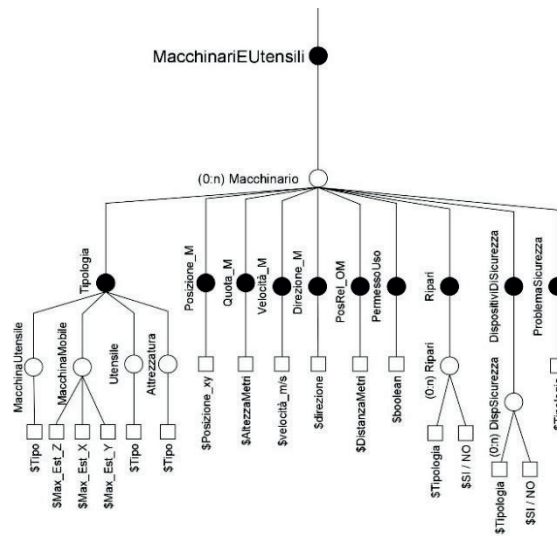


Figura 5 - CDT operatore per la rappresentazione di macchinari e utensili

2.1.4 Il modello CDT_Operatore, Infortunio

Al verificarsi di un evento incidentale, il modello propone i dati acquisiti relativi al contesto operatore-macchina-ambiente, oltre ai dati che descrivono più in dettaglio l'infortunio (vedi Figura 6), la situazione (attività svolta, DPI indossati) e le conseguenze riportate.

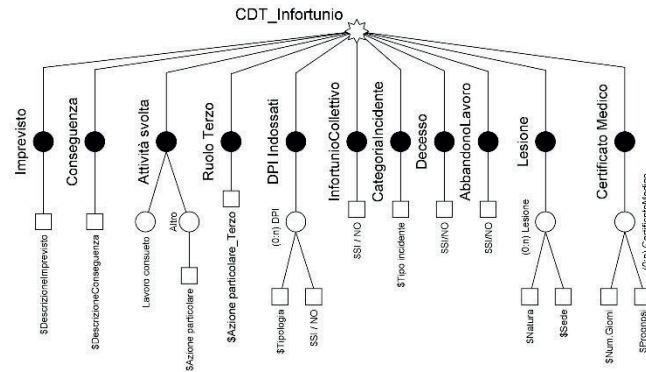


Figura 6 - CDT operatore per la rappresentazione di un infortunio

2.1.5 Il modello CDT_Operatore, NearMiss

Il modello relativo al near miss, come illustrato in Figura 7, riporta le dimensioni di imprevisto, attività svolta, ruolo terzo (cioè il lavoratore coinvolto nell'evento non è quello a rischio di infortunio), DPI indossati, categoria di rischio a cui il lavoratore è stato esposto.

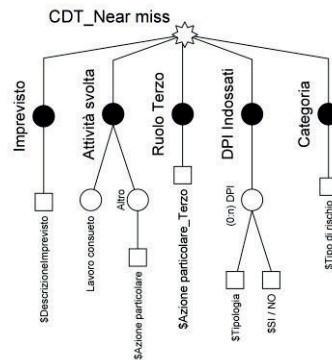


Figura 7 - CDT operatore per la rappresentazione di un near miss

2.2 Introduzione al linguaggio PerLa

La possibilità di specificare una struttura, quale il *Context Dimension Tree*, mediante linguaggi standard di rappresentazione di informazioni semi-strutturate usati in varie applicazioni, come ad esempio XML, permette di rilevare e gestire automaticamente i cambiamenti di contesto, adattando di conseguenza le azioni del sistema durante l'esecuzione. La scelta, tuttavia, è caduta sul linguaggio PerLa, sviluppato dal Politecnico di Milano. PerLa [SP-011] è un *framework* per configurare e gestire i moderni sistemi pervasivi e, in particolare, le reti di sensori wireless. Lo sviluppo del sistema PerLa è indirizzato alla progettazione e realizzazione delle seguenti caratteristiche:

- vista data-centrica dei sistemi pervasivi;
- interfaccia omogenea ad alto livello per i dispositivi (più in generale sorgenti di dati) eterogenei;
- supporto per reti altamente dinamiche (ad esempio, reti di sensori wireless);
- minimo sforzo di codifica per l'aggiunta di nuovi dispositivi.

I primi due obiettivi sono raggiunti utilizzando una tecnica già nota in letteratura: gestire il sistema pervasivo come un database. Il risultato di questo approccio è la possibilità di accedere a tutti i dati generati dalla rete di sensori tramite una query PerLa che permette agli utenti finali e alle applicazioni di alto livello di raccogliere ed elaborare informazioni senza alcuna conoscenza del sottostante sistema pervasivo. Il linguaggio PerLa è un linguaggio dichiarativo simile al linguaggio SQL.

L'altro componente fondamentale del sistema PerLa è il PerLa Middleware, la cui architettura espone due interfacce principali: un'interfaccia di alto livello, che permette l'inserimento di *query*, e un'interfaccia di basso livello che fornisce meccanismi *Plug&Play* per aggiungere continuamente nuovi dispositivi e supportare il risparmio energetico.

Il PerLa Middleware ha il ruolo fondamentale di nascondere agli utenti i dettagli tecnici sull'interazione con i nodi di rete e le diverse sorgenti di dati (caratteristiche hardware e software, protocolli di comunicazione, capacità di calcolo, ecc.), permettendo di concentrarsi sul significato applicativo delle operazioni; inoltre fornisce una grande scalabilità, in termini sia di numero sia di tipologie di nodi.

Nell'ambito del progetto, PerLa è stato scelto puramente come linguaggio di specifica per definire le modalità di interrogazione della base di dati sensoristici e di integrazione delle fonti eterogenee di dati.

Infatti, grazie alla componente del linguaggio PerLa dedicata al contesto, è anche possibile specificare il modello di contesto definito nel paragrafo precedente e definirne le modalità di gestione, quali l'attivazione dei contesti critici e il rispettivo comportamento del sistema (lancio di *query* specifiche di contesto, generazione di allarmi, ecc.).

2.2.1 L'istruzione Query PerLa continue

Le *query continue* sono eseguite in continuo sull'intera base di dati sensoristici e sull'insieme delle sorgenti di dati a disposizione e il risultato ottenuto permette di aggiornare in tempo reale i valori dei parametri e delle dimensioni del contesto definiti nel *Context Dimension Tree*.

Si riportano più sotto due esempi di *Low Level Query* (del tipo CREATE OUTPUT STREAM) che forniscono in output uno *stream* continuo di dati per raccogliere il flusso di dati provenienti dai sensori. La prima *query* avvia un'operazione di campionamento della posizione del lavoratore su tutti i dispositivi, ad esempio di tipo *smartwatch*, indossati dai lavoratori. La seconda invece avvia un'operazione di campionamento della posizione del macchinario su tutti i dispositivi installati sulle macchine mobili. In entrambe le *query* le letture dei dati sono raccolte ogni 10 secondi, come specificato nella clausola SAMPLING, e queste vengono sintetizzate in nuovi record di output ogni 30 secondi, come indicato nella dichiarazione EVERY che gestisce l'intera sezione di raccolta dei dati.

```
CREATE OUTPUT STREAM PosizioneBraccialeto (id_tag STRING, ts STRING, x
FLOAT, y FLOAT, z FLOAT) AS:
EVERY 30 seconds
SELECT id_tag, ts, x, y, z
SAMPLING EVERY 10 seconds
EXECUTE IF EXISTS id_tag, x, y, z
```

```
CREATE OUTPUT STREAM PosizioneMacchinario (id_macchinario STRING, ts
STRING, x FLOAT, y FLOAT, z FLOAT) AS:
EVERY 30 seconds
SELECT id_macchinario, ts, x, y, z
SAMPLING EVERY 10 seconds
EXECUTE IF EXISTS id_macchinario, x, y, z
```

2.2.2 PerLa Context Language

La sintassi del *PerLa Context Language* [SP-017, Sc-011a, Sc-011b] è composta da due parti: la prima (*CDT Declaration*) permette di dichiarare il modello di contesto; la seconda (*Context creation*) consente di creare specifici contesti sulla base del CDT definito.

2.2.3 Context Dimension Tree Declaration

In questa parte vengono riportati alcuni estratti della specifica del modello di contesto CDT. Le clausole CREATE DIMENSION, CREATE CONCEPT e CREATE ATTRIBUTE consentono di dichiarare i nodi dell'albero associando a ciascun nodo il suo tipo e il suo nome. Dove è presente la clausola EVALUATED ON, il valore associato all'attributo sarà pari al risultato della query contenuta nella clausola EVALUATED ON.

Di seguito, viene riportata la sintassi per la creazione delle dimensioni *Quota_O* (la quota dell'operatore), analogamente *Posizione_O*, e *Orario* (descritte nella sezione precedente e riportate in Figura 2):

```
CREATE DIMENSION Quota_O
  CREATE ATTRIBUTE $MetriQuota
  EVALUATED ON:
  SELECT z
  FROM PosizioneBraccialetto
CREATE DIMENSION Posizione_O
  CREATE ATTRIBUTE $Posizione_xy
  EVALUATED ON:
  SELECT x, y
  FROM PosizioneBraccialetto
CREATE DIMENSION Orario
  CREATE CONCEPT Diurno
  CREATE ATTRIBUTE ora
  EVALUATED ON:
  SELECT ora
  FROM Timestamp
  WHERE ora > 5:00 AND ora < 23:00
CREATE CONCEPT Notturmo
  CREATE ATTRIBUTE ora
  EVALUATED ON:
  SELECT ora
  FROM Timestamp
  WHERE ora > 23:00 AND ora < 5:00
```

2.2.4 Context Creation

Utilizzando il linguaggio PerLa è possibile dichiarare un contesto basato su un CDT predefinito, gestirne l'attivazione e definire le azioni conseguenti del sistema.

L'attivazione del contesto dipende dal verificarsi della condizione contenuta nella clausola ACTIVE IF; tale condizione viene definita sui parametri e le dimensioni del contesto considerando alcuni valori-soglia ritenuti critici. Per ogni parametro, il valore critico viene confrontato con il risultato ottenuto dalle *query* continue.

All'interno della clausola ON ENABLE si definiscono le azioni che il sistema deve compiere quando il contesto è attivo; esempi di azioni sono: *query* specifiche da inviare alla base di dati per ottenere le informazioni utili all'operatore che si trova nella situazione di rischio, attivazione di segnali di allarme, memorizzazione di dati utili, ecc.

Di seguito si riportano alcuni esempi di *query* specifiche di contesto per modellare alcuni rischi, in verde sono evidenziati i parametri e le dimensioni contenute nella clausola ACTIVE IF.

2.2.4.1 Rischio caduta dall'alto dell'operatore

Il contesto *CadutaDallAlto_Operatore* (illustrato in Figura 8) è attivo in corrispondenza della seguente configurazione: un operatore (Profilo EXISTS) che non è abilitato ad effettuare lavori in

quota (ProfiloSmartwatch_Name IS NOT 'Lavoro in quota') si trova ad una quota maggiore del valore soglia pari a 2 m (Quota_O > 2m).

```
CREATE CONTEXT CadutaDallAlto_Operatore
ACTIVE IF Profilo EXISTS AND
    Quota_O > 2m AND
    ProfiloSmartwatch_Nome IS NOT 'Lavoro in quota'

ON ENABLE (CadutaDallAlto_Operatore):
    SELECT ProfiloSmartwatch, Quota_O, Ambiente_Tipologia,
        Ambiente_Accesso,Macchinario_Tipologia,
        Macchinario_PermissioneUso
    SET PARAMETER 'warning message' = TRUE
    ACTIVATE ALARM

ON DISABLE
    DROP CadutaDallAlto_Operatore
    INSERT RECORD INTO CDT_Nearmiss
    SET PARAMETER 'warning message' = FALSE
REFRESH EVERY 5m
```

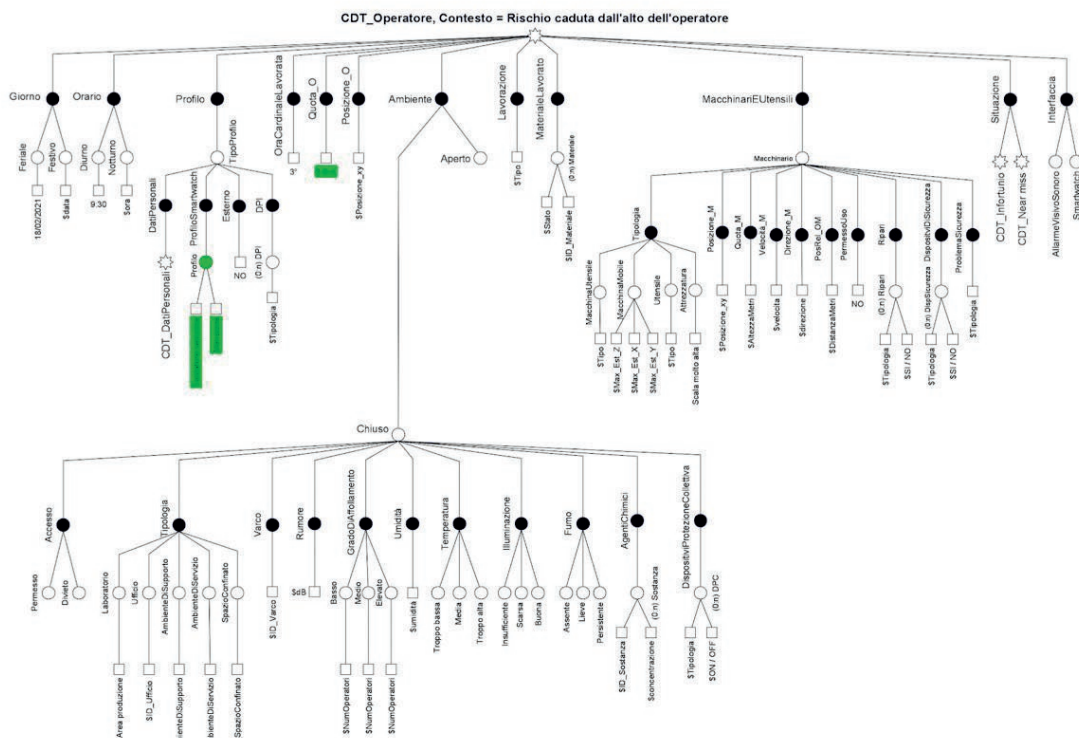


Figura 8 - Contesto: rischio caduta dall'alto dell'operatore

2.2.4.2 Posizione relativa dell'operatore con una macchina utensile

Il contesto *Operatore_MacchinaUtensile* (illustrato in Figura 9) è attivo in corrispondenza della

seguente configurazione: un operatore (Profilo EXISTS) si trova in prossimità di una macchina utensile (Macchinario_Tipologia = 'MacchinaUtensile') e la distanza tra l'operatore e la macchina è minore della distanza minima di sicurezza ($PosRel_OM < \$minimum_safe$). Si noti che la distanza minima di sicurezza può assumere un unico valore per tutti gli operatori, oppure, può dipendere dal profilo dell'operatore e in tal caso si dovranno considerare diverse distanze minime in base ai permessi degli operatori per l'utilizzo della macchina.

```

CREATE CONTEXT Operatore_MacchinaUtensile
ACTIVE IF Profilo EXISTS AND
    Macchinario_Tipologia = 'MacchinaUtensile' AND
    PosRel_OM < $minimum_safe

ON ENABLE (Operatore_MacchinaUtensile):
    SELECT ProfiloSmartwatch, PosRel_OM, Macchinario_PermissioneUso,
        Ambiente_Tipologia, Ambiente_Accesso
    SAMPLING EVERY 1m
    SET PARAMETER 'warning message' = TRUE ACTIVATE ALARM

ON DISABLE:
    DROP Operatore_MacchinaUtensile
    INSERT RECORD INTO CDT_Nearmiss
    SET PARAMETER 'warning message' = FALSE
    REFRESH EVERY 5m
    
```

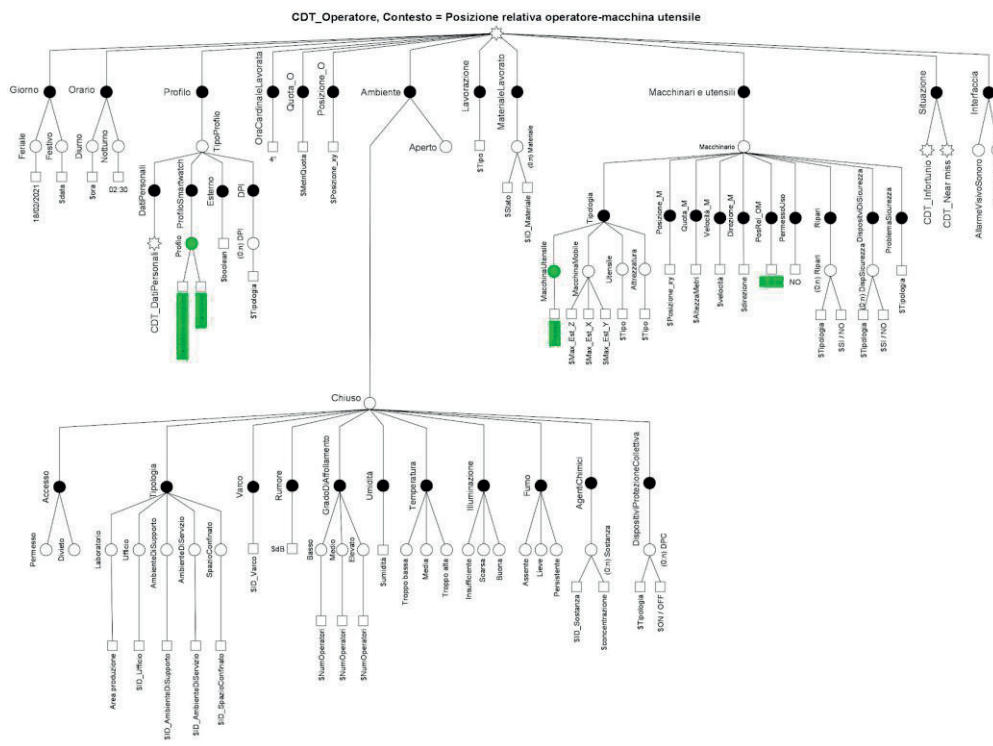


Figura 9 - Contesto: Posizione relativa operatore-macchina utensile

2.2.4.3 Posizione relativa dell'operatore con una macchina mobile

Il contesto *Operatore_MacchinaMobile* (illustrato in Figura 10) è attivo in corrispondenza della seguente è attivo in corrispondenza della seguente configurazione: un operatore (Profilo EXISTS) si trova in prossimità di una macchina mobile (Macchinario_Tipologia = 'MacchinaMobile'), la distanza tra l'operatore e la macchina è minore della distanza minima di sicurezza ($PosRel_OM < \$minimum_safe$) e la macchina è accesa/in movimento (Macchinario_Velocità_M > 0 m/s)

```
CREATE CONTEXT Operatore_MacchinaMobile
ACTIVE IF Profilo EXISTS AND
    Macchinario_Tipologia = 'MacchinaMobile' AND
    Macchinario_Velocità_M > 0 m/s AND
    PosRel_OM < $minimum_safe

ON ENABLE (Operatore_MacchinaMobile):
    SELECT ProfiloSmartwatch, PosRel_OM, Macchinario_PermessoUso,
        Macchinario_Velocità_M, Macchinario_Direzione_M,
        Ambiente_Tipologia, Ambiente_Accesso
    SAMPLING EVERY 1 min
    SET PARAMETER 'warning message' = TRUE
    ACTIVATE ALARM

ON DISABLE:
    DROP Operatore_MacchinaMobile
    INSERT RECORD INTO CDT_Nearmiss
    SET PARAMETER 'warning message' = FALSE
    REFRESH EVERY 5m
```

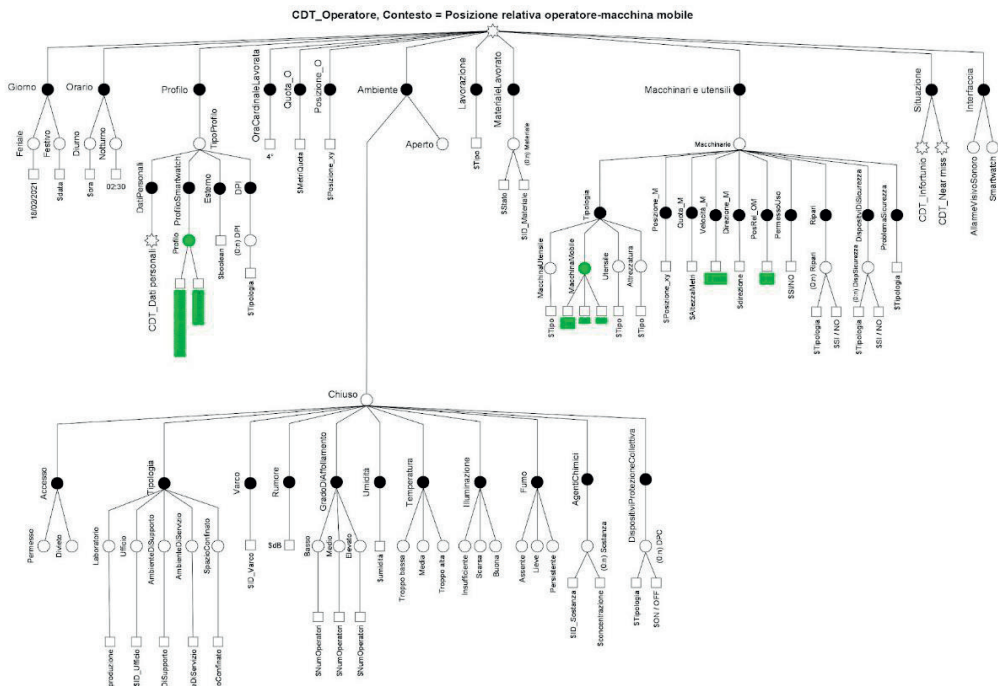


Figura 10 - Contesto: Posizione relativa operatore-macchina mobile

2.2.4.4 Accesso di un operatore in un'area ad accesso limitato

Il contesto *Varco_Operatore* (illustrato in Figura 11) è attivo in corrispondenza della seguente è attivo della seguente configurazione: un operatore (Profilo EXISTS) si trova in prossimità di un ambiente chiuso (Ambiente = 'Chiuso') ad accesso limitato e regolato da un varco (Ambiente_Varco EXISTS) e all'operatore è vietato l'accesso all'ambiente (Ambiente_Accesso = 'Divieto').

```
CREATE CONTEXT Varco_Operatore
ACTIVE IF Profilo EXISTS AND Ambiente = 'Chiuso' AND
Ambiente_Varco EXISTS AND Ambiente_Accesso = 'Divieto'

ON ENABLE (Varco_Operatore):
SELECT ProfiloSmartwatch, Posizione_O, Ambiente_Tipologia
SAMPLING EVERY 1 min
SET PARAMETER 'warning message' = TRUE
ACTIVATE ALARM

ON DISABLE:
DROP Varco_Operatore
INSERT RECORD INTO CDT_Nearmiss
SET PARAMETER 'warning' = FALSE
REFRESH EVERY 5m
```

