Università degli Studi di Padova

DIPARTIMENTO DI TECNICA E GESTIONE DEI SISTEMI INDUSTRIALE
Corso di Laurea Triennale in Ingegneria Meccatronica

Tesi di Laurea Triennale

L'era 4.0 della chirurgia robotica

Laureando:
Riccardo Saltarel

Relatore:
Prof. Giovanni Boschetti

Anno Accademico 2020/2021

Sommario

1.Introduzione, Robotica e Chirurgia 3
1.1.Perché i robot? 3
2.**Il robot stereotassico** 4
2.1.Robot endoscopici 5
2.2.L'operazione Lindenberg 6
2.3.Prospettive e attuali tecnologie 7
3.Il Da Vinci 8
3.1.Sistema di visione 9
3.2.Sistema di posizionamento e controllo 10
3.3.Strumentazione 10
3.4.Training 12
4.Image-guided robotic surgery 13
4.1.La navigazione intraoperativa 14
4.2.Computer vision surgery 15
4.3.L'apprendimento automatico 16
4.4.Identificazione automatica 16
5.Video processing 17
5.1.LSTM e selezione informazioni 18
5.2.Recurrent Neural Network (RNN) vs LSTM 19
5.3.Potenzialità e future direzioni 20
5.4.Machine Learning 21
5.5.Reti Neurali e deep learning 22
6.Computer vision 23
6.1.AlexNet 23
6.2.Convnet (CNN) 24
7.Conclusione 26
8.Bibliografia 27

1. Introduzione, Robotica e Chirurgia

Fin dall'antichità si è sempre cercato di ricorrere alla chirurgia lì dove il nostro corpo non era più in grado di ripararsi e ritornare alle condizioni ottimali. Le mani dell'uomo per definizione sono sempre state un valido ma non impeccabile strumento per finalizzare gli interventi più complicati di chirurgia. Con l'industria 4.0 nasce l'idea di rivolgersi all'affidabilità delle macchine software, all'accessibilità dei *big data* e alle innovazioni hardware per portare a termine i compiti sempre più delicati, lunghi e precisi. Ormai i robot non sono più qualcosa di lontano e con l'avanzamento delle nuove tecnologie come il *computer vision*, il *pattern recognition*, il *machine learning*, le macchine stanno acquisendo la facoltà di riconoscere oggetti, tracce, piani e di commettere specifiche decisioni. Il robot è un dispositivo meccatronico realizzato in uno o più bracci meccanici controllati attraverso circuiti integrati, sensori, trasduttori, attuatori, azionamenti e il proprio software. Può essere programmato per compiere istruzioni o tasks sequenziali automaticamente e può essere gestito attraverso interfacce meccaniche e computerizzate. Le nuove frontiere puntano all'intelligenza artificiale e all'automatizzazione delle facoltà decisionali di un automa. Utilizzarli in ambito sanitario può rivelarsi un grande vantaggio in termini di sicurezza, prestazioni, efficienza e ripetibilità.

1.1. Perché i robot?

Ai giorni d'oggi in chirurgia si evita il più delle volte che il corpo del paziente venga a contatto con l'ambiente esterno perché così lo si esporrebbe a notevoli rischi di tipo infettivo, scavando e spostando tessuti vi sarebbero grandi rischi di sanguinamento. Invece di operazioni dove il paziente viene operato venendo letteralmente dissezionato, si preferiscono altre alternative più recenti: le tecnologie di tipo MIS (*Minimal Invasive Surgical*). Camere ad alta definizione e microstrumenti ci permettono di agire in modo sempre meno invasivo, il paziente così oltre ad essere non più esposto a numerosi rischi può affidarsi a un macchinario che rende l'operazione più minuziosa e prevedibile: "*The benefits of these minimally invasive approaches included:* (i) *reduced wound access trauma,* (ii) *shorter hospital stay,* (iii) *improved visualisation,* (iv) *less postoperative wound complications (ranging from wound infections to incisional hernias), and* (v) *less disfigurement. As such they were designed to offer an equivalence to open surgery with less tissue trauma and speedier discharge that in turn was anticipated to offer* (vi*) increased cost-efficacy*" [i]