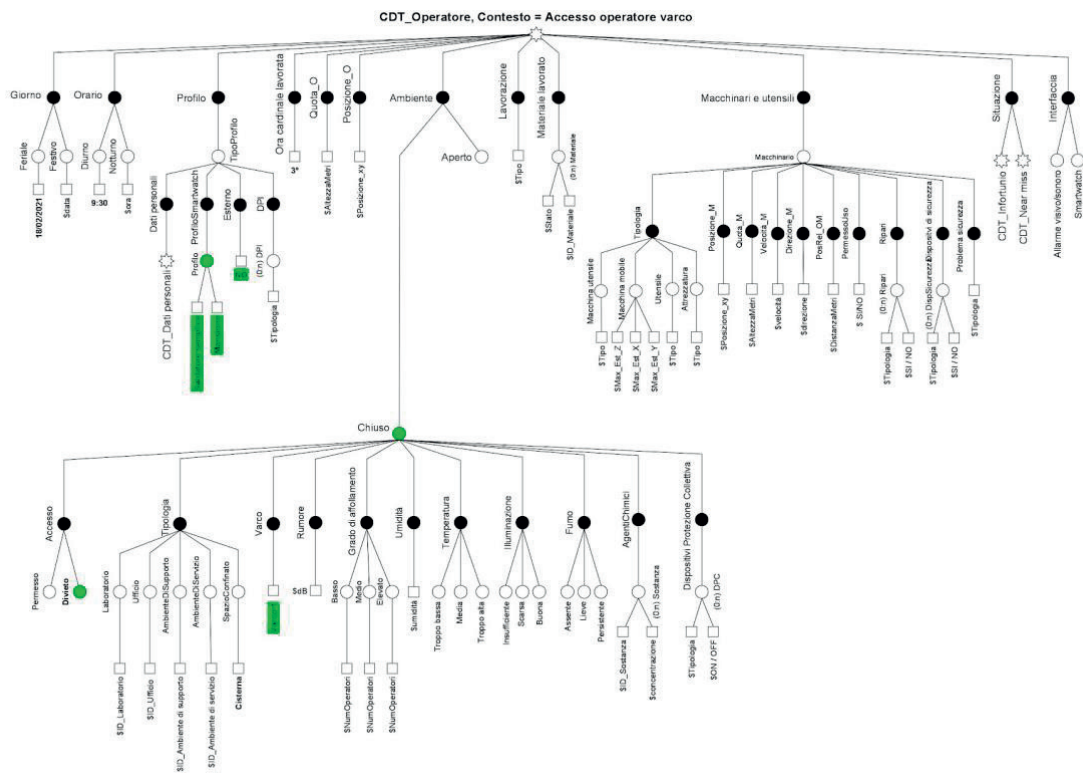


Figura 11 - Contesto: Accesso operatore in un'area ad accesso limitato

2.2.4.5 Interazione tra macchine mobili

Il contesto *MacchinaMobile_MacchinaMobile* (illustrato in Figura 12) è attivo in corrispondenza della seguente configurazione: due macchine mobili (*Macchinario1_Tipologia* = 'MacchinaMobile' AND *Macchinario2_tipologia* = 'MacchinaMobile') si trovano ad una distanza relativa minore della distanza minima ($Differenza(Posizione_M1, Posizione_M2) < \$minimum_safe$) e almeno una delle due macchine è accesa o in movimento ($Velocità_M1 > 0 \text{ m/s}$ OR $Velocità_M2 > 0 \text{ m/s}$). Si pensi al caso tipico di due muletti in movimento nella stessa area.

```

CREATE CONTEXT MacchinaMobile_MacchinaMobile
ACTIVE IF Macchinario1_Tipologia = 'Macchina mobile' AND
Macchinario2_tipologia = 'Macchina mobile' AND
Differenza (Posizione_M1, Posizione_M2) <
$minimum_safe AND ( Velocità_M1 > 0 m/s OR
Velocità_M2 > 0 m/s)

ON ENABLE (MacchinaMobile_MacchinaMobile):
SELECT Posizione_M1, Posizione_M2, Velocità_M1,
Velocità_M2, Direzione_M1, Direzione_M2
SAMPLING EVERY 1 min
SET PARAMETER 'warning' = TRUE
ACTIVATE ALARM

ON DISABLE:
DROP MacchinaMobile_MacchinaMobile
INSERT RECORD INTO CDT_Nearmiss
SET PARAMETER 'warning' = FALSE

REFRESH EVERY 5m
    
```

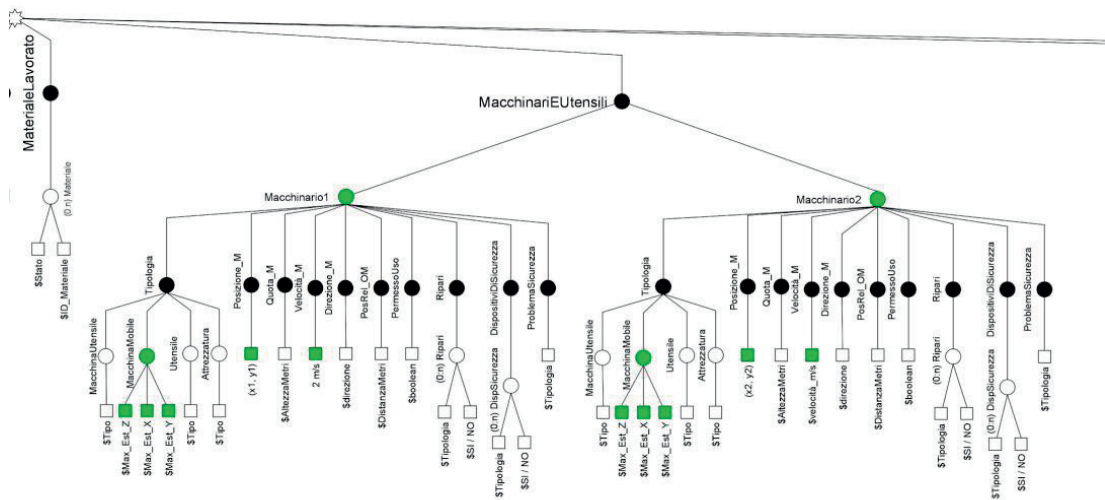


Figura 12 - Contesto: Interazione tra macchine mobili

2.3 Ontologie per la modellazione del contesto

Per la rappresentazione ed integrazione dei diversi contesti nel processo di analisi, si è ritenuto necessario progettare un'ontologia che si propone di modellare il contesto operatore-macchina-ambiente ricalcando le dimensioni principali del modello *Context Dimension Tree*, ma inserendo classi e relazioni basate sull'analisi dei dati testuali. In particolare, questo strumento ha permesso di analizzare alcune descrizioni della dinamica di incidenti, avvenuti in un arco temporale di 5 anni in Lombardia, estratti dalle banche dati Inail. Inoltre, per ottenere una descrizione più accurata delle conseguenze fisiche sui lavoratori, l'ontologia prodotta può essere integrata con ontologie mediche esistenti.

Oltre alla modellazione, scopo dell'ontologia è anche quello di integrare il contesto nelle operazioni di:

- definizione di processi di *decision-making* per la gestione del contesto e per l'identificazione di interventi.
- progettazione e sviluppo di algoritmi per individuare situazioni di rischio.

Dovendo quindi combinare rappresentazione del contesto e analisi di dati testuali, si è scelto di dividere l'ontologia in categorie grammaticali e classi semantiche. In questo modo è possibile scindere ruolo grammaticale e significato di ciascun termine, e utilizzare queste due informazioni indipendentemente o in maniera combinata, a seconda dell'analisi svolta.

Per quanto riguarda le categorie grammaticali, tenendo conto del pre-processing effettuato sui dati e degli algoritmi implementati per l'analisi (si veda la Sezione 3.3.1 per i dettagli), sono state inserite al momento le quattro classi più importanti, mostrate in Figura 13: (i) Sostantivo, (ii) Verbo, (iii) Aggettivo e (iv) Avverbio.

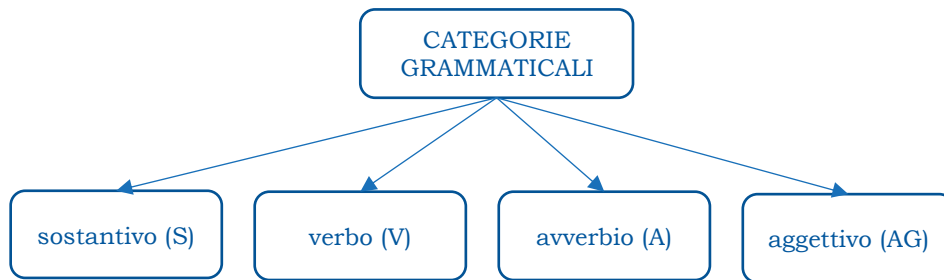


Figura 13 - Ontologia, categorie grammaticali

Per l'individuazione delle classi semantiche si è invece tenuto conto sia dell'albero di contesto sia dei dati testuali presenti nelle descrizioni degli incidenti analizzati. Seguendo quindi un approccio bottom-up, è stata svolta un'analisi delle occorrenze per individuare i termini più frequenti e definire le classi dell'ontologia sulle basi di questi. Queste classi sono poi state organizzate gerarchicamente tenendo conto della loro corrispondenza con le dimensioni dell'albero di contesto.

Come mostrato nella Figura 13, lo schema delle categorie grammaticali è molto semplice. Per definire la gerarchia delle classi semantiche, invece, è stato analizzato il dataset "Inail_Infortuni_Lombardi". Il dataset contiene, relativamente al dominio geografico e settoriale di interesse

(Lombardia, metalmeccanico), circa 20.000 denunce di infortunio sul lavoro all'anno, distribuite temporalmente negli anni 2013-2018.

La gerarchia delle classi semantiche, riportata nelle Figure 14 e 15, è stata ottenuta selezionando i primi cento termini più frequenti nell'attributo "r_descrizione_particolareggiata_lavorazione" del dataset.

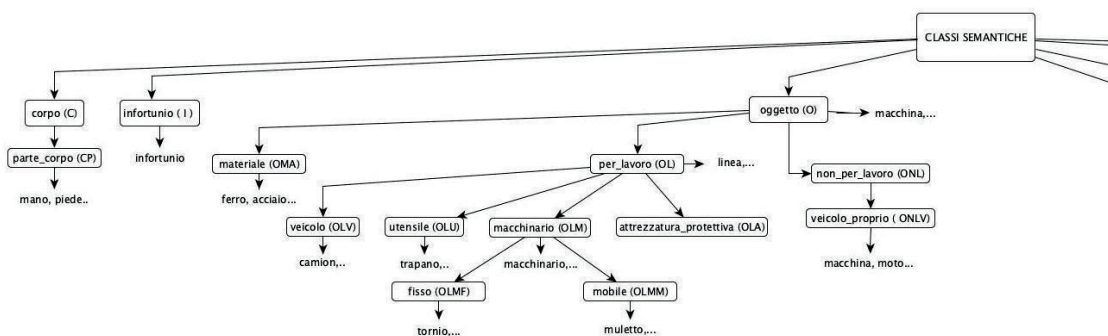


Figura 14 - Ontologia, classi semantiche (1)

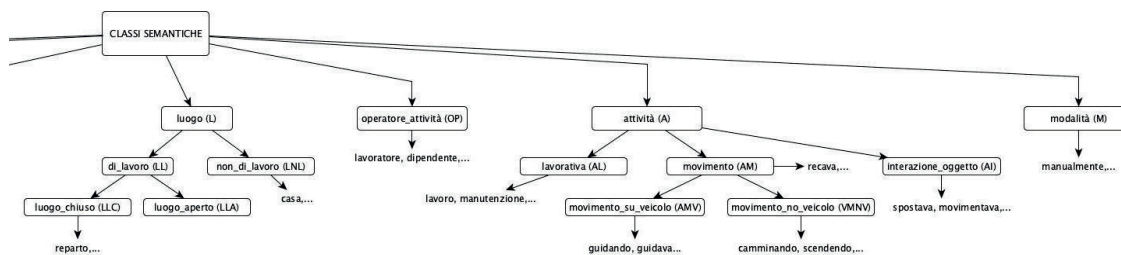


Figura 15 - Ontologia, classi semantiche (2)

È possibile vedere come, sebbene siano stati considerati solo i primi cento termini più frequenti, buona parte degli elementi che concorrono all'identificazione del contesto riescano a essere modellati dall'ontologia. L'ontologia, infatti, permette di rappresentare l'operatore, la sua attività corrente, il luogo e lo strumento che sta utilizzando, e anche a descrivere in parte l'infortunio.

Per quanto riguarda la combinazione di classi semantiche e categorie grammaticali, si è lavorato sulla formalizzazione di regole che ricalchino quelle della grammatica e dell'analisi logica italiana e le relazioni degli elementi all'interno del contesto di interesse. Ad esempio, per esprimere la "modalità" con cui viene svolta un'azione si dovrà combinare la classe semantica "modalità" con la categoria grammaticale "avverbio", o in altri casi "aggettivo" (analizzando più parole insieme questo concetto potrà essere espresso anche da una perifrasi). Sarà, inoltre, necessario esplicitare quali sono le combinazioni classi - categorie non ammesse per non creare ridondanze. Ad esempio, avendo creato una classe "attività lavorativa" la classe semantica "utensile" non dovrà mai essere combinata con la categoria "verbo" per rappresentare un verbo come "martellare".

Unified Medical Language System (UMLS) [Bo-004] è un compendio ontologico di diversi vocabolari delle scienze biomediche attraverso cui, sfruttando opportune tecniche di *mapping*, è possibile esplorare i vari sistemi ontologici inclusi, anche in più lingue.

Il paradigma di UMLS permette di superare due importanti criticità tipiche delle ontologie mediche: la diversità di rappresentazione di uno stesso concetto nei vari vocabolari e la specializzazione orizzontale delle singole ontologie. UMLS è strutturato in tre componenti:

- *Metatesauro*: comprende oltre un milione di concetti biomedici, espressi in cinque milioni di varianti derivate da 200 diversi vocabolari in più lingue, tra cui ICD 9-10⁵ e SNOMED CT⁶.
- *Rete semantica*: classifica in maniera coerente tutti i concetti inclusi nel metatesauro e fornisce le relazioni in essere tra questi. La rete comprende 133 tipi semantici (descritti anche in forma testuale) e 54 possibili relazioni.
- *Specialist lexicon*: comprende diversi strumenti di tipo Natural Language Process (NLP) che permettono di analizzare agilmente testi il cui lessico è significativamente di impronta biomedica.

Al fine di integrare diverse e molteplici basi di dati storiche di infortuni, sia di Inail che esterne, anche in diverse lingue, l'integrazione di un'ontologia come UMLS con l'ontologia costruita può apportare molti benefici.

2.4 Workflow per l'integrazione e l'elaborazione *context-aware* dei dati

La gestione del contesto guida la raccolta dei dati dal campo, la loro elaborazione e integrazione e, infine, la generazione degli allarmi.

Il processo si compone di 8 fasi, rappresentate in Figura 16.

Di seguito si illustrano le fasi attraverso l'analisi del seguente caso d'uso: “*Un lavoratore entra in un'area opportunamente delimitata Y, in cui è presente un macchinario fisso X per l'uso del quale sono richiesti specifici DPI non indossati dal lavoratore*”, che fa riferimento al contesto *Operatore_MacchinaUtensile*, illustrato nel § 2.2.4.2.

⁵ <https://www.cdc.gov/nchs/icd/index.htm>

⁶ <https://www.nlm.nih.gov/healthit/snomedct/index.html>

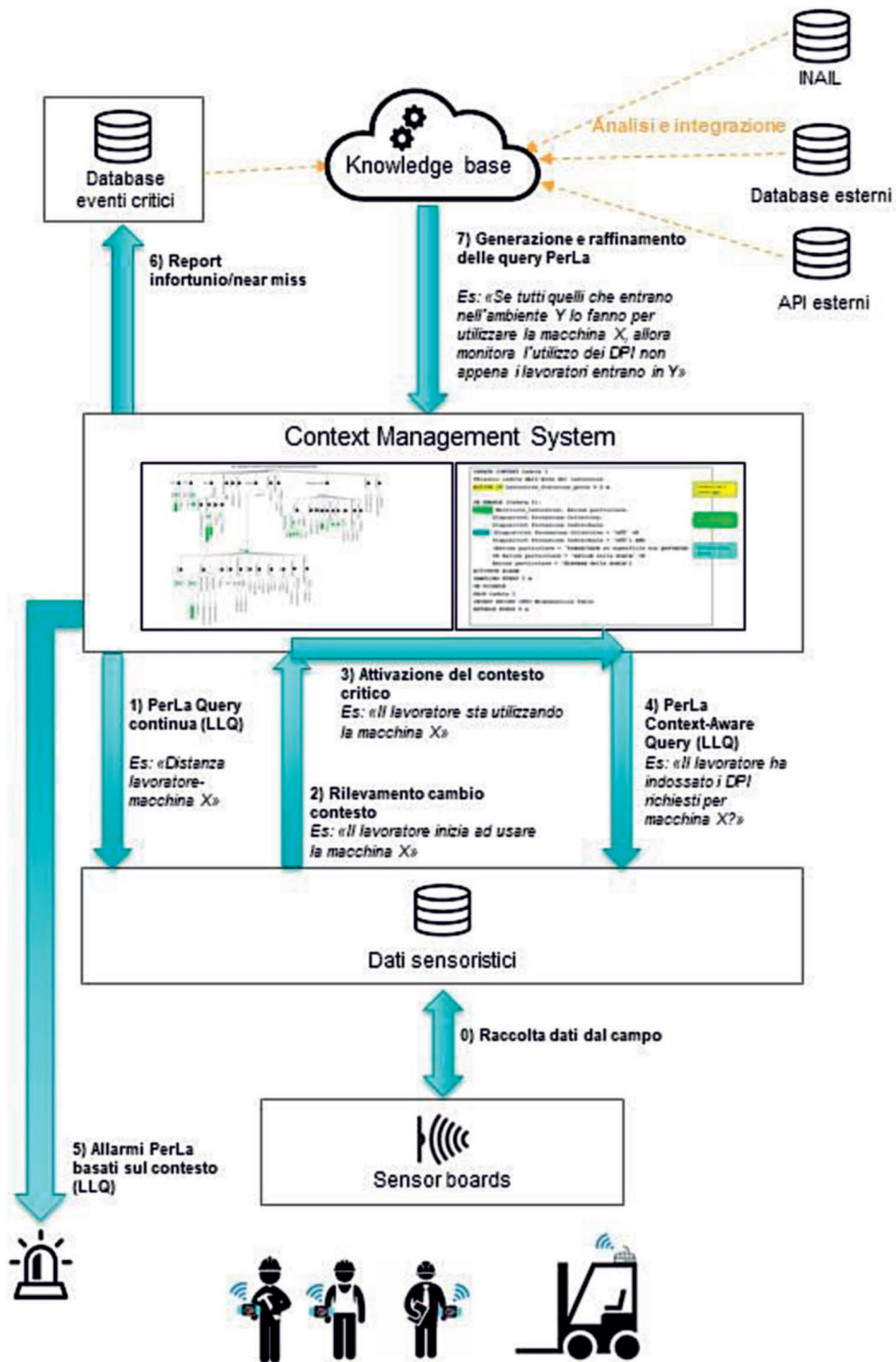


Figura 16 - Workflow per il monitoraggio della posizione relativa operatore - macchinario

Step 0: Raccolta dati dal campo

Come introdotto precedentemente, i primi livelli dell'architettura di sistema sono formati dall'in-

sieme dei dispositivi indossabili dai lavoratori, quelli installati sulle macchine, e dal sistema di localizzazione ed elaborazione immagini.

Lo Step 0 corrisponde alla fase di raccolta dei dati dal campo; perciò, sono considerate le misurazioni effettuate dai sensori installati, ad alto livello e senza perdita di generalità, come una base di dati.

Step 1: Query PerLa continue (LLQ)

Il monitoraggio si basa sull'esecuzione continua di un insieme di *query* del tipo Low Level Query (definite nelle sezioni precedenti). Tali *query* avviano delle operazioni di campionamento dei parametri principali del contesto (es: quota del lavoratore, posizione del lavoratore, ecc.) e permettono così di monitorare costantemente i cambiamenti di contesto.

Nel caso di esempio, la *query continue* che permette il monitoraggio è la seguente:

“qual è la distanza tra il lavoratore e il macchinario X?”, infatti, si vuole monitorare l'utilizzo della macchina X da parte del lavoratore.

Step 2: Rilevamento cambio di contesto

Il risultato della *query continue* (Step 1) evidenzia un cambiamento del valore del parametro presente nella *query* specifica di contesto (n.2 indicato in Figura 16).

Step 3: Attivazione di un contesto critico

L'aggiornamento del parametro può determinare l'attivazione di uno specifico contesto identificato come critico. Nell'esempio ha determinato il contesto *“il lavoratore sta utilizzando il macchinario X”*, per il quale sono necessarie ulteriori verifiche.

Step 4: PerLa Context-Aware Query (LLQ)

L'attivazione di un contesto critico determina l'esecuzione di una *query* specifica di contesto (o PerLa Context-Aware Query) mediante la quale è possibile verificare particolari condizioni del paradigma operatore-macchina-ambiente, generare segnali di allarme e infine definire quali sono i dati utili da collezionare.

Nel caso d'uso, una *query* specifica di contesto potrebbe essere: *“il lavoratore ha indossato tutti i DPI richiesti per l'utilizzo del macchinario X?”*

Step 5: Allarmi PerLa basati sul contesto

Come detto in precedenza, con le *query* specifiche di contesto è anche possibile generare allarmi che vengono attivati ogni volta che si verificano le condizioni critiche per il contesto considerato. Ad esempio, nel caso d'uso, se il lavoratore non indossa i DPI richiesti per l'utilizzo del macchinario viene attivato un allarme, e sono messe in atto le misure necessarie a prevenire un possibile infortunio.