2. Il robot stereotassico

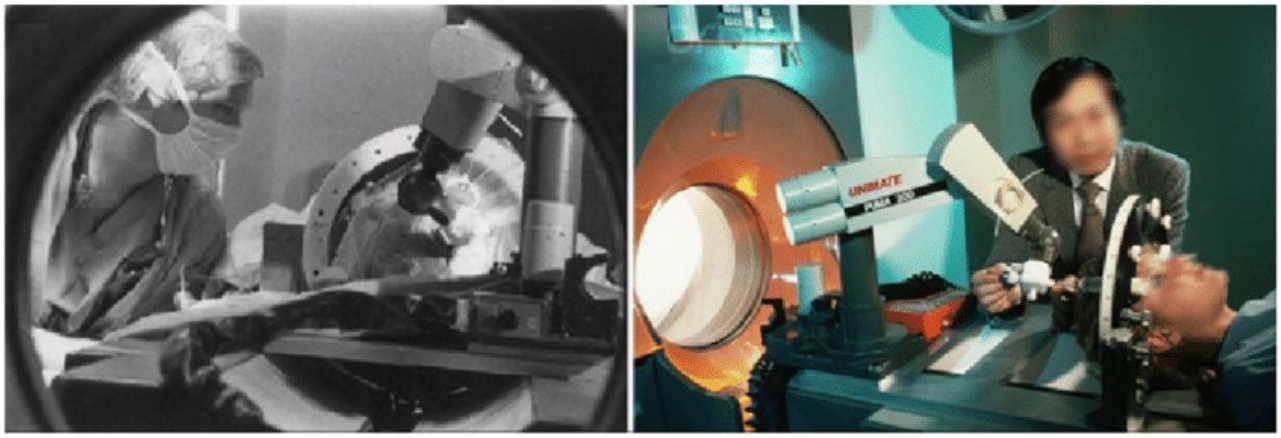

Figura 1

Il primo utilizzo di un robot a scopo clinico risale al 1985, quando per la prima volta fu utilizzato un Puma 200 (figura 1)[ii] per la biopsia di un cervello. L'utilizzo del robot permise di danneggiare meno tessuti sani possibili. La geometria del movimento e la precisione del taglio netto fu di importanza fondamentale per la riuscita dell'operazione. Questo robot però necessitava dell'ausilio dell'uomo per completare più compiti sequenziali differenti, non poteva procedere senza la supervisione di un operatore né poteva prendere in consegna tasks sequenziali senza essere controllato: poteva essere definito come una vera e propria estensione del suo supervisore. Un grande limite di questa macchina consisteva nel non poter operare in tessuti che fossero in movimento ma altresì statici senza alcuna pulsazione cioè come ossa e cervello[iii]. Il robot era in grado di fare semplici movimenti grazie a sistemi di guida CT oppure ad ultrasuoni. [1]Dopo diversi anni di evoluzione sono stati introdotti molti più vantaggi che superano il mero aspetto meccanico, vi era il bisogno di migliorare flessibilità e destrezza di operazione nei tessuti elastici e pulsanti.

[1] La tomografica computerizzata è una tecnica di diagnostica per immagini che consente di ricostruire un'immagine tridimensionale dei tessuti facendo passare dei raggi X attraverso la zona del corpo interessata, dopo aver rielaborato i dati con un computer.

2.1. Robot endoscopici

Solo alla luce del nuovo millennio vedremo i primi sistemi robotici endoscopici venire commercializzati[iv]. Inizialmente sponsorizzati dal fondo ricerca DARPA (*Defense Advanced Research Projects Agency)*, due principali robot furono sviluppati e approvati dalla *US Food And Drug Administration*. Questi ampliarono l'impiego della robotica a tutte quelle situazioni in cui si viene in contatto con tessuti che non sono più immobili ma che si spostano in base alla respirazione o il timbro cardiaco. Il primo fu Lo Zeus (figura 2), composto da 3 bracci robotici indipendenti, controllati dal chirurgo che osservava il campo chirurgico con un display in 2 dimensioni dalla console. I telemanipolatori master-slave legavano chirurgo e robot attraverso dei joystick in modo che i bracci fossero controllati attraverso un'interfaccia aptica che replica e filtra i movimenti delle mani dell'uomo eliminando tremolii e aumentando l'accuratezza di diversi fattori. Una delle braccia era controllata da un sistema di navigazione comandato a voce chiamato AESOP (*Automated Endoscopic System for Optimal Positioning*). Nonostante non risolse il problema della percezione di profondità che esisteva anche con la standard MIS, si rilevò un successo per la telechirurgia.