Con lo *step 5* si conclude la parte *“on line”* del processo *context-aware* che include la raccolta dei dati e il monitoraggio tramite *query*, l'attivazione dei contesti critici e la generazione degli allarmi.

Negli *step* successivi si attiva la parte *“off line”* del processo, che comporta la memorizzazione dei dati raccolti nel database degli eventi critici e la successiva integrazione dei dati con altri *dataset* al fine di creare una base di conoscenza e condurre analisi ad ampio spettro.

Step 6: Memorizzazione rapporto near-miss o infortunio

Il processo prevede la generazione (anche in maniera automatica) di un rapporto con i dettagli dell'incidente o del quasi-incidente, a partire dal modello dell'Albero di Contesto (vedi CDT_Operatore, CDT_Infortunio, CDT_near-miss).

Il report viene memorizzato nel database degli eventi critici, condivisibile (con opportune differenze e diversi livelli di privacy) con i database esterni a cui è collegata l'azienda per analisi di più alto livello.

Il formato del rapporto è di tipo strutturato o semi-strutturato, per facilitarne una successiva analisi automatizzata (Step 7).

Se ne riporta di seguito una possibile rappresentazione testuale relativa al caso d'uso:

“Un lavoratore, dopo essere entrato nell'area Y dell'azienda, ha iniziato a utilizzare il macchinario X non indossando i DPI (<elenco_DPI>) obbligatoriamente richiesti. Sono stati attivati gli allarmi (<elenco_allarmi>). Nessuna conseguenza per il lavoratore.”

Step 7: Generazione e affinamento delle query PerLa

Come riportato nella Figura 16, integrando e analizzando i diversi *dataset* storici di infortuni disponibili e i rapporti generati nello Step 6 nei vari *Case Labs* sarà possibile costruire una base di conoscenza allo scopo di generare nuove *query* PerLa e raffinare quelle esistenti (*continue* e specifiche di contesto, step 1 e 4). Il affinamento delle *query* permette di anticipare il rilevamento di situazioni di rischio e, conseguentemente, diminuire il rischio di infortunio.

Con riferimento al caso d'uso, dall'analisi dei dati si potrebbe evincere che molti lavoratori sono entrati nell'area Y per utilizzare il macchinario X.

Con questa nuova informazione è possibile *anticipare* la verifica che il lavoratore, già al momento del suo ingresso nell'area Y, indossi i DPI richiesti per usare il macchinario.

Nel caso in cui tali DPI non siano indossati all'ingresso dell'area Y, si attiva un allarme per ricordare al lavoratore la necessità di indossare i dispositivi previsti.

Per perseguire quanto descritto nel caso d'uso sono necessari almeno due differenti livelli di integrazione *context-aware* dei dati:

1. L'elaborazione integrata dei *data stream* dai sensori, durante la fase di raccolta dei dati dal campo (Step 0).
2. L'integrazione dei dati raccolti nel database degli eventi critici con altri dataset di infortuni e near-miss (dataset Inail, altri database e API esterni), con l'obiettivo di costruire una base di conoscenza e condurre successive analisi utili al miglioramento del sistema di monitoraggio *context-aware*.

3 RECKONition: l'intelligenza artificiale per l'analisi della dinamica degli infortuni

In questo capitolo si descrive RECKONition, lo strumento software sviluppato per l'analisi dei testi riguardanti la dinamica degli infortuni, rilasciato a Inail al termine del progetto RECKON. RECKONition rappresenta un importante risultato del progetto, è, infatti, un potente tool che ha guidato le scelte implementative effettuate nell'installazione del sistema nelle aziende pilota.

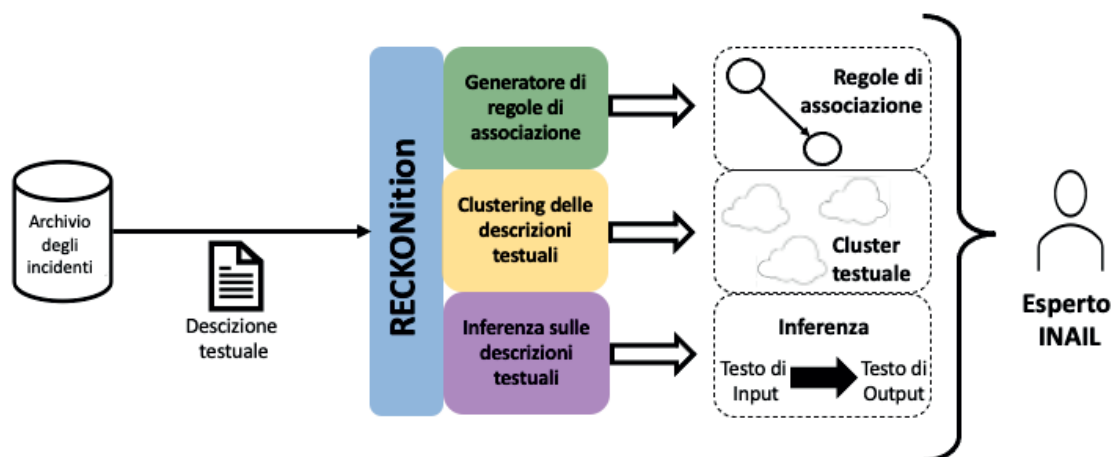


Figura 17 - RECKONition: overview del sistema

L'architettura del sistema RECKONition, illustrata nella Figura 17, comprende tre diversi moduli: il Generatore di regole di associazione, il modello di Clustering di descrizioni testuali e quello di Inferenza di descrizioni testuali. Tutti questi strumenti ricevono in ingresso un insieme di descrizioni testuali di incidenti sul lavoro e forniscono in output (i) regole di associazione, che servono per scoprire relazioni tra i termini, (ii) cluster di descrizioni testuali, che evidenziano gruppi di incidenti sul lavoro che condividono somiglianze nelle loro descrizioni, e (iii) previsioni di frasi successive estratte dalla descrizione testuale degli incidenti. Tutti gli output del sistema RECKONition rappresentano strumenti preziosi per l'esperto Inail per acquisire conoscenze sugli incidenti sul lavoro e definire politiche di prevenzione. I tre moduli del sistema RECKONition sono descritti in dettaglio nelle prossime sezioni.

3.1 Definizione del processo decision-making

Dall'analisi, svolta nel corso del progetto, degli algoritmi disponibili in letteratura per sviluppare tecniche di pre-processamento di dati testuali, l'attenzione si è concentrata sugli algoritmi di *clustering*. Questa è una tecnica di apprendimento non supervisionato che opera su dati non etichettati (di cui quindi non si conosce a priori la categorizzazione, o l'appartenenza a classi specifiche) e raggruppa dati simili in *cluster*. L'algoritmo ricerca, dunque, *pattern* nei campioni (*samples*) utilizzati per l'addestramento, evidenziando le loro similarità e differenze. Il primo step della *cluster analysis* è quello di definire il concetto di similitudine e, conseguentemente, di distanza all'interno del *dataset*. È stato scelto di incorporare nell'analisi l'ontologia descritta in precedenza (§ 2.3), valutando le occorrenze dei termini ed estraendo i 100 elementi più frequenti. Ognuna di queste parole è poi stata sostituita con un TAG, cioè un'etichetta che ne rappresentasse la classe ontologi-

ca di appartenenza. In questo modo termini appartenenti alla stessa classe e quindi riferiti ad un “incidente” simile, vengono rappresentati allo stesso modo e considerati uguali dall’algoritmo di *clustering*.

In particolare, per verificare l’efficacia generale del processo, il *dataset* è stato diviso in due insiemi:

- *training_set*: 90% delle descrizioni sulla base delle quali sono state calcolate le occorrenze dei termini e definiti i 100 più frequenti
- *test_set*: 10% delle descrizioni alle quali sono state applicate le sostituzioni e conseguentemente gli algoritmi di *clustering*.

I TAG utilizzati sono la combinazione dei TAG delle categorie grammaticali e delle classi semantiche riportati nell’ontologia. Per non introdurre errori nella *clusterizzazione* si è scelto di combinare TAG grammaticali e semantici e sostituire più parole possibili, lasciando invariati però quei termini che risultano ambigui se analizzati singolarmente. Ad esempio, il termine “taglio” può indicare sia una ferita che una lavorazione di un materiale e non è stato dunque sostituito con nessun TAG. Oltre ai termini ambigui, sono rimasti invariati anche quelli di cui risultasse evidente solo la categoria grammaticale.

Il *dataset* finale è composto da 12449 *samples* e 6662 parole presenti nel documento, per ciascuna delle quali è stata calcolata la funzione di peso TF_IDF (*Time Frequency - Inverse Document Frequency*) [LRU-014]. La TF_IDF è una funzione utilizzata in *information retrieval* per misurare l’importanza di un termine rispetto ad un documento o ad una collezione di documenti. Tale funzione aumenta proporzionalmente al numero di volte che il termine è contenuto nel documento, ma cresce in maniera inversamente proporzionale con la frequenza del termine nella collezione.

Nel nostro caso i *sample*, e quindi documenti, sono le descrizioni particolareggiate degli incidenti. L’idea alla base è quella di dare più importanza ai termini che compaiono nel documento, ma che in generale sono poco frequenti. La funzione può essere scomposta in due fattori. Il primo fattore (*Time Frequency*) indica il numero dei termini presenti nel singolo documento. In genere questo numero viene diviso per la lunghezza del documento stesso per evitare che siano privilegiati quelli più lunghi. L’altro fattore della funzione (*Inverse Document Frequency*) indica l’importanza generale del termine nella collezione. Il calcolo si basa sull’idea che più un termine è frequente nei diversi documenti, minore sarà la sua importanza del termine per identificare il singolo documento. I valori di ciascun termine saranno quindi i valori TF-IDF dei termini nelle diverse descrizioni.

Due descrizioni saranno dunque simili, e avranno quindi maggiore probabilità di essere *clusterizzate* insieme, più termini avranno in comune, ma ciascun termine avrà un peso differente sulla distanza.

Oltre alla definizione di similitudine/distanza è inoltre importante definire gli algoritmi di *clustering* specifici che si intende utilizzare. In generale gli algoritmi di *clustering* possono differire tra loro a seconda delle metriche utilizzate, dei diversi approcci e dei diversi parametri necessari.

3.1.1 Kmedoids

Kmedoids [HC-009] è un algoritmo che appartiene alla categoria dei “*representative based clustering*”, ovvero quegli algoritmi che permettono di individuare, per ciascun *cluster* formatosi, un rappresentante, un punto che riassume le caratteristiche del gruppo. In *Kmedoids* in particolare, questo è un elemento del *cluster* stesso e viene chiamato *medoide*. Un *medoide* può essere definito

come un elemento di un *cluster* la cui dissimilarità media rispetto a tutti gli oggetti nel *cluster* è minima, in questo modo esso sarà il punto più centrale di un dato insieme di punti. L'algoritmo è tra i più semplici e veloci tra quelli a disposizione, ha però tre svantaggi fondamentali, quali: necessita di conoscere a priori il numero di cluster da individuare, permette di individuare solo cluster convessi, è sensibile agli *outlier*.

Kmedoids infatti inizializza in maniera casuale ($n = \text{numero_di_cluster}$) punti come *medoidi*. Calcola poi per ogni punto la distanza dal *medoide* e lo assegna al *medoide*, e quindi al *cluster* più vicino; infine, dopo la formazione di questi primi *cluster* reinizializza i *medoidi*. L'algoritmo continua a iterare in questo modo finché i *medoidi* di ciascun *cluster* si stabilizzano. Non conoscendo a priori le tipologie di infortuni e quindi il numero esatto di *cluster*, l'algoritmo è stato eseguito con valori da 2 a 30.

Per valutare la qualità dei *cluster* ottenuti, è stata analizzata la capacità dell'algoritmo di raggruppare frasi contenenti parole con lo stesso significante, ma anche con lo stesso significato all'interno del contesto operatore-macchina-ambiente. Ovviamente, *cluster* che contengono frasi simili tra di loro, come ad esempio i cluster 27 e 28, sotto elencati, sono *cluster* significativi. Ciò che interessa però non è solo raggruppare descrizioni contenenti le stesse parole, ma raggruppare descrizioni che riguardano incidenti simili nel contesto analizzato. Sotto questo punto di vista, il cluster 27 è sì significativo, poiché identifica tutte le situazioni che vedono l'operatore utilizzare una mola, ma non raggruppa questo tipo di incidenti con altri incidenti che coinvolgono un utensile generico.

cluster #27

medoide: ['VAIO', 'mola'] | USAVA LA MOLA

['REGOLANDO LA MOLA', 'UTILIZZAVA LA MOLA', 'MOLAVA UN PEZZO DI FERRO CON UNA MOLA MANUALE', 'STAVA UTILIZZANDO UNA MOLA', 'USAVA LA MOLA PER LA RAGGIATURA DI UN CILINDRO', 'STAVA PRENDENDO LA MOLA', 'UTILIZZAVA LA MOLA', 'STAVA SP OSTANDO UNA MOLA DALLA MACCHINA AD UN BANCO ASPIRAZIONE', 'STAVA EFFETTUANDO UN OPERAZIONE DI MANUTENZIONE MOLA', 'UTILIZZAVA UNA MOLA', 'UTILIZZAVA LA MOLA SMERGLIATRICE', 'USAVA UNA MOLA', 'UTLIZZAVA UNA MOLA', 'STAVA USANDO L A MOLA CON UN PEZZO', 'PULIVA UN PEZZO DI METALLO CON MOLA', 'USAVA LA MOLA', 'IL LAVORATORE STAVA USANDO UNA MOLA FLESSIBILE', 'UTILIZZAVA LA MOLA', 'STAVA PROFILANDO LA MOLA']

cluster #28

medoide: ['non', 'conoscenza'] | NON NE SIAMO A CONOSCENZA

['NON NE SIAMO A CONOSCENZA', 'NON SIAMO A CONOSCENZA', 'NON NE SIAMO A CONOSCENZA', 'L'AZIENDA NON E' A CONOSCENZA DELLA DINAMICA DELL'EVENTO', 'L'AZIENDA NON E' A CONOSCENZA DELL'INFORMAZIONE', 'NON NE SIAMO A CONOSCENZA', 'NON NE SIAMO A CONOSCENZA', 'NON SIAMO A CONOSCENZA DELLE MODALITA DELL'ACCADUTO', 'NON È A CONOSCENZA DELLA DITTA', 'NON SIAMO A CONOSCENZA DEI FATTI', 'IL DATORE DI LAVORO NON E' A CONOSCENZA DELLA DINAMICA DELL'INFORTUNIO', 'NON SIAMO A CONOSCENZA', 'NON SIAMO A CONOSCENZA DI NESSUN EVENTO DI INFORTUNIO', 'NON A CONOSCENZA', 'IL DATORE DI LAVORO NON E' A CONOSCENZA DELL'AVVENUTO INFORTUNIO.', 'NON NE SIAMO A CONOSCENZA. NON HA AVVERTITO IL SUO RESPONSABILE DI LINEA.', 'NON NE SIAMO A CONOSCENZA', 'NON SIAMO A CONOSCENZA DELLA DINAMICA DELL'EVENTO', 'NON SIAMO A CONOSCENZA', 'NON SIAMO A CONOSCENZA', 'NON SIAMO A CONOSCENZA DI QUANTO SUCCESSO', 'NON SIAMO A CONOSCENZA DI QUANTO SUCCESSO', 'NON SIAMO A CONOSCENZA DELLA DINAMICA DELL'EVENTO.', 'NON SIAMO A CONOSCENZA DI COME SIA AVVENUTO L'INFORTUNIO IN QUANTO NON ABBIAMO AVUTO COMUNICAZIONE IN MERITO', 'NON NE SIAMO A CONOSCENZA']

Nei *cluster* precedenti dunque, sono state raggruppate descrizioni molto simili tra loro, ma l'impatto della sostituzione dei TAG è stato minimo. I *cluster* 12 e 25, riportati di seguito, illustrano esempi dove l'impatto delle sostituzioni risulta più evidente. I due *cluster* descrivono rispettivamente incidenti avvenuti utilizzando utensili (quali: trapano, taglierino, martello) e incidenti avvenuti al di fuori dell'ambiente di lavoro a bordo di veicoli propri (auto, moto, bici).

La Confidenza di una regola viene definita come:

$$Conf(A \Rightarrow B) = \frac{Supp(A \Rightarrow B)}{Supp(B)}$$

e indica quante volte gli *item* in B compaiono in transazioni che contengono A.

Infine, il *Lift* di una regola si definisce come:

$$Lift(A \Rightarrow B) = \frac{Supp(A \Rightarrow B)}{Supp(A) * Supp(B)}$$

Essa rappresenta la correlazione tra le occorrenze di A e B: un *Lift* maggiore di 1 identifica una correlazione positiva, ossia la probabilità di osservare B avendo già osservato A è maggiore della probabilità di osservare B.

Il calcolo delle metriche descritte viene abitualmente effettuato per gli *item* appartenenti all'insieme degli *item* più frequenti, indicato come *Frequent Item Set* (FIS). Un *item* A appartiene all'insieme FIS se il suo supporto è maggiore di una soglia minima, ossia se $Supp(A) > minsupp$.

La definizione del *Frequent Item Set* è alla base dell'algoritmo Apriori, un metodo di estrazione di regole di associazione ben noto in letteratura, che sfrutta le misure introdotte per esplorare l'insieme di elementi e costruire regole di associazione a partire da esse [Ag-994]. In alcuni contesti però, come ad esempio quello medico, non sono solo gli *item* più frequenti a produrre regole rilevanti, ma spesso gli *item* meno frequenti rivestono una maggiore importanza semantica (ad esempio, nomi di malattie o persino di lesioni specifiche).

In [MSG-014] dunque, viene presentata una versione dell'algoritmo Apriori estesa al caso degli *infrequent Itemset* (inFIS), chiamata FISinFIS Apriori. Seguendo questa intuizione, è stata realizzata una versione dell'algoritmo FISinFIS Apriori basata su operazioni tra matrici sparse, in cui le regole di associazione positive (PAR) e negative (NAR) sono costruite a partire sia dall'insieme di elementi frequenti sia dall'insieme di elementi infrequenti.

Come nell'algoritmo FISinFIS Apriori originale, l'algoritmo sviluppato nel progetto imposta un valore minimo di *Inverse Document Frequency* (IDF) come limite inferiore per filtrare gli *item*, cioè i termini che sono troppo frequentemente utilizzati nel database. Inoltre, è stata introdotta una soglia aggiuntiva all'IDF come limite superiore per gli elementi utilizzati troppo raramente. La logica alla base di questa decisione è quella di eliminare sia i termini utilizzati in tutte le descrizioni degli incidenti (e quindi poco rilevanti per la singola descrizione) sia i termini contenenti errori di battitura. Questi ultimi, infatti, costituiscono una parte non trascurabile del database considerato.

Nella prima fase di analisi, inoltre, è stato applicato un pre-processamento per eliminare le *stop word*, parole che svolgono la sola funzione di collegare i vari elementi di una frase e non sono indispensabili ai fini della definizione del significato di un periodo o di una proposizione, e svolgono, dunque, un ruolo principalmente sintattico, non semantico. È stato poi effettuato *stemming* delle parole rimanenti, ovvero la riduzione della forma flessa di una parola alla sua forma radice, detta "tema". In particolare, l'implementazione di *stemming* scelta è quella proposta in [Po-001], detta *Snowball Stemming*, che riconosce ed elimina nelle parole le desinenze (conosciute a priori dall'algoritmo). Grazie a questo pre-processing si è ottenuto un *item set* più scarno, privo di desinenze, punteggiatura e lettere maiuscole.

I risultati ottenuti dall'applicazione dell'algoritmo di *mining* al database di descrizioni di incidenti

sul lavoro consentono a RECKONition di individuare le proprietà o le azioni industriali che sono più probabilmente correlate a un infortunio o, con regole negative, alla sua prevenzione. Esempi di regole positive (PAR) sono quelle che legano i termini “incident”, “stradal”, e “guid”, da cui si deduce una correlazione tra l’occorrenza di un incidente e l’azione guidare. In Figura 18, invece, riportiamo più AR, sia positive che negative, che hanno in comune il termine “dpi”. Sebbene alcune relazioni siano poco significative, come “dpi” e “indoss”, quelle dedotte dalle NAR sottolineano, ad esempio, come ci sia una relazione tra la presenza di “dpi” e l’assenza di “contus” o l’assenza di “coinvolt”.

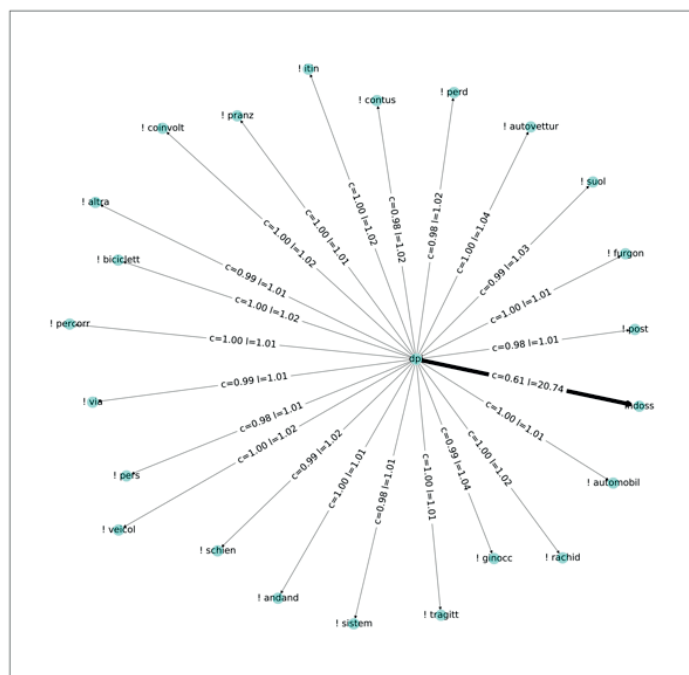


Figura 18 - Esempio di AR ottenuta con FISinFIS Apriori

3.3 Clustering del linguaggio naturale

Il *clustering* delle descrizioni testuali [LRU-014] viene eseguito per evidenziare le differenze e le somiglianze tra le descrizioni degli incidenti e raggruppare quelle simili, al fine di facilitare l’identificazione di interventi appropriati in base alla situazione descritta. Nelle sezioni seguenti verranno illustrati i due diversi approcci impiegati dal nostro sistema. Entrambi utilizzano l’algoritmo K-Medoids [PJ-009], ma si differenziano per le operazioni eseguite sui dati.

3.3.1 Clustering basato su TAG e occorrenze

Il primo approccio di *clustering* si basa sull’ontologia descritta nella Sezione 2.3. Il contesto di interesse è l’azienda metallurgica, perciò vengono utilizzati TAG che rappresentano classi ontologiche per integrare la conoscenza contestuale nell’analisi di *clustering*.

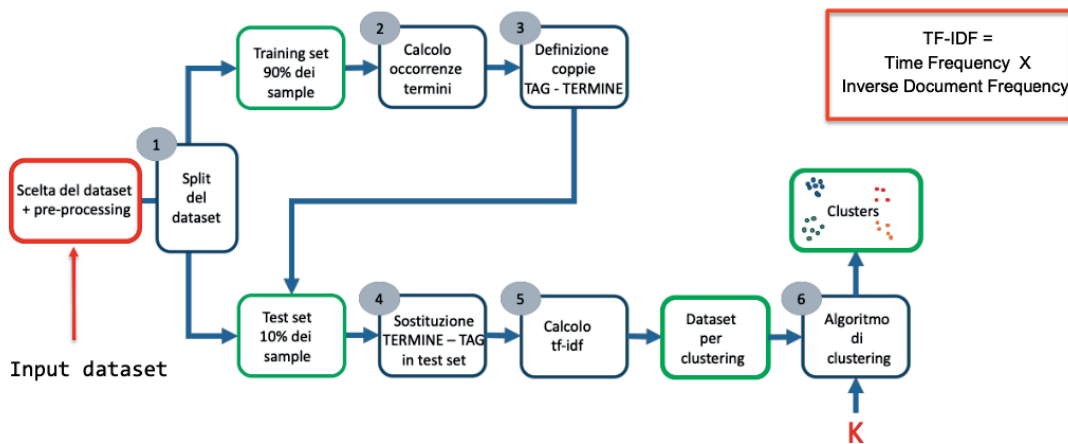


Figura 19 - Algoritmo per il clustering basato su TAG e occorrenze

La Figura 19 riassume i passaggi principali della metodologia usata. Inizialmente si calcolano le occorrenze di ogni parola, si identificano le cento parole più frequenti (2) e si definiscono le coppie TAG - TERMINE nel *training set* (3); si eseguono poi le cento sostituzioni nel *test set* (4). In questo modo, le parole che rappresentano concetti simili vengono lette dall'algoritmo come lo stesso termine. Come si può notare anche dal diagramma, prima di calcolare le occorrenze si effettuano tutte le necessarie operazioni di preelaborazione sull'intero *dataset*, e in particolare vengono eliminate le *stop word*. Quasi tutte le parole rimanenti sono risultate quindi descrittive del contesto, ed è stato possibile associarle a un TAG. Nei rari casi in cui ciò non è avvenuto, o la sostituzione ha introdotto ambiguità, i termini sono rimasti invariati.

Una volta effettuata la sostituzione, si calcola la *Time Frequency - Inverse Document Frequency* (TF-IDF) [LRU-014] di ciascun termine in ogni descrizione (5). Da questa operazione si ottiene un set di dati pronto per il *clustering*. nell'elaborazione dei dati disponibili si è ottenuto un insieme composto da 6.662 termini (*features*), i cui valori corrispondono al TF-IDF calcolato. Una volta ottenuto il set di dati, si esegue l'algoritmo K-Medoids.

Di seguito si riportano alcuni esempi di *cluster* ottenuti impostando il parametro $K = n_clusters$ pari a 30, e mostriamo come si è stati in grado di separare le diverse situazioni descritte.

```
Cluster 12: medoide: [USANDO IL TRAPANO]
elementi: [USANDO IL TUBO]; [USANDO IL TRAPANO]; [USANDO IL MARTELLO];
[USANDO LA FRESA];...
```

```
Cluster 17: medoide: [SCESO DAL CARRELLO ELEVATORE]
elementi: [SCENDERE DAL CARRELLO ELEVATORE]; [SALIRE SUL CARRELLO
ELEVATORE]; [SPOSTARE IL CARRELLO]; [GUIDARE IL CARRELLO];...
```

È importante notare che l'algoritmo non solo è in grado di raggruppare le descrizioni che contengono le stesse parole, ma anche di identificare quelle che riguardano incidenti simili nel contesto analizzato. Negli esempi mostrati, il cluster 12 rappresenta gli incidenti che coinvolgono strumenti

di lavoro (trapano, martello, fresa,...), mentre il cluster 17 rappresenta gli incidenti che si verificano a bordo di macchinari mobili.

3.3.2 Clustering basato su Transformers

Il secondo approccio al *clustering* adottato da RECKONition sfrutta i nuovi modelli *Encoder-Decoder* con meccanismi di *attention* e *self-attention* per costruire gruppi omogenei di descrizioni di incidenti [Ch-014, Be-003]. Queste tecniche si basano su reti neurali, e in particolare combinano diversi strati di *encoder-decoder* impilati, in modo da modellare diverse strutture sintattiche con moduli di *self-attention*, insieme a reti neurali *feed-forward*, con il risultato di milioni di parametri addestrabili. La versione base del *Bidirectional Encoder Representations from Transformers (BERT)* [Dev-018] presenta 12 strati di *encoder-decoder*, 768 nodi nascosti per ogni strato e 12 *attention head*, con il risultato di 110 milioni di parametri. Dato l'elevato numero dei parametri, RECKONition utilizza un modello BERT pre-addestrato su 13 GB di dati testuali italiani provenienti dai corpora di Wikipedia e OPUS3ed, e ne esegue un *fine-tuning* su un sottoinsieme del *dataset* fornito da Inail. Il modello perfezionato viene poi convalidato su un set di *hold-out* che è utilizzato per eseguire il *clustering* su descrizioni testuali non viste: durante il *forward step* di una nuova frase della rete, gli *hidden layer* dell'ultimo strato del *decoder* del modello vengono estratti e utilizzati come *feature* numeriche che rappresentano la frase.

Questa procedura viene eseguita per ogni descrizione presente nell'insieme di validazione, in modo da costruire un nuovo *dataset* con forma $v \times (768 * nt)$, essendo 768 il numero di stati nascosti di uno strato BERT, v il numero di frasi nell'insieme di validazione e nt il numero di lunghezze delle frasi *tokenizzate*.

RECKONition sfrutta l'*Incremental Principal Component Analysis (IPCA)* [Ch-014] per ridurre la dimensionalità delle caratteristiche rappresentative a un numero di componenti principali tale da esprimere almeno l'85% della variabilità dei dati ed esegue il *clustering* KMedoids per valori di $n_cluster$ k nell'intervallo [2, 100]. A differenza dell'approccio di *clustering* descritto nel paragrafo precedente, questo metodo può trovare efficacemente frasi con strutture sintattiche altamente correlate, trascurando però alle volte gli aspetti semantici. Di seguito mostriamo due esempi di cluster ottenuti con l'approccio di *clustering* basato sui trasformatori applicato a un set di dati campione.

```
Cluster 10: Medoide: "A CAVALLO DEL CARRELLO ELEVATORE"
```

```
Elementi: ["GUIDARE IL CARRELLO ELEVATORE", "GUIDARE L'AUTO", "GUIDARE IL CARRELLO ELEVATORE", "GUIDARE L'AUTO", "GUIDARE LA MOTO", "GUIDARE IL CARRELLO ELEVATORE", "RIPARARE IL CARRELLO ELEVATORE"].
```

```
Cluster 27: Medoide: "SALIRE LA SCALA"
```

```
Elementi: ["ARRAMPICARSI SULLA SCALA", "ARRAMPICARSI SULLA PARETE", "ARRAMPICARSI SULLA SCALA", "ARRAMPICARSI SUL TETTO", "ARRAMPICARSI SULLA SCALA", "ARRAMPICARSI SULLA SCALA", "RIMUOVERE IL DIFENSORE"].
```

3.4 Language model

Esistono numerosi contesti sperimentali in cui un *Language Model* (LM) [Be-003], cioè un modello in grado di elaborare sequenze testuali [SVL-014], può essere efficacemente applicato. Si va dall'estrazione di caratteristiche, come è stato fatto con il *clustering* basato su *Trasformers* (illustrato nel paragrafo precedente), alla classificazione di testi, alla risposta a domande, alla previsione di frasi successive, all'inferenza del linguaggio naturale o persino alla generazione di testi.

RECKONition comprende un LM personalizzato, che è stato opportunamente costruito e addestrato sul *dataset* di incidenti fornito da Inail. L'obiettivo è quello sviluppare un modello in grado di ricevere in input la descrizione della dinamica dell'infortunio e fornire in output una predizione delle conseguenze sui lavoratori coinvolti, espressa in linguaggio naturale.

Si sottolinea che la fase di tokenizzazione [WK-992], che permette di passare dalla rappresentazione simbolica (testuale) a quella numerica, richiede la definizione di una dimensione del vocabolario per fissare il numero di *token*, che il nostro modello fissa a 5000 elementi. Il livello di *embedding* viene quindi utilizzato per proiettare la rappresentazione rada ottenuta dalla fase di tokenizzazione in uno spazio denso di dimensione 128 *token*.

La capacità di apprendimento delle sequenze del modello risiede nello strato n mostrato in Figura 13, che è composto da due LSTM (*Long Short-Term Memory*) bidirezionali con 100 unità ciascuno [HS-997, GJM-013], uno strato di appiattimento, due strati densi con 50 neuroni ciascuno e attivazioni Rectified Linear Units (ReLU) e uno strato finale di *Dropout* [Sr-014] con un *drop rate* del 50% prima dello strato denso di uscita con 5000 neuroni (uno per ogni elemento) e attivazione sigmoideale.

Questo modello può essere addestrato in modo supervisionato [Bi-006] minimizzando la *Binary Cross Entropy* come *loss function* con *Adaptive Moment Estimation* (ADAM) [KB-014]. Ciò si ottiene fornendo le descrizioni delle dinamiche degli incidenti come caratteristiche in ingresso, insieme ai corrispondenti effetti e gli infortuni dei lavoratori coinvolti come target di input. Per una nuova frase non vista che descrive uno scenario lavorativo, il modello emette un vettore di probabilità sulle voci del vocabolario che definiscono quelle che meglio descrivono la conseguenza della descrizione in ingresso.

L'architettura di alto livello del LM è mostrata nella Figura 20 e rappresenta il processo adottato. Il modulo *Tokenizer* (t) trasforma una stringa (s) di input, espressa in linguaggio naturale, in un vettore numerico *rado* $t(s)$; il modulo di *incorporazione* (e) proietta l'output precedente in uno spazio numerico *denso* $e(t)$; infine, utilizzando la rete neurale (N), si estrae l'output numerico della rete neurale ricorrente $N(e)$, che a sua volta viene trasformato, mediante il decodificatore (d), in una stringa in linguaggio naturale $s'=d(N(e))$.

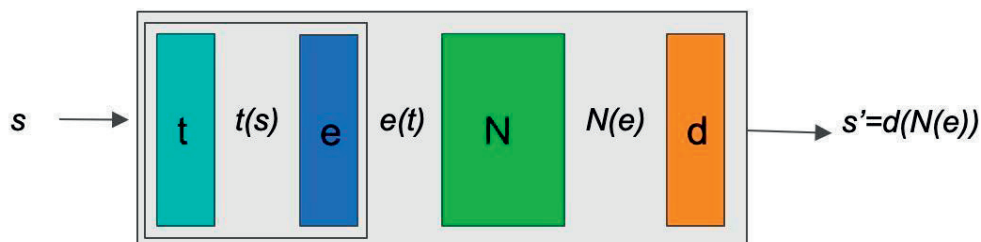


Figura 20 - Architettura di alto livello del modulo di Language Model di RECKONition

4 Dall'architettura al sistema

In questo capitolo sono descritte le tecniche studiate e implementate per la sensorizzazione del contesto e lo sviluppo di soluzioni *real-time* per l'individuazione delle situazioni di rischio. Verrà inoltre descritto nel dettaglio l'intero sistema installato nelle tre aziende (*Case Labs*) e verrà fornita una panoramica delle prime analisi eseguite sui dati dei primi tre mesi di funzionamento del sistema.

4.1 Sensorizzazione del contesto

L'attività di sensorizzazione si è articolata attraverso le seguenti fasi: la definizione, la progettazione di appropriati apparati hardware e la realizzazione di nuove soluzioni in funzione del contesto specifico. Nella fase preliminare del progetto è stato necessario analizzare la letteratura scientifica relativa al monitoraggio degli elementi di contesto (operatore-macchina-ambiente come riportato nell'esempio di Figura 21) e alla prevenzione degli incidenti in modo da avere una panoramica completa delle situazioni di pericolo. Tale analisi si è basata sulla banca dati Scopus⁷ di Elsevier B.V. e ha permesso di mettere in luce alcuni aspetti interessanti:

- la letteratura scientifica relativa al monitoraggio degli elementi di contesto si è concentrata sulla cantieristica e sul *mining* mentre è molto scarna per quanto riguarda il manifatturiero
- solo negli ultimi anni (a partire dal 2017) il monitoraggio della sicurezza nel mondo del manifatturiero sta attirando l'attenzione dei ricercatori; questa crescita di interesse per gli aspetti del monitoraggio e della sicurezza in ambito manifatturiero sembra legata alla disponibilità di nuovi sensori e di nuova tecnologia wireless
- i lavori scientifici si concentrano sul monitoraggio di alcune situazioni di pericolo mentre altre (tra quelle evidenziate dalle interviste fatte alle PMI) vengono completamente trascurate
- si osserva un approccio frammentato, ossia si tende ad analizzare, anche con grande dettaglio e approfondimento, una particolare situazione di pericolo con una determinata dotazione sensoristica ma si tende a perdere di vista la situazione complessiva e quindi non si riescono a cogliere situazioni complesse che generano situazioni di pericolo se più situazioni semplici si verificano contemporaneamente

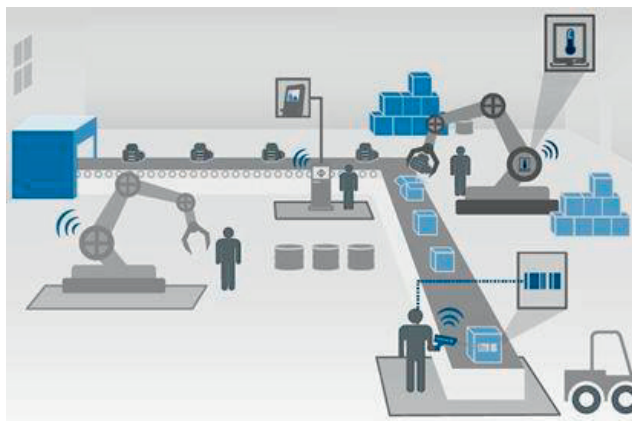


Figura 21 - Esempio di stazione di lavoro sensorizzata [Microsoft IoT in Manufacturing Infographics, 2015]

⁷ <https://www.scopus.com>

Oltre all'analisi della letteratura scientifica, si è fatto uno *scanning* approfondito delle tecnologie sensoristiche *off-the-shelf* che potessero coprire le diverse esigenze evidenziate durante le interviste e messe in luce dall'analisi di letteratura. Si sottolinea che tale *scanning* è stato rivolto alle soluzioni disponibili sul mercato, anche relative ad altri ambiti industriali, e non alle soluzioni prototipali o di ricerca (anche se con prestazioni attese migliori) in modo da poter pervenire a proposte di reti di sensori immediatamente installabili presso le aziende facenti parte del *Case Labs*. Abbiamo ritenuto infatti che fosse prioritaria la disponibilità della tecnologia piuttosto che le potenzialità di nuove tecnologie essendo la sicurezza un'emergenza immediata.

Altro aspetto fondamentale dell'analisi del mercato è stata quella di individuare soluzioni che potessero coprire più situazioni di pericolo. Infatti, individuare una soluzione che funziona solo in una particolare situazione non è adeguato all'idea fondante del progetto, ossia quella di predisporre un sistema di monitoraggio che possa essere facilmente adattato alle più svariate situazioni di pericolo. Tra le soluzioni sensoristiche commerciali individuate, abbiamo quindi preferito quelle caratterizzate da maggiore trasversalità in termini di situazioni coperte. Questo ha permesso di ridurre non solo la sensoristica necessaria a coprire le situazioni di pericolo, ma anche tutta l'infrastruttura hardware e software retrostante per la gestione della sensoristica stessa. Ovviamente, l'adozione di sensori "trasversali" comporta anche degli svantaggi: se un sensore è in grado di rilevare più grandezze contemporaneamente, per distinguere le grandezze di interesse (quelle che distinguono la situazione di pericolo da una situazione normale) è necessario implementare algoritmi complessi che "filtrino" una situazione di interesse rispetto a situazioni che non lo sono. Se quindi da un lato l'*hardware* e il *software* sono stati semplificati, in quanto occorrerà gestire un numero inferiore di sensori diversi, dall'altro la potenza di calcolo e la complessità degli algoritmi di elaborazione dei segnali sono aumentate.

Il terzo *driver* per la definizione della sensoristica appropriata è stato l'apertura del sistema; infatti, se è vero che si è data priorità alle soluzioni sensoristiche e *hardware* esistenti, si è però anche messo a punto un *framework* che possa scalare con le dimensioni dell'azienda, e quindi con la maggiore o minore esigenza di sensoristica, e che possa integrare in maniera *seamless* eventuali nuove soluzioni/nuovi sensori promettenti ed affidabili che potranno essere resi disponibili sul mercato nei prossimi anni. Inoltre, il sistema è stato pensato perché possa adattarsi alle specifiche situazioni aziendali in termini di infrastruttura informatica (ad es. database gestionale) e architettura hardware. Insomma, una piattaforma aperta alle nuove soluzioni, e plasmabile rispetto alle diverse condizioni al contorno sulle quali dovrà innestarsi.

4.2 Sensoristica per il monitoraggio di macchinari pericolosi

Al fine di individuare la sensoristica più appropriata per monitorare le principali macchine causa di incidenti e/o gli elementi più critici di macchine/strutture, nonché gli operatori stessi e l'ambiente, ci si è basati sull'analisi della letteratura scientifica.

Per la ricerca bibliografica sono state utilizzate le seguenti parole chiave: «risk», «safety», «monitoring», «prevention», «operator», «worker», «sensor». Sono stati quindi individuati ca. 800 articoli scientifici, tra i quali si sono selezionati quelli relativi al contesto industriale (manifatturiero) oggetto del progetto RECKON. Strettamente legati al contesto, abbiamo individuato solo 16 articoli distribuiti temporalmente come riportato in Figura 22.

Entrando più nel dettaglio della ricerca bibliografica, si nota come alcuni articoli ([Xu-014, Fa-019,

De-018, Ba-015]) siano relativi al sistema di monitoraggio con particolare riferimento alla posizione degli operatori all'interno dell'ambiente di lavoro. I sistemi proposti si basano sul paradigma dell'IoT (*Internet of Things*) e prendono in considerazione diverse tecnologie di comunicazione *wireless* che vanno dal *Radio-FrequencyID* (RFID) al più recente *Real-Time high-precision indoor Location* (RTLS), valutando come errori di comunicazione possano inficiare la valutazione della posizione relativa e quindi il sistema di monitoraggio.

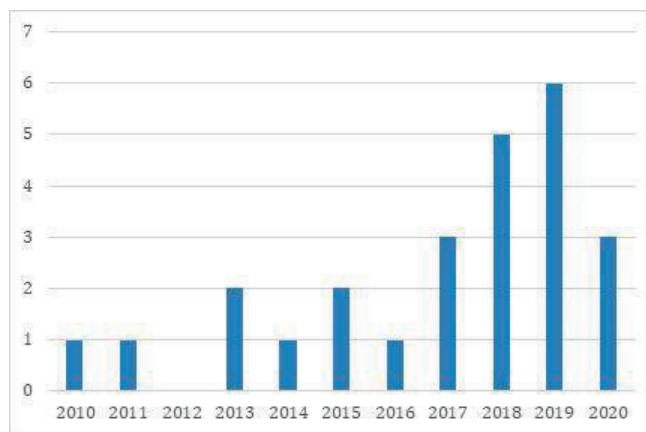


Figura 22 - Distribuzione temporale pubblicazioni scientifiche in ambito industria manifatturiera

Particolare attenzione viene dedicata al pericolo di caduta dell'operatore. Mentre i lavori più datati ([Ko-015, Yi-016a]) non considerano specificatamente l'ambiente lavorativo (ad esempio [Yi-016a] si focalizza sul pericolo di caduta di persone anziane all'interno delle loro abitazioni e solo marginalmente affronta le differenze che ci potrebbero essere se tale sistema venisse implementato in ambienti diversi da quello casalingo). [Ko-019] affronta direttamente il pericolo di caduta in ambiente industriale mettendo in evidenza le sue particolarità e spiegando come, per tale ambiente, siano necessarie soluzioni specifiche.

Vi sono poi alcuni articoli ([By-018, Yi-016b, Ce-013]) che indagano, nello specifico, le reti di sensori per la valutazione del rischio di collisione tra operatori e macchine in movimento. Tali sensori sono essenzialmente sensori di visione (telecamere nel visibile o nel vicino infrarosso) che permettono di tenere sotto controllo un ambiente dinamico in evoluzione come quello che si ha con persone e macchine in movimento.

Oltre alla potenziale collisione tra operatori e macchine, si considera come potenzialmente rischiosa l'operazione di sollevamento carichi ([Pa-018, Kh-018]) con eventuale ondeggiamento del carico stesso in condizioni di non perfetta visibilità (a causa di occlusioni dovute all'ambiente o al carico stesso). Anche in questo caso vengono analizzati sistemi basati su visione, eventualmente accoppiati con modelli fisici in grado di valutare le caratteristiche del carico sollevato.

L'articolo [Me-000] affronta invece il problema del monitoraggio della sicurezza in ambienti caldi, e come la strumentazione da adottarsi debba essere scelta per sopravvivere a questi ambienti, ma soprattutto quali siano i parametri cui fare attenzione e come tenerli sotto controllo. Si evidenzia la necessità di monitorare non solo i relativi movimenti, ma anche la differenza di temperatura dei vari oggetti che potrebbero venire in contatto con l'operatore e dell'ambiente stesso, che determina una massima durata di permanenza dell'operatore in quell'ambiente.

Da questa breve panoramica della letteratura esistente in quest'ambito si evince come essa sia molto parziale, e come ci siano ampi margini di pubblicazione sui diversi aspetti che vengono affrontati all'interno del progetto RECKON. Rispetto all'individuazione delle situazioni di pericolo considerate vengono però suggerite (e analizzate) soluzioni sensoristiche interessanti, e le conclusioni sono state utilizzate per la definizione della sensoristica appropriata ai settori industriali selezionati.