Figura 2

2.2. L'operazione Lindenberg

Questi concluse con successo il primo intervento alle coronarie a cuore pulsante durante l'operazione Lindenberg (figura 3)[v]. Lo scopo era quello di brevettare una teleoperazione a centinaia di chilometri di distanza. L'impresa consisteva in un intervento telerobotico tra un paziente in Francia a Strasburgo e il chirurgo a New York. La velocità della trasmissione dati e latenza di connessione poteva influenzare pericolosamente la riuscita dell'intervento. Nel caso in cui ci fosse stato un ritardo maggiore di 250ms sarebbe stato notevolmente incrementato il tempo per il completamento dei tasks, comparati al tempo reale[vi]. In collaborazione con la France telecom, si è utilizzato una fibra ottica ad alta velocità che ha permesso la fattibilità dell'impresa.

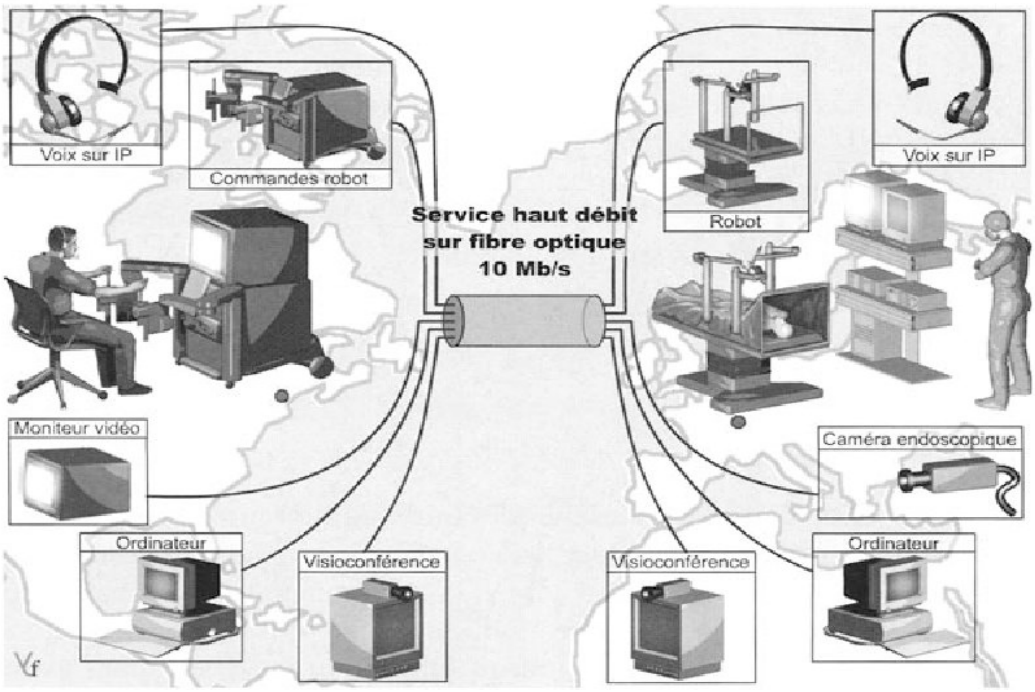

Figura 3

2.3. Prospettive e attuali tecnologie

La miglior soluzione per sorpassare il limite della telechirurgia e migliorare ulteriormente le procedure è la vera e propria automazione. Questa sarebbe possibile attraverso l'integrazione di sistemi di registrazione automatici e real-time tra *imaging* e sistemi robotici con una piattaforma informatica che autonomamente riconosce i tipi di procedura e le specifiche anatomiche del paziente e predice spostamenti e alterazioni fisiche degli organi [vii] .Le istruzioni più semplici possono essere eseguite semi-automaticamente attraverso un interfaccia computer-uomo assistita che utilizza modelli stocastici di analisi delle traiettorie: "*Semiautonomous telerobotic atrial fibrillation ablation was made possible over a distance between Boston and Milan using real-time electromagnetic navigation and preoperative CT images of the patient. The system created the surgical plane autonomously, using a stochastic model based on an anatomical atlas of 10 000 patients. Surgical steps can be recognized automatically and, to some extent, predicted. Organ displacements owing to breathing motion and heartbeat can also be predicted effectively, with a relatively small error margin. Additionally, ongoing research is bringing exciting results on organ deformation based on biomechanical properties. The process has now begun; however, many improvements are still required to conceive a full multistep complex procedure automatically performed by machine*" [viii]