I sensori sono stati classificati nelle seguenti categorie:

- in grado di rilevare una singola grandezza fisica in uno specifico punto (ad es. l'accelerometro, la termocoppia, ...)
- in grado di rilevare più grandezze fisiche in uno specifico punto (ad es. l'Unità di Misura Inerziale (IMU), la coppia di fotocellule, ...)
- in grado di rilevare una singola grandezza fisica in un'area (ad es. la telecamera, la termocamera, ...)
- in grado di rilevare più grandezze fisiche in un'area (ad es. la camera iperspettrale)

Sulla base dell'analisi della letteratura e sulla base dell'analisi delle tecnologie sensoristiche disponibili sul mercato, al fine di ridurre il numero di sensori e di rendere più «generalizzabile» possibile la determinazione della situazione di rischio, si sono considerate tipologie di sensori in grado di rilevare più grandezze fisiche e/o in un'area.

4.3 Progettazione della sensoristica

Il *framework* sensoristico sviluppato prevede quindi l'utilizzo di:

- *smart watch* in dotazione agli operatori: lo *smart watch* fungerà sia da *input* per il sistema di monitoraggio, indicando lo stato e la posizione dell'operatore nell'ambiente di lavoro e relativamente ad oggetti/macchine, sia da *output* del sistema di monitoraggio, fornendo all'operatore messaggi di allarme, siano essi semplici vibrazioni, suoni o messaggi più complessi per guidare l'operatore in situazioni di rischio;
- sensori in radiofrequenza (Bluetooth Low Energy BLE, Wireless Local Area Network WLAN) in grado di determinare la distanza, la direzione di movimento e la velocità di uno *smart watch* rispetto al sensore stesso; pertanto, installando tali sensori in posizioni critiche e su macchine in movimento, sarà possibile risolvere situazioni di pericolo quali, ad esempio,
 - l'operatore entra in zone a rischio e/o dove non è autorizzato ad entrare,
 - l'operatore corre il rischio di essere investito da macchine in movimento,
 - l'operatore attraversa barriere (ad esempio le porte di emergenza in caso di evacuazione per incendio);
- combo telecamera-termocamera (dove necessaria) con *post-processing* delle immagini a bordo per il monitoraggio di aree estese (dove non sarebbe conveniente l'utilizzo di sensori in radiofrequenza), per gestire situazioni in cui l'operatore o il visitatore non è dotato di *smart watch*, nonché per monitorare il corretto impiego dei DPI in particolari situazioni. L'elaborazione a bordo della telecamera-termocamera risulta di fondamentale importanza per anonimizzare il dato, per ridurre il flusso di dati verso il sistema di elaborazione centralizzato e per riuscire ad intervenire in tempo reale in situazioni di emergenza, dove la latenza della comunicazione prevede un intervento tempestivo.

4.4 Sistemi di localizzazione INDOOR

La diffusione dei sistemi di localizzazione indoor nasce dalla necessità di risolvere il problema di raggiungibilità dei segnali satellitari all'interno di locali chiusi e dal bisogno di incrementare la risoluzione della posizione calcolata. In quest'ottica, la ricerca ha contribuito alla nascita di sistemi innovativi di localizzazione *indoor* basati su tecnologie radio, ottiche, magnetiche e acustiche. Nella Tabella 1 si riportano le tecnologie emergenti suddivise per tipologia e standard di comunicazione. In questa parte del documento prenderemo in considerazione le tecnologie Radio, in particolare la tecnologia Bluetooth low energy.

Tabella 1 - Principali tecnologie usate nei sistemi di localizzazione indoor

Radio	Ottiche	Magnetiche	Acustiche
UWB (Ultra-wideband)	Video Camera	Magnetic strength	Ultrasound
Bluetooth	Lidar (<i>Light Detection and Ranging</i>)		
UHF RFID	Infrared light pulses		
Wi-Fi (<i>Wireless Fidelity</i>)	VLC (<i>Visible light communication</i>)		

4.4.1 Tecnologie per sistemi di localizzazione

Di seguito si forniscono le descrizioni delle tecnologie considerate per la scelta di sensori per la funzionalità di localizzazione.

4.4.1.1 Tecnologie Radio.

In questa categoria ricadono tutti quei sistemi di localizzazione indoor che utilizzano moduli radio come elementi di identificazione e di tracciamento. In questi sistemi si distinguono due tipologie di moduli in base alla loro funzione: i TAGS e le ANCORE. I Tags sono dispositivi agganciati ai target (oggetti/persona) di cui si vuole monitorare la posizione, mentre le Ancore sono dei moduli radio fissati alla struttura dell'edificio usati come punti di riferimento assoluto. Tag e Ancore instaurano dei *link* di comunicazione attraverso i quali è possibile stimare la distanza in modulo. Le Ancore sono connesse ad un sistema di elaborazione che utilizza le singole distanze Tag-to-Ancora, per localizzare il target mediante algoritmi basati su tecniche di triangolazione. Il calcolo può essere fatto dalla singola ancora o da un server a cui tutte le ancore sono connesse.

I metodi utilizzati per il calcolo della distanza si distinguono in Range-based e Angle-based.

Range-based

I metodi permettono una stima del modulo della distanza tra l'elemento trasmittente e quello ricevente basata sulle informazioni relative alla potenza del segnale radio Received Signal Strength Indication (RSSI) o al tempo di volo del segnale Time of Flight (TOF).

RSSI è uno dei metodi più utilizzati, come approfonditamente descritto in [Lu-011, Se-013, Pa-009]. Il parametro è fornito dal modulo ricevente, in particolare dallo stadio RF che, oltre ad implementare le funzioni di modulazione e demodulazione, misura la potenza del segnale radio rice-

vuto. Come noto dalla fisica, la potenza del segnale radio diminuisce all'aumentare della distanza tra modulo TX e RX secondo l'equazione di Friis.

L'utilizzo dell'RSSI è affetto da due tipologie d'errore:

- Riflessione: il segnale, propagandosi, sbatte e si riflette su vari ostacoli, seguendo più percorsi (vedi Figura 23)
- Assorbimento: alterazione del normale decadimento dell'intensità del segnale che si avrebbe nello spazio libero da ostacoli.

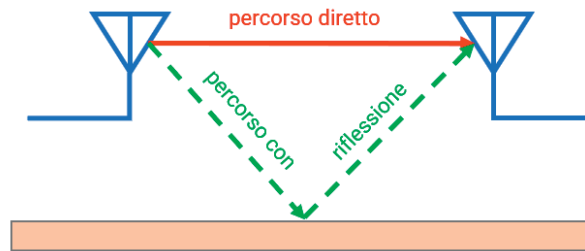


Figura 23 - Problema della riflessione

Il metodo TOF (Time of Flight) prevede la misura del tempo di volo di un segnale radio trasmesso dal tag ad ogni singola ancora. Questo tempo viene usato per calcolare la distanza di ciascun tag con le relative ancore. Per poter ottenere una buona risoluzione della distanza è fondamentale avere un *clock* molto preciso ed utilizzare delle tecniche di misura molto robuste. Tra queste le più diffuse sono Two Way Ranging (TWR) e la Time Difference of Arrival (TDoA).

La tecnica TWR prevede un'interazione tra Tag e Ancora in 3 fasi: POLLING (tag2ancora), RECEIVE (ancora2tag) FINAL (tag2ancora). Per ciascuna interazione viene monitorato il *timestamp* di avvio e ricezione delle diverse fasi. Lato Ancora il vettore di *timestamp* di partenza del messaggio e di arrivo viene inviato al sistema d'elaborazione il quale ricava la distanza di ciascun Tag con le relative Ancore. Conoscendo le distanze il sistema procede con la localizzazione del Tag.

La tecnica TDoA prevede l'invio di segnali radio periodici (denominati *Blink Signals*) da parte del Tag a tutte le ancore del sistema. Questi segnali sono degli impulsi molto brevi che arrivano a tutte le ancore in istanti differenti (in funzione della distanza reale di ciascun Tag con la relativa ancora). Il sistema di elaborazione, connesso alle ancore utilizza questi *timestamps* per calcolare le distanze e per effettuare la localizzazione finale del TAG. Questo approccio, a differenza del TWR richiede uno *slot* temporale della comunicazione inferiore, riducendo i consumi del Tag. Per questa modalità è però necessario che tutte le ancore siano interconnesse tra loro da una rete al fine di garantire un sincronismo spinto.

Angle based.

Sebbene l'indicatore dell'intensità del segnale ricevuto (RSSI) possa essere utilizzato per stimare la distanza da un punto fisso noto, spesso questa tecnica non è abbastanza precisa per applicazioni come un sistema di localizzazione indoor (IPS) e la tracciabilità delle risorse. Il nuovo standard Bluetooth 5.1 ha introdotto nuove funzioni di radiogoniometria: Angolo di arrivo (AoA) e Angolo di partenza (AoD), che consentono di determinare con precisione la posizione di un trasmettitore Bluetooth in due o tre dimensioni.

La tecnica radiogoniometria della radiofrequenza (RF) basata su RSSI fornisce la distanza appross-

simata in base all'intensità del segnale. Effettuando più misurazioni della distanza da punti diversi si può ottenere una maggiore precisione. L'RSSI ha il grande vantaggio di richiedere una sola antenna per dispositivo, eliminando la complessità, i costi e le dimensioni degli array di antenne. Lo svantaggio è una mancanza di precisione, che con questa tecnica è limitata a 3-5 metri.

AoA - Metodo di radiogoniometria AoA (Angolo di arrivo). Una seconda tecnica comune di radiogoniometria è nota come Tempo di arrivo (ToA), ovvero il tempo che un segnale radio impiega per arrivare da un singolo trasmettitore a un singolo ricevitore remoto. Anche questo metodo si avvale di una sola antenna per dispositivo, ma richiede che ogni dispositivo abbia un clock sincronizzato estremamente preciso, e questo è il rovescio della medaglia. La precisione della posizione per i sistemi ToA può avvicinarsi a 1 m. Con il rilascio della specifica Bluetooth 5.1, il Bluetooth Special Interest Group (SIG) ha scelto di supportare una terza tecnica di radiogoniometria basata su AoA e AoD. Con AoA, un dispositivo ricevente traccia gli angoli di arrivo per i singoli oggetti, mentre con AoD calcola la propria posizione nello spazio usando gli angoli di più beacon e le loro posizioni (Figura 24).

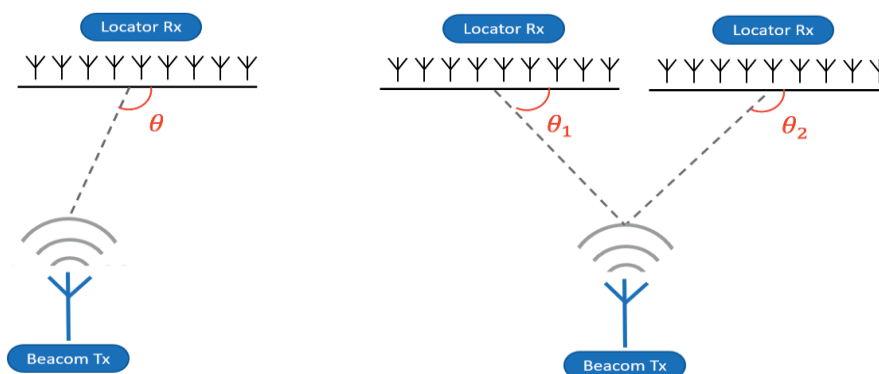


Figura 24 - principio di funzionamento dell'AoA (sinistra) e AoD (destra)]

Nel metodo di radiogoniometria AoA (Figura 24 a sinistra), le risorse trasmettono (TX) la propria posizione a un localizzatore AoA che misura l'angolo di arrivo del segnale. Con il metodo AoD (Figura 24 a destra), i beacon trasmettono informazioni AoD mentre un dispositivo mobile riceve (RX) i segnali dei beacon e calcola la posizione. In ogni caso, è il dispositivo ricevente ad aver bisogno della potenza computazionale per calcolare la direzione del trasmettitore.

La decisione di includere una funzione di radiogoniometria in Bluetooth 5.1 è fondata, in parte, sul fatto che esistono già alcune aziende che offrono soluzioni AoA e AoD proprietarie per prodotti Bluetooth Low Energy (BLE). Bluetooth 5.1 permette agli sviluppatori di sfruttare più facilmente la radiogoniometria RF grazie all'aggiornamento alle Core Specification mirato a semplificare l'estrazione dei dati del segnale "IQ" (informazioni in fase e in quadratura) dai pacchetti BLE. In questo modo, è per loro più semplice implementare le applicazioni comprendenti servizi di localizzazione.

Il metodo AoA è idoneo per tracciare un *transceiver* BLE che sta trasmettendo. Utilizzando una singola antenna, il *transceiver* invia pacchetti abilitati alla radiogoniometria che vengono ricevuti da un "localizzatore" a più antenne. Il localizzatore campiona i dati IQ dai pacchetti di segnali mentre passa da tutte le antenne attive dell'array. Questo gli permette di rilevare la differenza di fase del segnale dovuta alla differenza della distanza da ogni antenna dell'array alla singola anten-

na trasmittente. Il motore di localizzazione utilizza quindi le informazioni sulla differenza di fase per determinare l'angolo da cui sono stati ricevuti i segnali e quindi la direzione del trasmettitore (Figura 25).

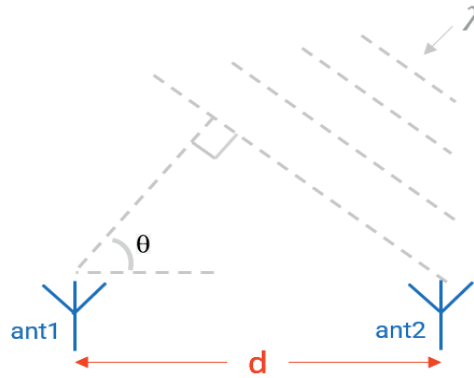


Figura 25 - L'angolo di arrivo di un segnale radio può essere calcolato se si conoscono la fase del segnale (θ) di ogni antenna, la lunghezza d'onda (λ) e la distanza (d) tra antenne adiacenti

La combinazione della direzione dei segnali calcolata da due o più localizzatori consente di individuare un trasmettitore (Figura 26).

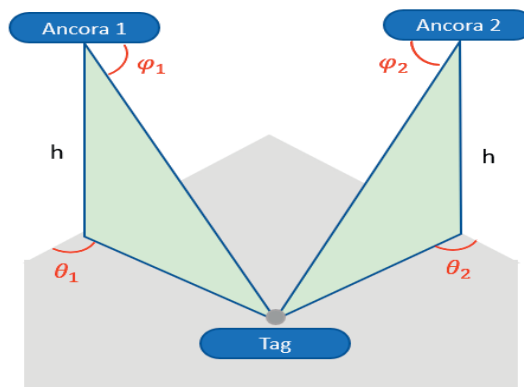


Figura 26 - esempio di localizzazione 3D

Calcolando l'AoA dei segnali in due localizzatori fissi, si può calcolare la posizione tridimensionale di una risorsa che sta trasmettendo. Se si conoscono le coordinate assolute dei localizzatori, è possibile calcolare anche le coordinate assolute della risorsa che trasmette.

AoD - Metodo di radiogoniometria AoD (Angolo di partenza). Nel caso del metodo AoD si ha la situazione inversa. In questo scenario, il dispositivo con l'array di antenne invia un segnale attraverso ognuna delle sue antenne. Quando ogni pacchetto di segnali delle antenne nell'array arriva alla singola antenna del ricevitore, a causa della diversa distanza percorsa a partire dal trasmettitore, risulta con una fase spostata rispetto al segnale precedente (Figura 27).

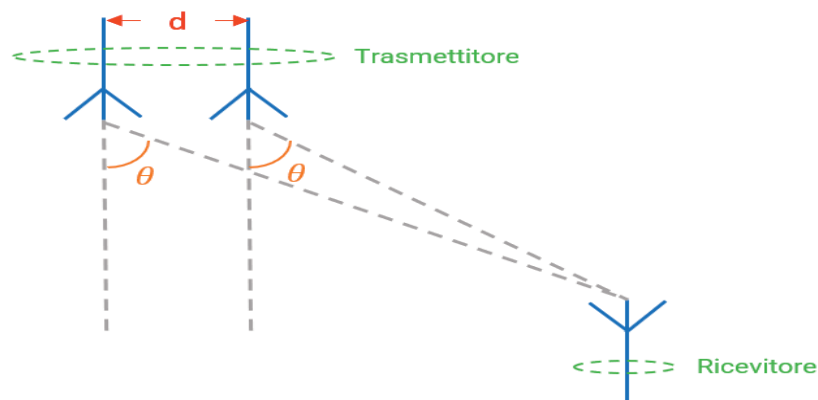


Figura 27 - Con il metodo AoD, quando ogni pacchetto di segnali delle antenne nell'array arriva alla singola antenna del ricevitore, a causa della diversa distanza percorsa a partire dal trasmettitore, risulta con una fase spostata rispetto al segnale precedente.

L'antenna del dispositivo ricevente preleva campioni di IQ dai pacchetti di segnali e li inoltra al motore di posizionamento, che utilizza poi i dati per determinare l'angolo da cui sono stati ricevuti i segnali e quindi la direzione del trasmettitore. Questo sistema è idoneo per applicazioni come la navigazione indoor in cui il trasmettitore è un punto di riferimento fisso e il ricevitore è, ad esempio, lo smartphone di un consumatore.

4.5 Sviluppo soluzioni real-time per situazioni a rischio

Nella progettazione del sistema complessivo si deve tenere conto non solo della rete di sensori necessari, ma dell'infrastruttura hardware complessiva necessaria alla trasmissione, *storage* e *processing* dei segnali. In particolare, per ridurre al minimo l'utilizzo di banda per l'invio dati verso il *cloud*, per rispettare la *privacy* ed implementare allarmi *real-time*, l'elaborazione dei segnali è stata progettata per essere effettuata in loco. Grazie ad una abbondanza di risorse in fase di dimensionamento, sono state implementate anche tecniche di *machine learning* per l'estrazione di *features* dalle immagini.

Tra le attività realizzate, in aggiunta sono state considerate la gestione del flusso di dati tra i vari sistemi di elaborazione (locali o su *cloud*) nonché alcune operazioni di *post-processing* (ad es. estrazione delle *feature* dal flusso video, reportistica dei dati acquisiti). La letteratura scientifica offre, in tale ambito, interessanti lavori ed esempi quali [Si-022, Bu-021, La-021, Ga-020, De-018, Bi-016, Mu-014, Bu-014, Ch-009] da cui è stato tratto spunto per la definizione dei sistemi sviluppati.

4.5.1 Metodologie di analisi dati per derivare informazioni sul contesto

Da letteratura, gli algoritmi per il controllo *real-time* di comportamenti a rischio e utilizzi non consoni di spazi e apparecchiature si distinguono nelle seguenti categorie (riportate nella Figura 28): single input booleano (A), single input proporzionale (C), multiple input booleano (B), multiple input proporzionale (D).

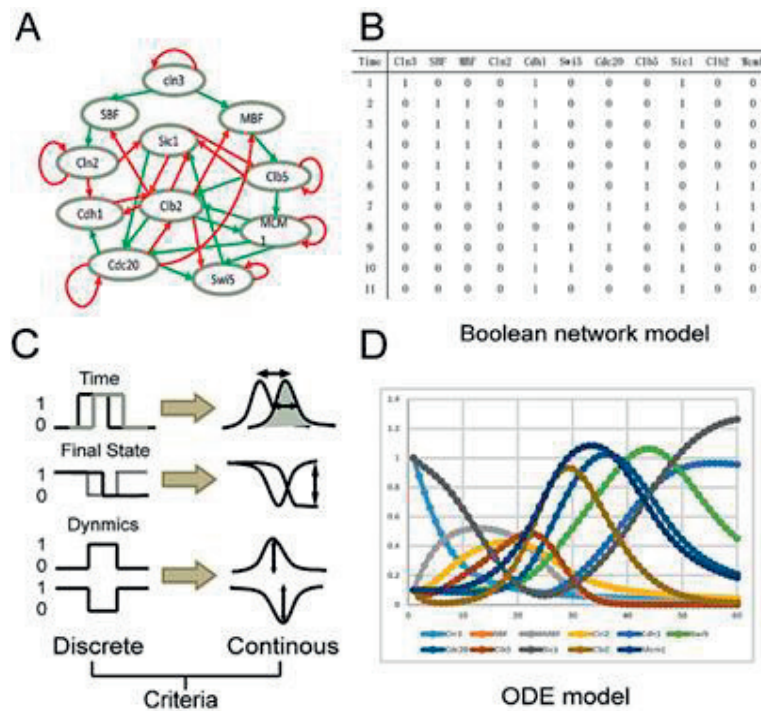


Figura 28 - Le quattro famiglie di algoritmi per il controllo real-time di comportamenti a rischio e utilizzi non consueti di spazi e apparecchiature

Gli algoritmi, rielaborando i dati dei sensori, sono in grado di rilevare e fornire un *warning* in tutte le situazioni di pericolo evidenziate dalle tre aziende prese in considerazione per i Case Lab. In particolare:

- nelle interazioni uomo-macchine mobili (segnalate da tutte e tre le aziende) si sfrutta l'algoritmo **single e multiple input proporzionale**, in quanto si individuano la posizione relativa e la velocità di avvicinamento tra il singolo operatore e il macchinario mobile in modo da identificare l'eventuale collisione e l'intervallo di tempo nel quale tale collisione potrà avvenire;
- nelle interazioni uomo-macchine fisse (segnalate da tutte e tre le aziende) si utilizza l'algoritmo **single input booleano** in quanto l'obiettivo è quello di identificare se l'operatore (eventualmente non autorizzato) si avvicina troppo a una macchina fissa; nel qual caso si manda all'operatore stesso un messaggio di *warning*;
- anche nel caso di ingresso in aree ad accesso limitato o aree pericolose vietate a determinate categorie di persone (segnalate da due delle tre aziende considerate come Case Labs) si sfrutta l'algoritmo **single input booleano**; in effetti questo è un caso molto simile al precedente (se all'area ad accesso limitato si sostituisce la macchina fissa ricadiamo nel caso precedente);
- nel caso di lavoro in quota (segnalato da due delle tre aziende considerate) si adotta l'algoritmo **single input booleano**; rispetto ai due casi precedenti si sfrutta la capacità del sistema di monitoraggio di fornire non solo la posizione nel piano ma anche la sua elevazione; è quindi possibile, inizializzando opportunamente delle soglie, introdurre un segnale di allarme robusto in grado di segnalare quando l'operatore sta salendo in quota, non quando si avvicina alla scala;
- nel caso di movimenti/azioni scorrette o parametri vitali scorretti in relazione all'attività svolta si utilizza l'algoritmo **single input proporzionale** che, permettendo di seguire il movimento

- dell'operatore, è in grado di valutare tali azioni scorrette in prossimità di macchinari pericolosi e/o aree critiche; si noti che il monitoraggio del movimento avviene in corrispondenza del braccialeto sensorizzato; è quindi parziale (non si monitora il movimento degli altri arti) ma fornisce comunque una valida indicazione se qualcosa non sta andando come dovrebbe;
- nel caso di mancata localizzazione di personale esterno (auditor, squadre di manutenzione, ...) o personale interno (lavoratore in solitaria, ...) si utilizzando le telecamere opportunamente posizionate e l'algoritmo **multiple input booleano** basato su *machine learning* per l'estrazione delle informazioni salienti dal flusso video; in particolare, individuata la persona nel campo visivo, si verifica che ci sia un segnale del braccialeto in corrispondenza di quella zona; in assenza di corrispondenza di segnale video e braccialeto, viene emesso un segnale di warning per richiamare l'attenzione a chi entra nella zona supervisionata di dotarsi di braccialeto;
 - anche nel caso di mancato uso o uso scorretto di DPI (segnalato da due aziende) si utilizzando le telecamere in maniera analoga al caso precedente; in questo caso però, si verifica che la persona individuata abbia i DPI prescritti; va sottolineato che tale *feature* non è stata completamente implementata (anche se sarebbe banale farlo) a causa della necessità di telecamere ad alta risoluzione e a maggiore copertura, aspetto questo attualmente non consentito dagli accordi con le parti sociali;
 - nel caso di mancata localizzazione in condizioni di emergenza (situazione critica evidenziata da un'azienda) si è adottato l'algoritmo **multiple input proporzionale**, che è in grado di rilevare la posizione di tutti i braccialetti nella zona supervisionata; tale sistema è ovviamente efficace qualora non ci siano persone prive di Tag all'interno di tale zona; ecco perché va accoppiato al sistema di visione degli accessi precedentemente descritto;
 - nel caso di uso illecito di determinati macchinari (anche questo aspetto è stato segnalato da un'azienda) si è adottato l'algoritmo **single input booleano**, che consente di verificare che l'operatore che sta lavorando su una data macchina abbia le abilitazioni per il suo utilizzo; per evitare falsi positivi, si considera non solo la vicinanza relativa ma anche il tempo di vicinanza;
 - infine, nel caso di non corretta messa in sicurezza dei macchinari (altro aspetto segnalato da un'azienda) si adottano telecamere e una logica di **multiple input proporzionale** che si attiva ad orari prestabiliti (o a fronte di date azioni degli operatori) e, analizzando il flusso video, consente di verificare che tutte le sicurezze della macchina monitorata siano attive.

Gli algoritmi di controllo rientranti nelle prime tre categorie (1,2,3) sono stati immediatamente implementati, mentre quelli rientranti nella categoria (4) hanno richiesto un'indagine preliminare dei dati acquisiti prima di una effettiva implementazione.

Si noti come il sistema di sensorizzazione e gli algoritmi sviluppati siano molto più generali rispetto alle situazioni di pericoli evidenziate dalle aziende. Infatti, con piccole modifiche e ri-tarature dei parametri (in particolare delle soglie di intervento) è possibile adattare l'algoritmo a innumerevoli altre situazioni di pericolo, anche tenendo conto dei tempi di intervento dei sensori e/o dell'accuratezza della localizzazione che si ha a disposizione con il sistema implementato.

4.5.2 Progettazione di sistemi di allarme real-time

Nella progettazione dei sistemi di controllo e allarme in tempo reale sono state prese in considerazione due differenti situazioni: quella che vede come soggetto dell'attività a rischio un operatore e

quella che invece vede la causa della situazione di rischio in una macchina operatrice (Figura 29). Se il comportamento a rischio e l'utilizzo non consoni di spazi e apparecchiature sono dovuti ad un operatore, *l'output* dell'algoritmo di controllo è diretto all'operatore stesso e l'evento anonimizzato viene registrato su *cloud*. Questo *feedback* è stato ritenuto il più consoni alla situazione. La registrazione su *cloud* permette ai gestori della sicurezza di mettere a punto strategie di miglioramento della sicurezza nel caso in cui un comportamento a rischio si verifichi più volte.

Se il rischio è relativo a macchinari, *l'output* dell'algoritmo di controllo è diretto a tutti gli operatori nei pressi del macchinario e al responsabile manutenzione/sicurezza. Comunque, anche qui l'evento viene registrato su *cloud*, in questo caso con l'obiettivo di permettere la definizione di procedure che riducano le attività a rischio.



Figura 29 - Uno scenario di applicazione degli algoritmi per il controllo real-time di comportamenti a rischio e utilizzi non consoni di spazi e apparecchiature [shutterstock]

4.6 Sensorizzazione del contesto nei Case Labs

Con l'obiettivo di installare il *backbone* dell'infrastruttura sensoristica nell'arco temporale del progetto, si è optato per una rete di sensori commerciali, che però avesse sufficiente flessibilità ed apertura per poter accogliere anche sensori di prossima generazione, siano questi ancora di tipo commerciale o sviluppati custom per specifiche grandezze da misurare o situazioni molto particolari da monitorare. A titolo meramente esemplificativo si pensi a sensori in grado di rilevare direttamente forti campi magnetici, che possono essere presenti in particolari macchine o in situazioni di malfunzionamento delle macchine stesse.

Le flessibilità e apertura della rete di sensori è ottenuta adottando un *bus* di campo asincrono (con *timestamp*) per la comunicazione dei dati acquisiti dai singoli sensori. Pertanto, purché il

sensores abbia un'uscita digitale (aspetto praticamente universale per i sensori oggi in commercio) e permetta un minimo di programmazione del protocollo di comunicazione, qualsivoglia sensore può essere collegato al *backbone* sviluppato. Inoltre, l'adozione di un bus digitale: (i) rende il segnale proveniente dai sensori praticamente immune da problemi di lunghezza dei cavi e di rumori elettrici e magnetici che potrebbero disturbare la trasmissione di segnali analogici; (ii) permette la supervisione del corretto funzionamento dei sensori e della rete rendendo così robusta l'implementazione a danni/difetti della stessa; (iii) consente la tracciabilità delle informazioni, ognuna dotata di *timestamp* e *recordID*. Infine, il bus digitale non richiede un supporto fisico, come invece richiesto nel caso di rete analogica: la rete potrebbe essere sia cablata fisicamente che *wireless* - laddove il cablaggio risultasse non conveniente per i più svariati motivi.

Per quanto concerne l'alimentazione dei sensori stessi, avendo, come detto, deciso di adottare soluzioni commerciali, essa si basa sulla rete elettrica già disponibile in loco. Chiaramente, laddove l'installazione della rete di sensori avvenisse in ambienti dove una rete elettrica non è disponibile (ad esempio in ambienti esterni o remoti), è indispensabile aggiungere alla rete stessa un sistema di alimentazione della sensoristica. Tale sistema di alimentazione va studiato caso per caso, in modo da sfruttare al meglio le diverse forme di energia disponibili: se ad esempio alcuni sensori venissero installati all'esterno si potrebbe sfruttare la luce solare; se alcuni sensori venissero installati su macchine vibranti si potrebbe recuperare l'energia per alimentarli dall'energia meccanica vibrazionale della macchina stessa; etc.

Grazie a diversi sopralluoghi svolti presso i Case Labs, è stato possibile realizzare un progetto esecutivo della rete di sensori e dell'*hardware* per l'elaborazione dei segnali, tradotto in capitolato e proposto a società che vendono e installano sensori (*system integrators*) con la richiesta di un preventivo.

A titolo di esempio, la Figura 30 riporta lo schema della disposizione dei sensori presso un'azienda dislocata su due stabilimenti adiacenti e con la necessità di transito di merci e persone da un edificio all'altro.

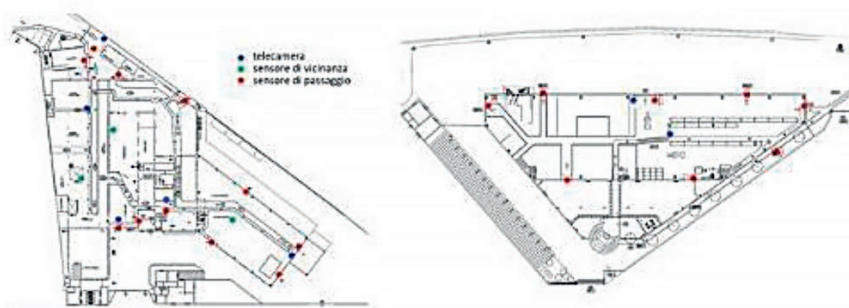


Figura 30 - Schema esempio di una possibile disposizione dei sensori presso un'azienda dislocata su due stabilimenti. L'azienda in questione non fa parte dei casi studio del progetto.

Come anticipato la sensorizzazione è pensata per essere in grado di individuare le situazioni di pericolo elencate precedentemente. Sono quindi stati installati sensori di distanza con tag ID presso:

- tutti gli ingressi e le uscite dei capannoni,

- gli incroci ciechi,
- i macchinari fissi che permettono lavorazioni in quota,
- i macchinari fissi che generano rumore tale da richiedere l'utilizzo di specifici DPI,
- i macchinari fissi che generano campi elettromagnetici pericolosi per persone con problemi cardiaci,
- i macchinari in movimento che devono essere manovrati da personale specifico.

Per garantire inoltre che tutto il personale che entra nei capannoni sia dotato di sistema di riconoscimento e verifica degli opportuni DPI indossati, in corrispondenza degli ingressi è prevista l'installazione di telecamere con elaborazione in locale dello *streaming* video come da specifiche descritte in precedenza.

Infine, all'interno dei capannoni, di comune accordo con il titolare dell'impresa e il responsabile della sicurezza, si devono individuare le postazioni dei sistemi hardware di post-processing dei dati provenienti dai sensori distribuiti, onde minimizzare i tempi di intervento per gli allarmi in tempo reale e filtrare le informazioni non necessarie prima di inviare sul *cloud* i dati che verranno utilizzati per la definizione di nuovi paradigmi di sicurezza.

4.7 Descrizione del sistema di localizzazione indoor fornito da NewEN.

A valle delle analisi fatte sui potenziali infortuni sul lavoro legati alle collisioni tra operatori e mezzi di movimentazione (carrelli e muletti), si è deciso di procedere con la selezione di un sistema di localizzazione indoor non invasivo, in grado di monitorare, in tempo reale, la posizione degli operatori e delle macchine di movimentazione. In questa parte del documento presentiamo un'analisi dello stato dell'arte dei sistemi di localizzazione indoor radio e una descrizione del sistema finale installato dall'azienda NewEN nelle aziende coinvolte nel progetto RECKON.

Nell'ambito del progetto RECKON, l'azienda NewEN ha svolto un ruolo di fondamentale importanza nella fornitura ed installazione di 3 sistemi di localizzazione indoor. Ciascun sistema è basato su moduli radio BLE (Bluetooth Low Energy) che utilizzano contemporaneamente le tecniche del Ranging e AoA per la localizzazione dei tags.

L'intero sistema è composto da: Tags attivi BLE, Ancore, Server.

I Tags attivi BLE sono piccoli sistemi *embedded ultra-low-power*, composti da un modulo radio BLE (con antenna realizzata su pcb) una MCU, una IMU ed un segnalatore acustico - visivo. Alcuni di questi Tags sono indossabili (*wearable*) dagli operatori per il loro tracciamento. Altri invece vengono usati per monitorare la posizione dei mezzi di movimentazione.

Le Ancore contengono al loro interno dei moduli radio (vedi Figura 31) che utilizzano array di antenne per la stima del RSSI e dell'AoA. Le ancore sono connesse ad una rete LAN attraverso la quale vengono anche alimentate mediante l'utilizzo del PoE.

Le Ancore, attraverso la rete LAN, comunicano con un Server principale che si occupa dell'elaborazione delle informazioni fornite da ciascuna Ancora. L'elaborazione porta come risultato la ricostruzione della posizione del Tag in uno spazio 2D con una risoluzione dell'ordine dei 10cm ed un aggiornamento ogni 100ms.



Figura 31 - Tags e antenne fornite dall'azienda NewEn

Le aree delle 3 aziende selezionate (Figura 32) hanno dimensioni diverse e caratteristiche strutturali differenti. Questo ha reso possibile dimostrare l'efficacia del sistema in contesti eterogenei, aventi condizioni operative e strutturali diverse.

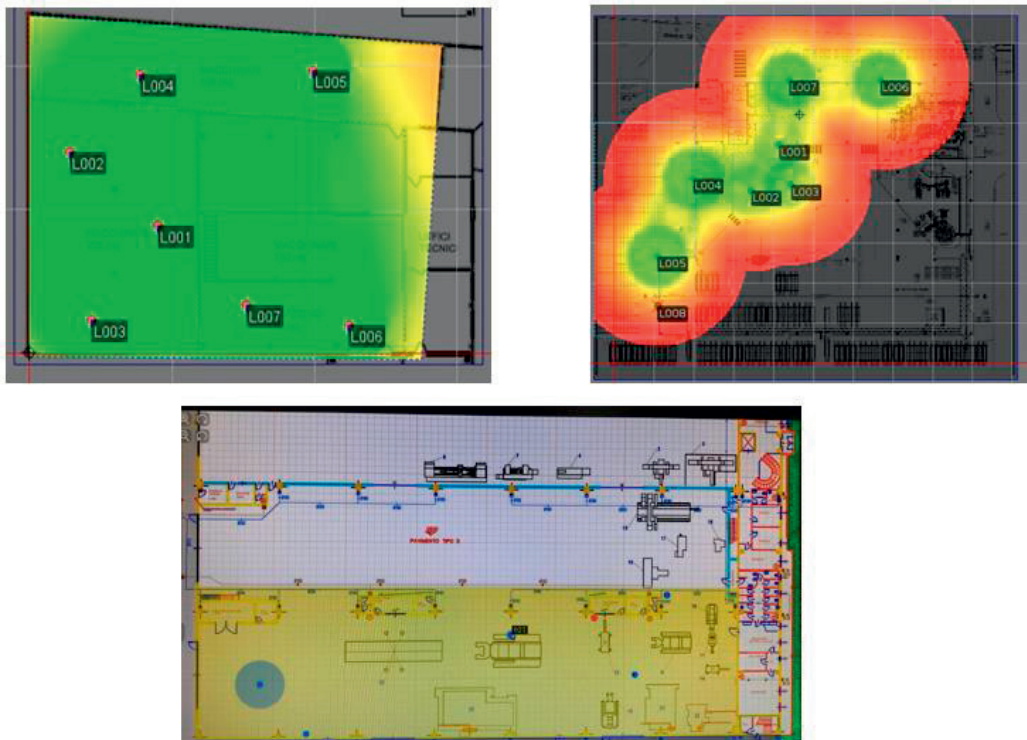


Figura 32 - mappature delle aree aziendali soggette al monitoraggio

5 Casi Studio e sperimentazione

In questo capitolo si intende descrivere la sperimentazione presso le aziende dei risultati del progetto RECKON.

La scelta e il coinvolgimento delle PMI nel settore metalmeccanico sono stati resi possibili grazie alla presenza di ApiTech tra i partner del progetto RECKON. Sono state selezionate tre aziende per far parte dei Case Labs su cui sperimentare i risultati del progetto, cioè fornire la sensorizzazione del contesto operativo. Tutte le situazioni di pericolo mappate dalla letteratura e dall'analisi degli incidenti hanno trovato riscontro nei Case Lab.

5.1 L'implementazione di RECKON nelle singole aziende

Di seguito si riportano i tre Case Lab e le situazioni di pericolo in essi individuate. La sensorizzazione di queste tre realtà industriali pone diverse sfide, sia di tipo logistico che di tipo tecnologico.

Azienda 1. (Case Lab 1)

L'Azienda 1 è organizzata su due stabilimenti (1 e 2) che impiegano rispettivamente 100 e 145 persone. Il Case Lab viene sviluppato a partire dallo stabilimento 1, a sua volta dislocato su due edifici adiacenti. Il settore di riferimento è quello della lavorazione della lamiera, caratterizzata nel caso dell'Azienda 1 da operazioni quali tranciatura, piegatura, imbutitura e foratura.

- Interazioni uomo - macchine mobili (focus carrelli)
- Interazioni uomo - macchine fisse (focus su area di trattamento termico)
- Mancata localizzazione di personale esterno (auditors, squadre di manutenzione, ...) o personale interno (lavoratore in solitaria, ...) (focus lavoratore in solitaria)
- Lavoro in quota (caduta dall'alto) (focus su operazioni in area trattamento termico)
- Mancato uso o uso scorretto di DPI nonostante l'informazione e la formazione ricevute dall'azienda (focus su DPI per area di trattamento termico)
- Movimenti/azioni scorrette o a parametri vitali scorretti in relazione all'attività svolta (focus lavoratore in solitaria)
- Parametri ambientali che hanno registrato valori profondamente diversi da quelli normalmente mantenuti in azienda (focus area sotterranea in cui operano i lavoratori in solitaria)

Azienda 2. (Case Lab 2)

L'Azienda 2 è caratterizzata da un unico stabilimento che impiega 185 persone. Il settore di riferimento per l'Azienda 2 è quello della produzione di cuscinetti. In particolare, realizza cuscinetti a rulli SRB, con una capacità produttiva di circa 900.000 cuscinetti/anno.

Situazioni di pericolo:

- Interazioni uomo-macchine mobili (focus manipolatori)
- Interazioni uomo-macchine fisse (focus su presse)
- Ingresso in aree ad accesso limitato o aree pericolose vietate a determinate categorie di persone (focus aree raffreddamento semilavorati)
- Lavoro in quota (caduta dall'alto) (focus su operazioni di manutenzione in quota su - singola macchina).
- Movimenti/azioni scorrette o a parametri vitali scorretti in relazione all'attività svolta.

Azienda 3. (Case Lab 3)

L'Azienda 3 è organizzata su due stabilimenti (1 e 2). Il Case Lab in questo caso è rappresentato dal secondo stabilimento al cui interno sono impiegati 30 dipendenti.

Il settore di riferimento per l'Azienda 3 è quello della forgiatura e delle successive lavorazioni meccaniche su una vasta gamma di materiali, inclusive delle operazioni di tornitura, fresatura, foratura, alesatura, rettifica. L'azienda è inoltre attrezzata per svolgere attività di riparazione ed assemblaggio, ma anche trattamenti termici quali ricotture, normalizzazioni, rinvenimenti, bonifiche e solubilizzazioni.

Situazioni di pericolo:

- Interazioni uomo-macchine mobili (focus su carrelli).
- Interazioni uomo-macchine fisse (focus presse).
- Ingresso in aree ad accesso limitato o aree pericolose vietate a determinate categorie di persone (focus su area con alto campo magnetico).
- Uso illecito di determinati macchinari (focus su carrelli).
- Mancata localizzazione di personale esterno (auditors, squadre di manutenzione, ...) o personale interno (lavoratore in solitaria, ...).
- Mancata localizzazione nel caso di emergenze (emergenze = incendi, terremoti, esalazioni di gas tossici, ...).
- Lavoro in quota (caduta dall'alto) (focus su operazioni di manutenzione in quota su singola macchina).
- Non corretta messa in sicurezza dei macchinari (focus su presse).
- Mancato uso o uso scorretto di DPI nonostante l'informazione e la formazione ricevute dall'azienda.

5.2 Sensorizzazione di un Case Lab

In questo paragrafo analizziamo nel dettaglio la sensorizzazione di una delle tre aziende che chiameremo Azienda 3. Descriveremo in particolare il contesto dell'azienda, i suoi punti critici e la predisposizione della sensoristica prevista al suo interno.

Dal punto di vista logistico, Azienda 3 presenta un capannone "aperto" ossia con possibilità di accesso da innumerevoli punti: aperture necessarie per consentire la movimentazione di oggetti di grandissime dimensioni come quelli lavorati dall'azienda. Dal punto di vista tecnologico, invece, Azienda 3 è dotata di forni per scaldare i particolari, e di forge di dimensioni rilevanti che non solo creano stress termico alla strumentazione ma generano anche schermature e angoli ciechi e richiedono una copertura ad hoc.

La verifica della sensorizzazione è stato un aspetto chiave dell'installazione dei sistemi per la localizzazione. Questa attività ha richiesto la definizione di una procedura di verifica del sistema che è stata poi applicata a valle dell'installazione per validarne l'efficacia e l'efficienza. In questa fase è stata fondamentale una stretta integrazione con le aziende coinvolte.

L'Azienda 3, come già detto, è dislocata su un solo capannone; l'area di interesse per il monitoraggio non ricopre tutto il capannone ma solo una sua parte, in quanto l'altra parte è adibita a zona di stoccaggio o piccole lavorazioni meccaniche. Questa interessante caratteristica permette di verificare la flessibilità del sistema proposto, che consente di monitorare sotto-aree anche potenzialmente "aperte": questa possibilità riveste grande importanza in diversi ambienti lavorativi. Le Figure 33 e 34 mostrano gli accessi alla zona da monitorare, caratterizzata dalla presenza di due forni per il riscaldamento dei pezzi e di due forge di cui una da più di 30.000 tonnellate.

Come si può osservare, gli accessi alla porzione di capannone di interesse per il monitoraggio sono

9, indicati dai triangoli blu. Si notino i tre triangoli blu nella parte superiore dell'immagine, che monitorano l'accesso alla zona di interesse rispetto alla restante parte del capannone non oggetto del monitoraggio. Le paratie divisorie tra queste due porzioni di capannone possono essere movimentate, le telecamere, quindi, andranno posizionate in modo tale da riuscire a rilevare l'intera parete divisoria.

La zona di interesse viene coperta mediante un sistema di localizzazione, tra operatori e tra operatori e macchine, con antenne posizionate in modo tale da avere due diverse accuratèzze di localizzazione: 50 cm nelle zone a piú elevato rischio (ossia quelle in prossimità dei forni, delle forge e dei muletti che movimentano i pezzi) e 1 m in tutte le altre zone.