Figura 4

3. Il Da Vinci

Nel 2003 la *Surgical Intuitive* acquistò *Computer Motion* e si guadagnò il monopolio del mercato e lo sviluppo dello Zeus venne interrotto per lasciare spazio al Da Vinci (figura 4). A differenza del precedente consisteva in una piattaforma molto più compatta e leggera con un braccio in più e dotato di ruote (figura 5). Vantava di offrire una visione 3D stereoscopica grazie un nuovo modello di telecamera (figura 6).
Ulteriori comandi posizionati sia agli estremi del supporto per gli avambracci, sia sulla pedaliera completano la funzionalità della console (5 pedali che offrono la possibilità di manovrare la telecamera, focalizzare l'immagine, disattivare gli strumenti, azionare la coagulazione dell'elettro-bisturi o muovere l'eventuale braccio ausiliario).

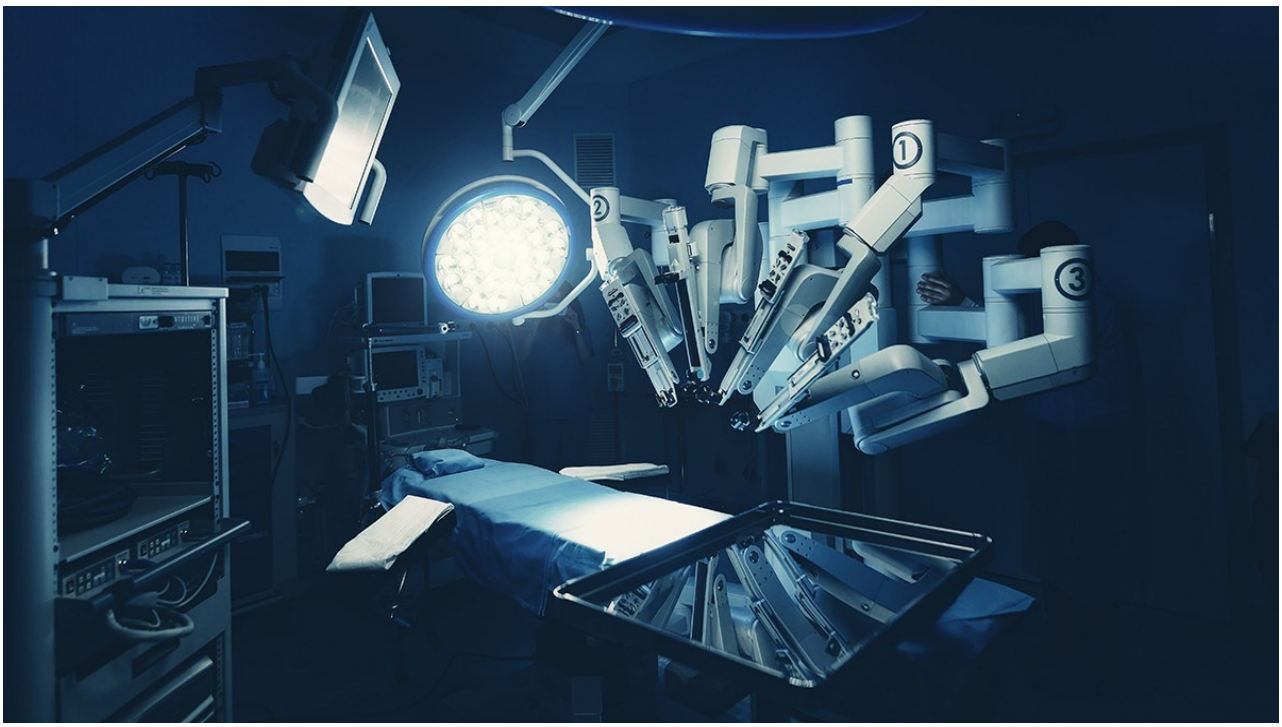

Figura 5

3.1. Sistema di visione

Il sistema di visione InSite costituisce uno dei punti caratteristici di questo sistema. Si tratta di un endoscopio ad alta risoluzione con visione tridimensionale. All'interno della cannula dell'endoscopio sono alloggiati due canali ottici, che trasmettono, attraverso due videocamere separate e indipendenti, l'immagine del campo operatorio a due monitor ad alta risoluzione. L'immagine è inoltre elaborata da un sistema di accentuazione del contrasto marginale e di riduzione del rumore elettronico di fondo. Il chirurgo ha la sensazione di esser quasi immerso nel campo operatorio, con una forte percezione della sua tridimensionalità. Tale effetto si ottiene grazie ai due monitor montati nella console, ognuno dei quali trasmette l'immagine proveniente da ogni singolo canale a ciascun occhio separatamente. La naturale profondità di campo, ottenuta grazie a questo artificio, il contrasto particolarmente elevato e la magnificazione, consentono una discriminazione spaziale assolutamente superiore a quella ottenibile con un sistema di visione tradizionale bidimensionale. L'inquadratura del campo operatorio è controllata direttamente dall'operatore che può modificarla attivando il braccio robotico endoscopico tramite il pedale di comando.