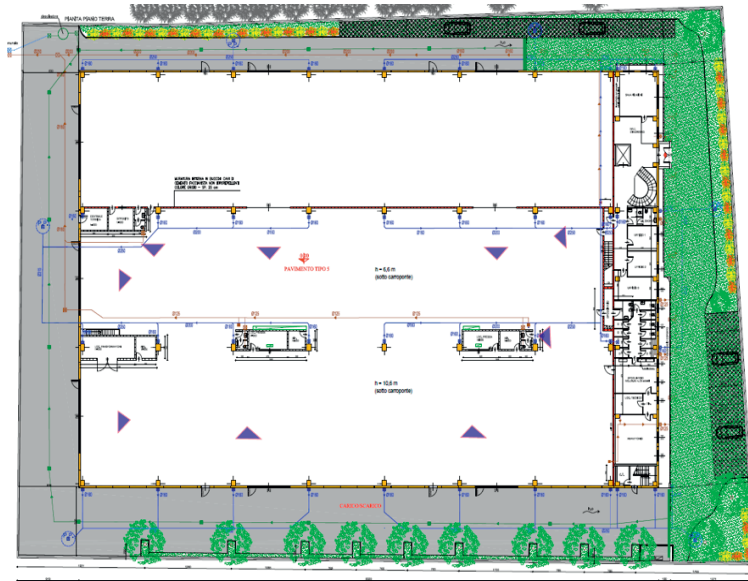


Figura 33 - Vista in scala dell'Azienda 3

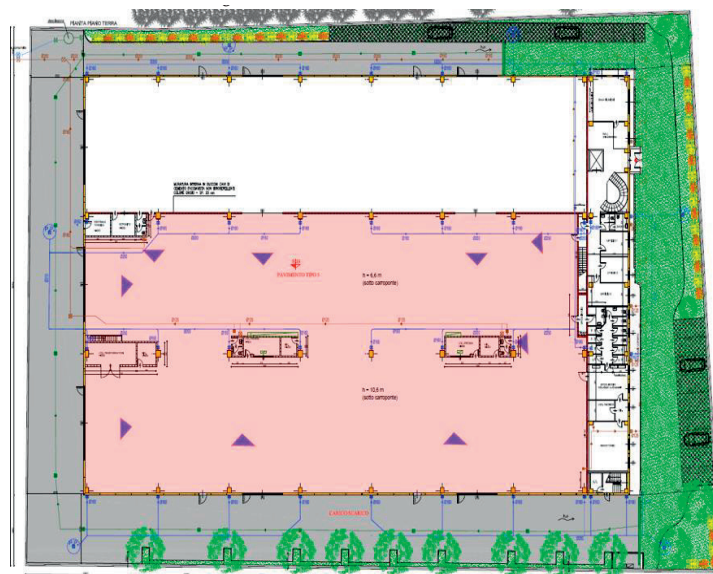


Figura 34 - Sensorizzazione dell'Azienda 3

5.2.1 Osservazioni

Le attività descritte si sono concentrate sulla formalizzazione della componente sensoristica del sistema e sulla definizione delle installazioni presso i Case Labs.

In particolare, per quanto riguarda i sensori, l'attività di ricerca si è concentrata sulla localizzazione tramite tecnologia *bluetooth*. Altre soluzioni tecnologiche sono state prese inizialmente in considerazione, ma la scelta è poi andata proprio sulla tecnologia *bluetooth* per rispondere ai vincoli applicativi e tecnologici del progetto e per accogliere gli spunti arrivati dai Case Labs.

Per quanto riguarda le aziende coinvolte nei Case Labs, l'attività di ricerca ha portato alla finalizzazione degli scenari applicativi e alla definizione delle installazioni, che si appoggiano alla soluzione tecnologica definita a partire dalle scelte sulla sensoristica precedentemente descritte.

5.3 Implementazione delle soluzioni nei Case Labs

Gli algoritmi descritti in precedenza sono stati implementati, integralmente o in parte, nei diversi Case Labs in funzione della dotazione sensoristica utilizzata.

Va sottolineato come, a causa delle diverse soluzioni tecnologiche implementate al fine di vagliare il rapporto costi-benefici delle stesse, pur essendo gli algoritmi essenzialmente gli stessi per le tre applicazioni, è stato necessario modificare le soglie di intervento. Ad esempio, poiché la risoluzione spaziale e il tempo di *refresh* del sistema di localizzazione *indoor* in ciascuna azienda sono diversi (vedi Figura 35), è stato necessario prevedere un intervento del sistema di allerta anticipato, nonché distanze relative maggiorate. Ciò nonostante, gli algoritmi si sono dimostrati robusti e flessibili. In conclusione, si può quindi affermare che, indipendentemente dell'investimento che l'azienda decide di sostenere, il *software* di controllo resta sostanzialmente invariato ma viene richiesto un tempo minimo di messa a punto (*fine tuning*).



Figura 35 - Esempio di monitoraggio della posizione relativa degli operatori autorizzati rispetto ad una macchina fissa (forgia) che lavora elementi incandescenti [wikipedia].

5.4 Verifica delle soluzioni real-time nei Case Labs

In questo paragrafo anticipiamo che i risultati evidenziano come gli algoritmi messi a punto siano molto robusti al variare della soluzione tecnologica adottata e della situazione di rischio da monitorare. Si conclude quindi che la scelta fatta di adottare strumentazione *off-the-shelf* e ridondante (ogni operatore e ogni macchina mobile devono essere dotati di Tag, ogni accesso e ogni macchinario critico o che richieda particolari DPI deve essere monitorato da un sistema di telecamere) ha permesso di sviluppare un sistema di monitoraggio estremamente flessibile e trasversale.

5.4.1 Osservazioni

Le attività mostrate hanno permesso la messa a punto e la verifica di una serie di algoritmi di controllo in tempo reale per allarmi in caso di situazioni di rischio imminente. È stato ideato un *framework* logico per la gestione di questi allarmi che consente grande robustezza in termini di soluzione tecnologica adottata (risoluzione spaziale della localizzazione e ritardo nella messa a disposizione dell'informazione di distanza e velocità relativa) e grande flessibilità nel monitorare le diverse situazioni di rischio. Ovviamente, le soglie di intervento del sistema dipendono dalla tipologia e sensibilità della sensoristica adottata nonché dall'ambiente che deve essere coperto da tale sistema.

5.5 Approfondimento tecnologico: WebApp e casi di studio

Il Sistema sopra descritto si completa infine di una *WebApp* messa a disposizione degli operatori Inail, alla quale è possibile accedere tramite username e password.

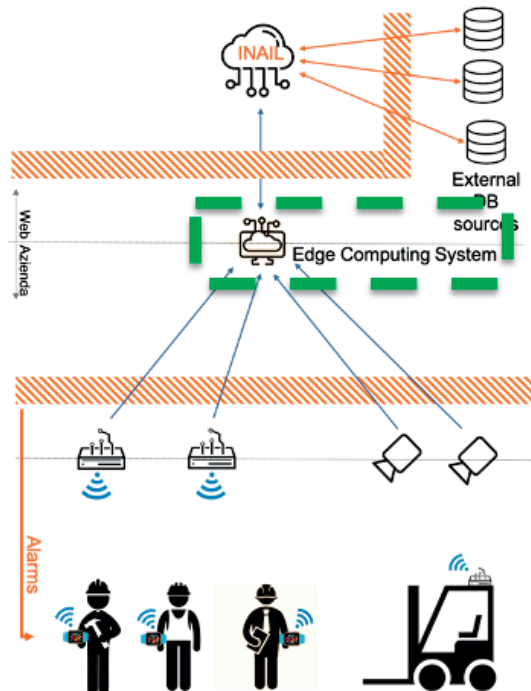


Figura 36 - Architettura RECKON. La WebApp lavora a livello dell'Edge Computing System

La *WebApp* ha le seguenti funzionalità:

- visualizzazione su mappa della posizione dei tag e dei muletti con colori diversi e delle aree a rischio
- invio notifiche in caso di situazioni pericolose (es. ingresso di un tag nel raggio di azione di un macchinario a rischio)
- invio comandi ai tag per vibrare e attivare il buzzer
- registrazione delle traiettorie seguite dalle persone e dai muletti

La Figura 36 mostra le principali interazioni della *WebApp* con il sistema di monitoraggio.

5.5.1 Esplorazione dei dati sensoristici e implementazione query PerLa

Come spiegato nella Sezione 2.2, le query PerLa possono essere tradotte in query SQL per interrogare i database forniti dal sistema descritto. Il linguaggio PerLa, infatti, consente di specificare le modalità di interrogazione della base di dati sensoristici, indipendentemente dalla natura delle fonti, spesso eterogenee. Nel nostro caso, le specifiche definite con le query PerLa hanno guidato la realizzazione di un database Mariadb (MySQL).

All'interno del database ci sono cinque viste:

- *vs_alarms*: contiene tutti gli allarmi di collisione per tutte e tre le aziende
- *vs_alarms_device*: aggiunge l'informazione sul tipo di device coinvolto nella possibile collisione
- *vs_rawDataUno*: contiene i dati grezzi di rilevazione tag per Azienda 1
- *vs_rawDataDue*: contiene i dati grezzi di rilevazione tag per Azienda 2, *vs_rawDataTre*: contiene i dati grezzi di rilevazione tag per Azienda 3

La Figura 37 mostra il risultato della query:

```
1) select *
from Vs_rawDataAzienda
order by tag, timestamp desc limit 100
```

che permette di visualizzare tutti i dati (compresa la posizione nel tempo) relativi ai Tag dell'azienda.

timestamp	tag	x	y	area	color	name	type	use
1658935292306	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292299	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292296	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292270	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292255	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292251	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292229	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292210	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292206	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292193	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292189	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292176	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292156	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292139	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292117	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier
1658935292102	ca0400048304	61.7300	21.8600	Produzione	#0066FF	015	Forze+	carrier

Figura 37 - Esempio query sul database. Informazioni sui Tag di una singola azienda

Grazie alle viste attualmente presenti nel dataset è inoltre possibile ottenere informazioni utili per analisi a posteriori. Ad esempio, una possibile query sulla vista *vs_alarms_devices*, è la seguente:

```
2) select idAlarm, area, device, deviceUse
from vs_alarms_devices
where deviceUse = 'people'
order by idAlarm desc limit 100;
```

Con questa query si possono estrarre, per ogni potenziale collisione che ha coinvolto almeno una persona, l'area in cui la collisione stava avvenendo e l'id del dispositivo indossato dalla persona. La Figura 38 riporta un esempio di risultato per una delle tre aziende.

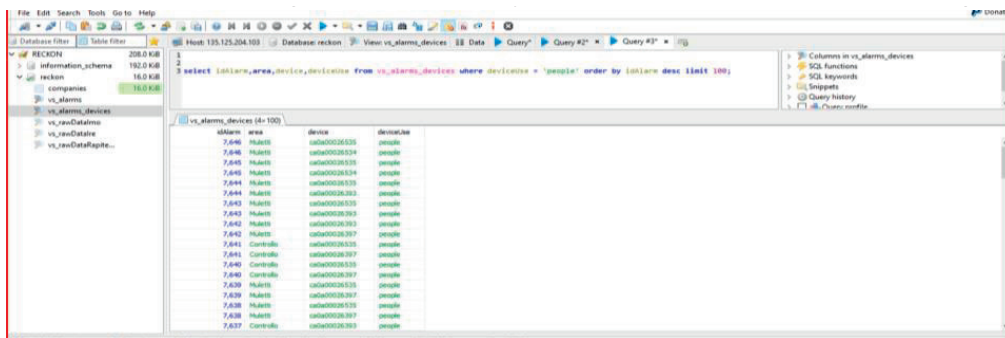


Figura 38 - Esempio query sul database. Informazioni su possibili collisioni singola azienda

A titolo esemplificativo viene mostrata in Figura 39 la mappa delle possibili collisioni di una delle tre aziende monitorate durante il progetto. È molto interessante notare che, come atteso, i risultati mostrati dalla *WebApp* evidenzino come in azienda siano presenti delle aree critiche maggiormente soggette a potenziali collisioni. Questo è esattamente il tipo di risultato atteso dal progetto che può servire sia nella prevenzione real-time che come analisi strategica da parte dell'azienda.

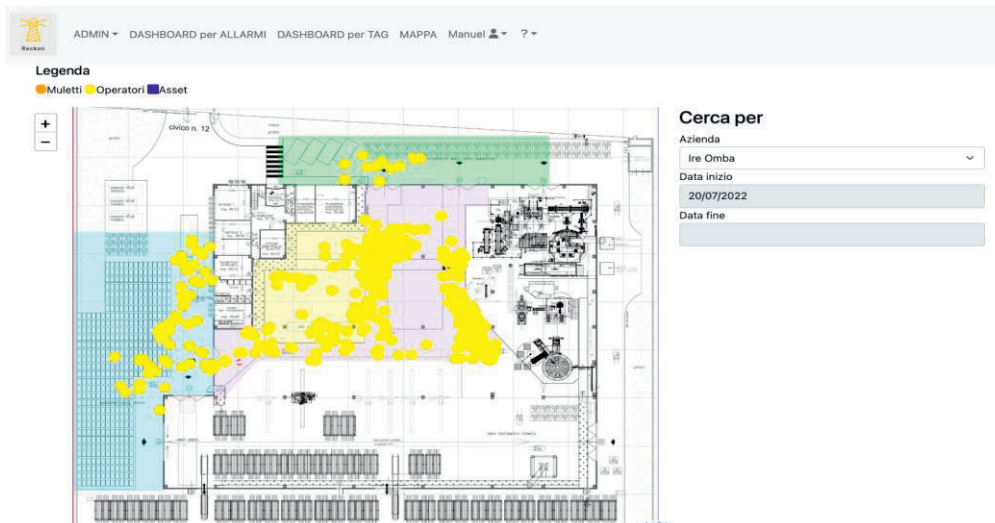


Figura 39 - Le possibili collisioni registrate in una singola azienda

Utilizzando i dati provenienti dalla vista *vs_alarms_devices* sono state prodotte delle analisi preliminari sui dati ottenuti nel periodo da luglio a ottobre 2022 per le tre aziende. In particolare, si è cercato di rispondere a tre domande per una visione strategica:

1. Quali **ore del giorno** sono più critiche?
2. Quali **giorni della settimana** sono più critici?
3. Quali **zone dell'azienda** sono più critiche?

I risultati ottenuti vengono descritti nei paragrafi seguenti.

1. Quali **ore del giorno** sono più critiche?

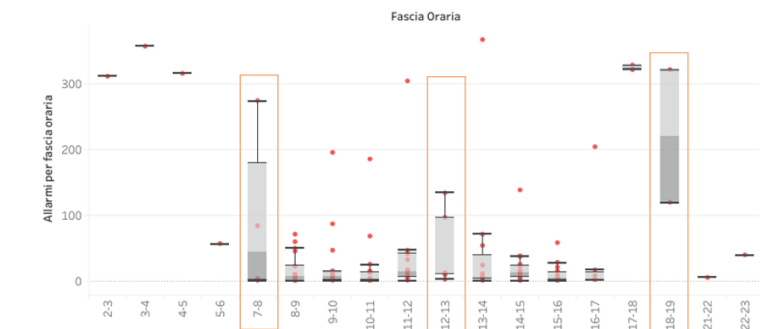


Figura 40 - Quali ore del giorno sono più critiche? Azienda 1

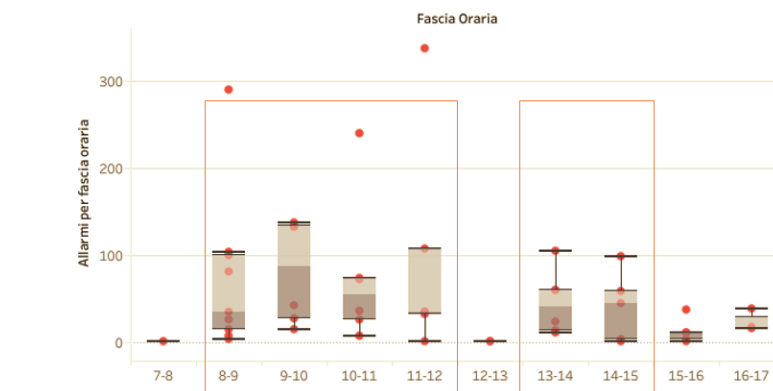


Figura 41 - Quali ore del giorno sono più critiche? Azienda 2

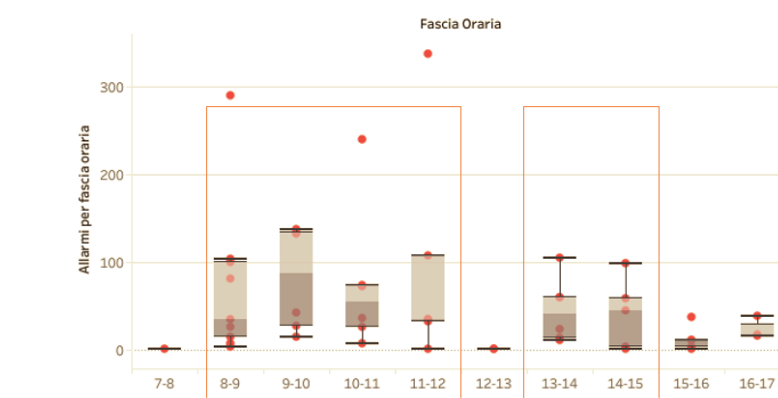


Figura 42 - Quali ore del giorno sono più critiche? Azienda 3

Le Figure 40, 41 e 42 mostrano la distribuzione degli allarmi di possibili collisioni nel tempo per le tre aziende. Come si evince dalle immagini, la prima e la seconda azienda presentano un comportamento opposto. Se nella prima, infatti, il maggior numero di possibili collisioni si registra a inizio e fine giornata e durante la pausa pranzo, nella seconda queste sono le ore in cui non sono stati registrati allarmi. Quest'ultimo andamento è simile a quello dell'azienda numero tre dove, inoltre, in diverse fasce orarie si registra un elevato numero di allarmi, spesso di gran lunga superiore alla mediana della fascia oraria stessa.

2. Quali **giorni della settimana** sono più critici?

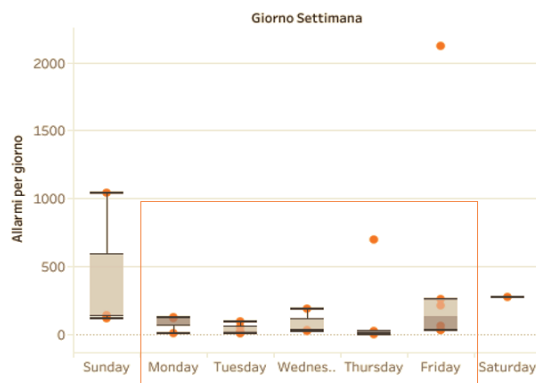


Figura 43 - Quali giorni della settimana sono più critici? Azienda 1

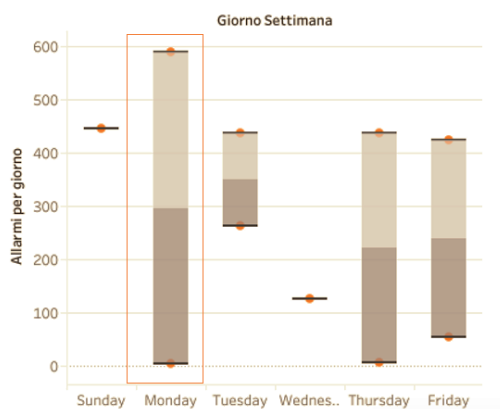


Figura 44 - Quali giorni della settimana sono più critici? Azienda 2

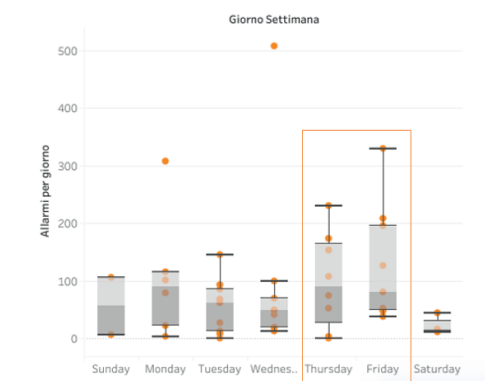


Figura 45 - Quali giorni della settimana sono più critici? Azienda 3

Le Figure 43, 44 e 45 mostrano la distribuzione degli allarmi di possibili collisioni nel tempo per le tre aziende e i risultati vengono mostrati raggruppati per giorni della settimana. Si nota che, nell'azienda uno la distribuzione è abbastanza omogenea, fatta eccezione per un picco del numero di allarmi la domenica e pochissimi outlier. Al contrario, nell'azienda due, la domenica non si registrano allarmi, mentre inizio e fine della settimana lavorativa sembrano essere i momenti più critici. Anche per l'azienda tre, il giovedì e il venerdì sono i giorni più critici.

3. Quali zone dell'azienda sono più critiche?



Figura 46 - Quali zone dell'azienda sono più critiche? Azienda 1

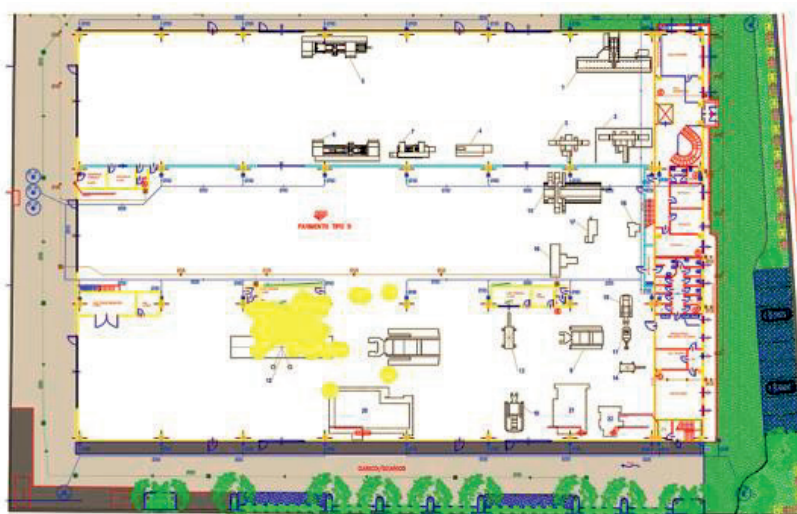


Figura 47 - Quali zone dell'azienda sono più critiche? Azienda 2

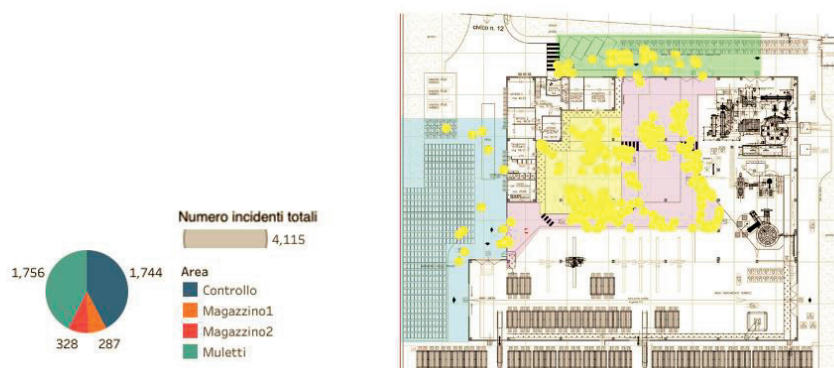


Figura 48 - Quali zone dell'azienda sono più critiche? Azienda 3

Le Figure 46, 47 e 48 rappresentano le planimetrie delle tre aziende con l'indicazione della posizione dei possibili incidenti e l'analisi della distribuzione di questi nello spazio. Ovviamente, i risultati dipendono dalla natura delle singole aziende e dalle principali attività svolte in ciascuna di esse. Dagli allarmi analizzati, comunque, si possono ricavare le seguenti informazioni. L'azienda uno vede una concentrazione di possibili collisioni nell'area magazzino, mentre nell'azienda tre le aree magazzino sembrano essere le meno problematiche e la concentrazione di possibili incidenti è nell'area controllo e muletti. Nell'azienda due l'unica zona interessata da possibili collisioni è invece l'area produzione. Le analisi sopra descritte forniscono un'idea delle potenzialità del sistema progettato e implementato. Le osservazioni fatte possono essere utilizzate dalle aziende per definire interventi per migliorare la sicurezza dei propri lavoratori, ad esempio ridistribuendo il personale durante le ore pomeridiane, o monitorando gli accessi/transiti nelle aree critiche. D'altra parte, la comunicazione con le aziende permetterà di migliorare il sistema stesso, ad esempio, riducendo il valore di distanza minima di sicurezza in alcune aziende dove il numero di possibili incidenti è molto alto, in modo da ridurre il numero di segnalazioni e tag coinvolti e delineare più chiaramente le effettive situazioni di rischio.