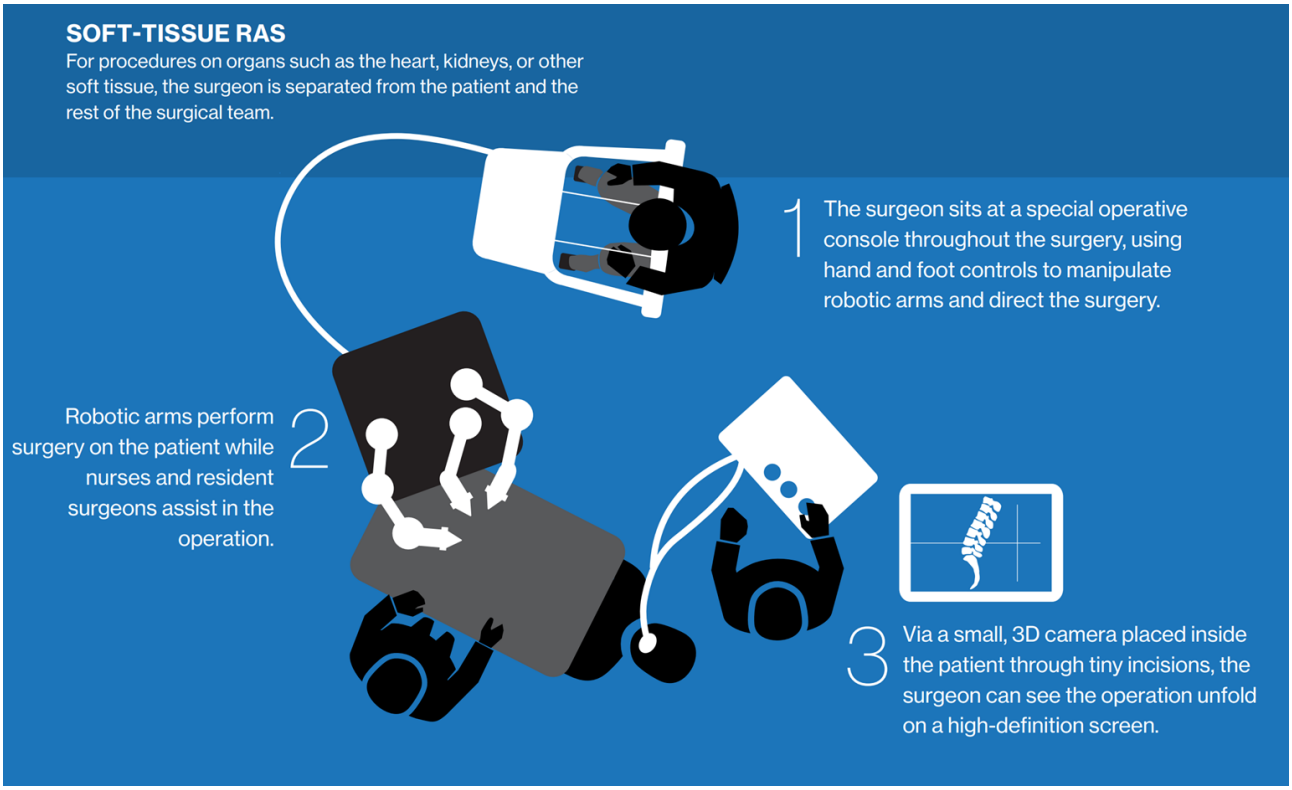

Figura 6

3.2. Sistema di posizionamento e controllo

Azionando i due master, è possibile spostare la posizione dell'ottica indietro e in avanti, a destra e a sinistra, in alto e in basso. È da rilevare che la qualità di visione e la tridimensionalità della stessa, oltre a consentire l'esecuzione di manovre delicate anche in spazi ristretti, può, in molte circostanze, compensare la mancanza di feedback tattile del sistema, come nella valutazione della tensione di rottura dei fili di sutura. Uno schermo touchscreen è montato sulla barra soffice su cui il chirurgo poggia gli avambracci durante la procedura chirurgica, serve per controllare ed impostare il sistema (calibrare il sistema di visione, impostare il rapporto di movimento, ecc.) e per determinarne l'accensione, l'avvio, l'arresto di emergenza e lo spegnimento, oltre che per gestire eventuali console accessorie o sistemi di proiezione di imaging. L'operatore della console chirurgica ha inoltre la possibilità di passare dalla vista a schermo intero ad una modalità a più immagini (visualizzazione TilePro™), che mostra l'immagine 3DHD del campo operatorio insieme ad altre due immagini (ecografo, ECG...) fornite da ingressi ausiliari.

3.3. Strumentazione

Oltre ad offrire una vasta gamma di strumenti intercambiabili a seconda della destinazione d'uso definita (clampaggi, suture, manipolazioni di tessuti), la Intuitive Surgical ha progettato questi stessi strumenti in modo da garantire la massima manovrabilità possibile, attraverso una modellazione basata sul polso umano.
Secondo l'azienda produttrice il sistema EndoWrist® consente al robot di raggiungere un grado di precisione chirurgica anche superiore alle capacità umane. Di fatto, grazie alla struttura realizzata con micro-tiranti in acciaio che riproducono la funzione e la direzione delle articolazioni del polso, è possibile riprodurre, all'interno del campo operatorio, sulla punta di strumenti del calibro di 5-8 mm, tutti i movimenti realizzabili con una mano.

Figura 7

Gli strumenti (figura 7) consentono rotazioni da 90° a 360° su 7 gradi di libertà (DOF, *Degrees Of Free*), permettendo traslazione, rotazione, primo snodo della testa dello strumento, secondo snodo della sola parte applicata della testa dello strumento e capacità di presa della parte applicata della testa dello strumento. Grazie al meccanismo di soppressione del tremore e al rapporto scalare dell'ampiezza dei movimenti inoltre è stato possibile limitare sempre più gli errori di movimento che affliggevano la più tradizionale tecnica laparoscopica.