6 Conclusioni

L'elevato numero di incidenti sul lavoro registrati all'interno delle PMI operanti nel settore metalmeccanico rende fondamentale agire per migliorare le condizioni di salute e sicurezza dei lavoratori. L'avvento di Industria 4.0, grazie all'introduzione di soluzioni tecnologiche innovative, può rappresentare un'importante leva per la riduzione degli incidenti. Tuttavia, a differenza delle aziende di grandi dimensioni, per le PMI si registrano limitati livelli di adozione di tali soluzioni tecnologiche. Questo differente - e meno efficiente - approccio verso Industria 4.0, qui definito *digital divide*, potrebbe ulteriormente acuire le differenze in termini di salute e sicurezza già esistenti tra PMI e grandi aziende, creando un ulteriore incremento del *safety divide*.

In questo documento abbiamo presentato RECKON, progetto di ricerca sviluppato su più unità laboratoriali del Politecnico di Milano, vincitore del bando Inail BRiC 2018 ID09. I risultati principali del progetto possono riassumersi come segue:

- Sviluppo di un Sistema di sensorizzazione *low cost*, minimamente invasivo ed estremamente flessibile basato su telecamere e tag RFID per la localizzazione di operatori, macchinari e muletti nell'ambiente di lavoro;
- Sviluppo di algoritmi di controllo in tempo reale per la verifica dell'utilizzo consono di spazi e apparecchiature e per la prevenzione di comportamenti a rischio;
- Sviluppo di algoritmi di controllo in tempo reale per l'identificazione di situazioni anomale di macchinari o loro elementi;
- Sviluppo di algoritmi di intelligenza artificiale per l'analisi dei testi nelle descrizioni di Inail di infortuni sul lavoro per identificare pattern ed eventi ripetuti.

Inoltre, in questo documento abbiamo riassunto lo studio metodologico alla base delle scelte implementative, studio che rende il sistema RECKON altamente generalizzabile ed estendibile a diversi casi di studio. Per quanto riguarda i casi di studio portati avanti durante i due anni di progetto, i primi risultati ottenuti, e illustrati nella Sezione 5 di questo documento, evidenziano le potenzialità del sistema, sia per il monitoraggio degli incidenti sull'ambiente di lavoro, sia per il miglioramento del sistema stesso.

Ringraziamenti

Gli autori desiderano ringraziare Silvia Negri e Irene Tagliaro di API (Associazione Piccole e Medie Industrie della provincia di Lecco e Sondrio), l'azienda NewEn s.r.l., le Aziende che hanno ospitato i prototipi sviluppati rendendo possibile la sperimentazione e tutti gli altri ricercatori coinvolti nel progetto, in modo particolare: Fabio Alberto Schreiber, Alessandra Scutti e Mahsa Shekari.

Riferimenti bibliografici

[Ab-999] Abowd, Gregory D., et al. "Towards a better understanding of context and context-awareness." International symposium on handheld and ubiquitous computing. Springer, Berlin, Heidelberg, 1999.

[Ag-994] R. Agrawal, et al. "Fast algorithms for mining association rules," in: Proc. 20th int. conf. very large data bases, VLDB, volume 1215, Citeseer, 1994, pp. 487-499.

[AIS-993] R. Agrawal, T. Imieliński, A. Swami, "Mining association rules between sets of items in large databases," in: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 1993, pp. 207-216.

[BBS-018] Badri, A., Boudreau-Trudel, B., Souissi, A.S. Occupational health and safety in the industry 4.0 era: A cause for major concern? Safety Science, 2018; 109:403-11. doi:10.1016/j.ssci.2018.06.012.

[Ba-015] Barral V, Suárez-Casal P, Escudero CJ, García-Naya JA. Multi-sensor accurate forklift location and tracking simulation in industrial indoor environments. Electron 2019;8. doi:10.3390/electronics8101152.

[Be-003] Y. Bengio, et al. "A neural probabilistic language model," The journal of machine learning research 3 (2003) 1137-1155.

[Bi-006] C. M. Bishop, "Pattern recognition and machine learning," springer, 2006.

[Bi-016] Bissert, P.T., Carr, J.L., Du Carme, J.P.; Proximity detection zones: Designs to prevent fatalities around continuous mining machines, Professional Safety, 61 (6), pp. 72-77, 2016.

[Bo-004] Bodenreider, Olivier. "The unified medical language system (UMLS): integrating biomedical terminology." Nucleic acids research 32.suppl_1 (2004): D267-D270.

[Bo-009] Bolchini, Cristiana, et al. "And what can context do for data?." Communications of the ACM 52.11 (2009): 136-140.

[Bu-014] Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A.; A privacy-preserving solution for tracking people in critical environments, Proceedings - IEEE COMPSACW 2014, art. no. 6903120, pp. 146-151, 2014.

[Bu-021] Buffi, A., Bernardini, F., Nepa, P., Marracci, M., Tellini, B., Di Donato, L., Pirozzi, M., Tomassini, L., Ferraro, A.; RFID-Based Localization Enables a Smart System for Worker Safety, RTSI 2021, pp. 513-518, 2021.

[By-018] Byeon J, Jang M-S, Choi S-W, Yoo HD, Lee E-H. A Study on Smart Helmet to Efficiently Cope with the Operation and Safety of Workers in Industrial Settings. Int J Control Autom 2018;11:169-78. doi:10.14257/ijca.2018.11.3.15.

- [Ce-013] Cheng T, Teizer J. Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. *Autom Constr* 2013;34:3-15. doi:10.1016/j.autcon.2012.10.017.
- [Ch-009] Cheng, C.-S., Chang, H.H., Chen, Y.-T., Lin, T.H., Chen, P.C., Huang, C.M., Yuan, H.S., Chu, W.C.; Accurate location tracking based on active RFID for health and safety monitoring, *iCBBE* 2009, 5162266, 2009.
- [Ch-014] K. Cho, et al. "Learning phrase representations using rnn encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078 (2014).
- [CA-018] Confartigianato. STUDI 2018. [https://www.confartigianato.it/2019/05/studi-nel-2018-valore-aggiunto-della-manifattura-a-21-ritmo-doppio-del-totale-economia-italia-prima-in-ue-per-occupati-nelle-piccole-imprese-della-manifattura-2-milioni-di-addetti-405-mila-in/#:~:text=L'Italia conta](https://www.confartigianato.it/2019/05/studi-nel-2018-valore-aggiunto-della-manifattura-a-21-ritmo-doppio-del-totale-economia-italia-prima-in-ue-per-occupati-nelle-piccole-imprese-della-manifattura-2-milioni-di-addetti-405-mila-in/#:~:text=L'Italia%20conta) (accessed June 4, 2020).
- [Da-012] Davis, Jim, et al. "Smart manufacturing, manufacturing intelligence and demand-dynamic performance." *Computers & Chemical Engineering* 47 (2012): pp. 145-156.
- [DDF-003] Davies, John, Dieter Fensel, and Frank Van Harmelen. "Towards the semantic web." Chichester, John Wiley & Sons (2003).
- [De-018] Depari A, Flammini A, Fogli D, Magrino P. Indoor Localization for Evacuation Management in Emergency Scenarios. 2018 *Work Metrol Ind 40 IoT, MetroInd 40 IoT* 2018 - doi:10.1109/METROI4.2018.8428343, pp.146-50.
- [Dev-018] J. Devlin, et al. Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
- [Di-009] Dikaiakos, Marios D., et al. "Cloud computing: Distributed internet computing for IT and scientific research." *IEEE Internet computing* 13.5 (2009): 10-13.
- [EC-003] European Commission. Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. 2003.
- [ES-019] Eurostat. Accidents at work statistics. 2019. https://ec.europa.eu/eurostat/statistics-explained/index.php/Accidents_at_work_statistics (accessed June 4, 2020).
- [FCP-004] Fabiano, B, Currò, F, Pastorino, R. A study of the relationship between occupational injuries and firm size and type in the Italian industry. *Saf Sci* 2004; 42:587-600. doi:10.1016/j.ssci.2003.09.003
- [Fa-019] Faramondi L, Bragatto P, Fioravanti C, Gnoni MG, Guarino S, Setola R. A Wearable Platform to Identify Workers Unsafety Situations. 2019 *IEEE Int Work Metrol Ind 40 IoT, MetroInd 40 IoT* 2019 - Proc 2019:339-43. doi:10.1109/METROI4.2019.8792857.
- [FPT-020] Fantini, Paola, Pinzone Marta, and Taisch Marco. "Placing the operator at the centre of Industry 4.0 design: Modelling and assessing human activities within cyber-physical systems."

Computers & Industrial Engineering 139 (2020): 105058. [Ga-020] Gavuji, P., Kaza, N.; Industrial Safety System using Internet of Things, International Journal of Scientific Research and Engineering Trends, 6 (3), pp. 1084-1088, 2020.

[GJM-013] A. Graves, N. Jaitly, A.-r. Mohamed, Hybrid speech recognition with deep bidirectional lstm, in: 2013 IEEE workshop on automatic speech recognition and understanding, IEEE, 2013, pp. 273-278.

[HC-009] Hae-Sang Park, Chi-Hyuck Jun. A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications. Volume 36, Issue 2, Part 2, 2009.

[HS-997] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735-1780.

[Inail-021] Inail, Andamento degli infortuni sul lavoro e delle malattie professionali, <https://www.inail.it/cs/internet/docs/alg-dati-inail-2021-gennaio-pdf.pdf>, 2021.

[Istat-019] ISTAT - Istituto nazionale di statistica. Imprese - Annuario statistico I. 2019.

[Istat-020] ISTAT - Imprese e addetti: classi di addetti, settori economici, 2020. <http://dati.istat.it/Index.aspx?QueryId=20596> (accessed June 24, 2020).

[KB-014] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

[Kh-018] Khurana C, Ahluwalia P, Varshney M, Pandey S. Surveyance of Ambient Conditions in Mines Using Intelligent Sensor Nodes. Proc 2nd Int Conf Trends Electron Informatics, ICOEI 2018 2018:1126-31. doi:10.1109/ICOEI.2018.8553962.

[Ko-015] Kozlovsky M, Pavlinić DZ, Oder A, Fehér G, Bogdanov P. Situation and location awareness in harsh environment. 2015 38th Int Conv Inf Commun Technol Electron Microelectron MIPRO 2015 - Proc 2015:328-32. doi:10.1109/MIPRO.2015.7160288.

[Ko-019] Komane BL, Mathonsi TE. A Review of Wireless Sensor Networks: Early Accident Detection Models for South African Mine Industries. 2019 Open Innov Conf OI 2019 2019:36-41. doi:10.1109/OI.2019.8908192.

[La-021] Landi, L., Buffi, A., Stecconi, A., Marracci, M., Di Leone, P., Bernardini, F., Di Donato, L.; Localization systems for Safety Applications in Industrial Scenarios, ESLER, 2021.

[LBH-015] Lee, Jay, Behrad Bagheri, and Hung-An Kao. "A cyber-physical systems architecture for industry 4.0-based manufacturing systems." Manufacturing letters 3 (2015): 18-23.

[LRU-014] Leskovec, J., Rajaraman, A., Ullman, J.. Mining of Massive Datasets (2nd ed.). Cambridge: Cambridge University Press. 2014. doi:10.1017/CBO9781139924801.

[Lu-011] Lui, Gough; Gallagher, Thomas; Binghao, Li (2011). "Differences in RSSI readings made

by different Wi-Fi chipsets: A limitation of WLAN localization”. 2011 International Conference on Localization and GNSS (ICL-GNSS). Localization and GNSS (ICL-GNSS), 2011 International Conference on. pp. 53-57.

[Lu-017] Lu, Yang. “Industry 4.0: A survey on technologies, applications and open research issues.” *Journal of industrial information integration* 6 (2017): 1-10.

[Me-000] Mehata K, Shankar S, Karthikeyan N, Nandhinee K, Hedwig R. IoT Based Safety and Health Monitoring for Construction Workers. 2019 1st Int. Conf. Innov. Inf. Commun., IEEE; 2000, p. 890-2. doi:10.1007/978-3-662-07677-4_120.

[MSG-014] S. Mahmood, M. Shahbaz, A. Guergachi, “Negative and positive association rules mining from text using frequent and infrequent itemsets,” *The Scientific World Journal* 2014 (2014).

[Mu-014] Musu, C., Popescu, V., Giusto, D.; Workplace safety monitoring using RFID sensors, TELFOR 2014, art. no. 7034494, pp. 656-659, 2014.

[Pa-009] Parameswaran, Ambili Thottam; Husain, M, I.; Upadhyaya, S. “Is RSSI a Reliable Parameter in Sensor Localization Algorithms - An Experimental Study” (PDF). September 2009. 28th International Symposium On Reliable Distributed Systems, New York.

[Pa-014] Paelke, Volker. “Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment.” *Proceedings of the 2014 IEEE emerging technology and factory automation (ETFa)*. IEEE, 2014.

[Pa-018] Park J, Cho YK, Khodabandelu A. Sensor-based safety performance assessment of individual construction workers. *Sensors (Switzerland)* 2018;18. doi:10.3390/s18113897.

[PJ-009] H.-S. Park, C.-H. Jun, “A simple and fast algorithm for k-medoids clustering,” *Expert systems with applications* 36 (2009) 3336-3341.

[Po-001] Porter, M. F. (2001).” *Snowball: A language for stemming algorithms*”.

[Sc-011a] Schreiber, Fabio A., et al. “Towards autonomic pervasive systems: the PerLa context language.” *Electronic Proceedings of the 6th International Workshop on Networking Meets Databases (Co-located with SIGMOD 2011)*, Athens. 2011.

[Sc-011b] Schreiber, Fabio A., et al. “Perla: A language and middleware architecture for data management and integration in pervasive information systems.” *IEEE Transactions on Software Engineering* 38.2 (2011): 478-496.

[Se-013] Sen, Souvik; Lee, Jeongkeun; Kim, Kyu-Han; Congdon, Paul. “Avoiding Multipath to Revive Inbuilding WiFi Localization”. 2013.

[Si-022] Siddharth, S.G., Tamilselvan, G.M., Venkatesh, C.; Location Prediction for Improved Human Safety at Complex Environments, *Computers, Materials and Continua*, 71(2), pp. 5219-5234, 2022.

- [SP-017] Schreiber, Fabio Alberto, and Emanuele Panigati. "Context-aware self adapting systems: a ground for the cooperation of data, software, and services." *International Journal of Next-Generation Computing* (2017): 32-61.
- [So-015] Sommer, L. Industrial revolution - Industry 4.0: Are German manufacturing SMEs the first victims of this revolution?. *J Ind Eng Manag* 2015; 8:1512-32. doi:10.3926/jiem.1470.
- [Sr-014] N. Srivastava, et al. "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research* 15 (2014) 1929-1958.
- [SVL-014] I. Sutskever, O. Vinyals, Q. V. Le, "Sequence to sequence learning with neural networks," *arXiv preprint arXiv:1409.3215* (2014).
- [TM-007] Tanenbaum, Andrew S., and Maarten Van Steen. *Distributed systems: principles and paradigms*. Prentice-Hall, 2007.
- [WK-992] J. J. Webster, C. Kit, "Tokenization as the initial phase in nlp," in: *COLING 1992 Volume 4: The 15th International Conference on Computational Linguistics*, 1992.
- [WW-016] Walters, D., Wadsworth, E. Contexts and arrangements for occupational safety and health in micro and small enterprises in the EU - SESAME project European Risk Observatory. 2016. doi:10.2802/665614.
- [Xu-014] Xu L Da, He W, Li S. Internet of things in industries: A survey. *IEEE Trans Ind Informatics* 2014;10:2233-43. doi:10.1109/TII.2014.2300753.
- [Yi-016a] Yi W, Chan APC, Wang X, Wang J. Development of an early-warning system for site work in hot and humid environments: A case study. *Autom Constr* 2016;62:101-13. doi:10.1016/j.autcon.2015.11.003.
- [Yi-016b] Yi W, Zhu J, Liu X, Wang X, Chan APC. A framework for establishing early warning system for working in hot environments. *ISARC 2016 - 33rd Int Symp Autom Robot Constr* 2016:1025-31. doi:10.22260/isarc2016/0123.

I Quaderni di ricerca

- 1 - G. Ciriello, M. De Felice, R. Mosca, M. Veltroni, *Infortuni sul lavoro. Un modello di lettura (della numerosità) su "open data" dell'Inail*
- 2 - L. Cortis, F. M. Fabiani, C. Ratti, L. Rossi, D. G. Svampa, C. Vitale, *Analisi e valutazione del comportamento dell'assorbitore di energia nei sistemi di arresto caduta con differenti masse di prova*
- 3 - P. Rossi, *Danno biologico: monitoraggio, criticità e prospettive d'interesse medico-legale a dieci anni dall'introduzione della nuova disciplina*
- 4 - M. De Felice, A. Goggiamani, R. Mosca, M. Veltroni, *Malattie professionali. Un modello di lettura (della numerosità) su "open data" dell'Inail*
- 5 - L. Cortis, F. M. Fabiani, L. Rossi, D. G. Svampa, *Nuovi strumenti per la valutazione dello scivolamento e ribaltamento delle scale portatili*
- 6 - L. Cortis, F. M. Fabiani, L. Rossi, D. G. Svampa, *Determinazione della resistenza caratteristica dei dispositivi di collegamento montante-traverso nei ponteggi metallici fissi di facciata*
- 7 - S. Carruba, R. Grimaldi, M. Sparro, S. Tomasini, *Vendor Rating: il modello di valutazione delle forniture IT dell'Inail*
- 8 - P. Agnello, S. M. Ansaldi, *Una proposta di ontologia per gli Open Data Inail sugli infortuni: dalla numerosità dei dati alla gestione della conoscenza*
- 9 - A. Alvino, A. Antonini, C. Delle Site, A. Di Loreto, F. Giacobbe, D. Lega, E. Pichini, M.A. Pierdominici, L. Ricciardi, A. Tonti, *Analisi delle cause di rottura di materiali metallici: esempi di Applicazione*
- 10 - L. Cortis, F. M. Fabiani, L. Rossi, D. G. Svampa, *Light workers e criticità dell'assorbitore di energia nei sistemi di arresto caduta*
- 11 - S. Carruba, R. Grimaldi, M. Sparro, S. Tomasini, *Vendor Rating: la valutazione delle forniture IT dell'Inail per il 2015*
- 12 - L. Cortis, F. M. Fabiani, L. Rossi, D. G. Svampa, *Metodologia per prove di rigidità e resistenza per i dispositivi di collegamento montante-traverso nei ponteggi metallici fissi di facciata*
- 13 - R. Giovinazzo, E. Guerrera, M. Mameli, D. Sarto, M. Veltroni, *Analisi della contaminazione microbiologica degli ambienti di lavoro. Valutazione della qualità del dato analitico nel conteggio microbico su piastra e nelle prestazioni dell'operatore*
- 14 - P. Agnello, F. Bracco, C. Brunel, M. Masini, T. F. Piccinno, A. Sedaoui, D. Tazi, *Promuovere la sicurezza nelle organizzazioni attraverso manager resilienti*

- 15 - L. Rossi, *Esecuzione in sicurezza dei lavori in copertura. Misure di prevenzione e protezione*
- 16 - F. Antonello, P. Bragatto, F. Di Maio, S. M. Hoseyni, E. Zio, *Stima adattiva del rischio di rottura di attrezzature in pressione, sulla base dei dati di monitoraggio*
- 17 - J. Cecca, N. L'Episcopo, L. Mainero Rocca, *Nuove tecniche analitiche al servizio della sicurezza dei lavoratori: sviluppo di un metodo di analisi diretta di fungicidi e insetticidi sulla superficie delle foglie*
- 18 - A. Di Francesco, P. Iacono, A. Nebbioso, A. Pirone, G. Saputi, C. Simeoni, M.R. Vallerotonda, *Settore Seveso: contesto normativo e caratterizzazione statistica degli infortuni*
- 19 - F. M. Fabiani, L. Rossi, *Criticità delle reti di sicurezza di piccole dimensioni*
- 20 - Bragatto, C. Mennuti, M. F. Milazzo, E. Proverbio, *Tecniche avanzate per la gestione dei serbatoi di stoccaggio atmosferico*
- 21 - E. Badellino, T. P. Baccolo, M. R. Marchetti, A. Papale, I. Rossi, E. Rotoli, S. Signorini
Il progetto Inail: La prevenzione alimentare al lavoro in ottica bio-psico-sociale - I risultati della ricerca-intervento
- 22 - A. Abrardo, P. Agnello, S. M. Ansaldi, L. Belli, P. Bragatto, L. Davoli, F. M. Fabiani, G. Ferrari, L. Parri, *CP-SEC: Sistema Cyber-Fisico per la sicurezza dei lavoratori in presenza di sostanze pericolose*