Figura 8

3.4. Training

La possibilità di operare in sinergia con un altro chirurgo garantisce anche il corretto apprendimento delle procedure cliniche, riducendo la *learning curve* (curva di apprendimento) da parte degli specialisti che desiderano approcciarsi alla tecnica chirurgica robotica. L'enorme rivoluzione apportata dal sistema "Da Vinci" nel mondo delle scienze mediche suscitò moltissima diffidenza nella comunità dei chirurghi, da sempre abituati a 'lavorare' in un certo modo e soprattutto con certi strumenti. Per questo motivo la *Intuitive Surgical* ha posto grande attenzione alla problematica dell'apprendimento e dell'addestramento. Ad oggi è possibile predisporre di simulatori virtuali (pacchetti hardware e software) che consentono ai chirurghi giovani e meno giovani di migliorare la propria abilità con il robot (figura 9). Il simulatore prevede una varietà di esercizi incentrati allo sviluppo di abilità specifiche (gestione dei comandi della console, corretto sfruttamento di tutte le potenzialità degli strumenti robotici, etc.…). Alla fine di ogni esercizio il sistema assegna un punteggio all'operatore, consentendogli di migliorare i risultati ottenuti monitorando di volta in volta i propri progressi.[ix]

Figura 9

4. Image-guided robotic surgery

La chirurgia cibernetica si affida all'utilizzo della robotica e della guida fornita da immagini in tempo reale. La combinazione di queste tecniche permette di ottimizzare l'accuratezza e la sicurezza delle operazioni chirurgiche. Il VR è un simulatore reale informatico in sviluppo. VR è un software medico che può elaborare un modello virtuale del paziente da un'immagine CT o MRI (Risonanza magnetica tomografica).[x] Gli operatori hanno la possibilità di allenarsi attraverso simulatori tridimensionali immersivi. In questo modo si pianifica un modello per ricostruire virtualmente la seduta prima dell'operazione vera e propria. I modelli 3D virtuali possono essere ottenuti su misura per ogni paziente, migliorando la visualizzazione dei dettagli anatomici che in altre circostanze verrebbero sottostimati.

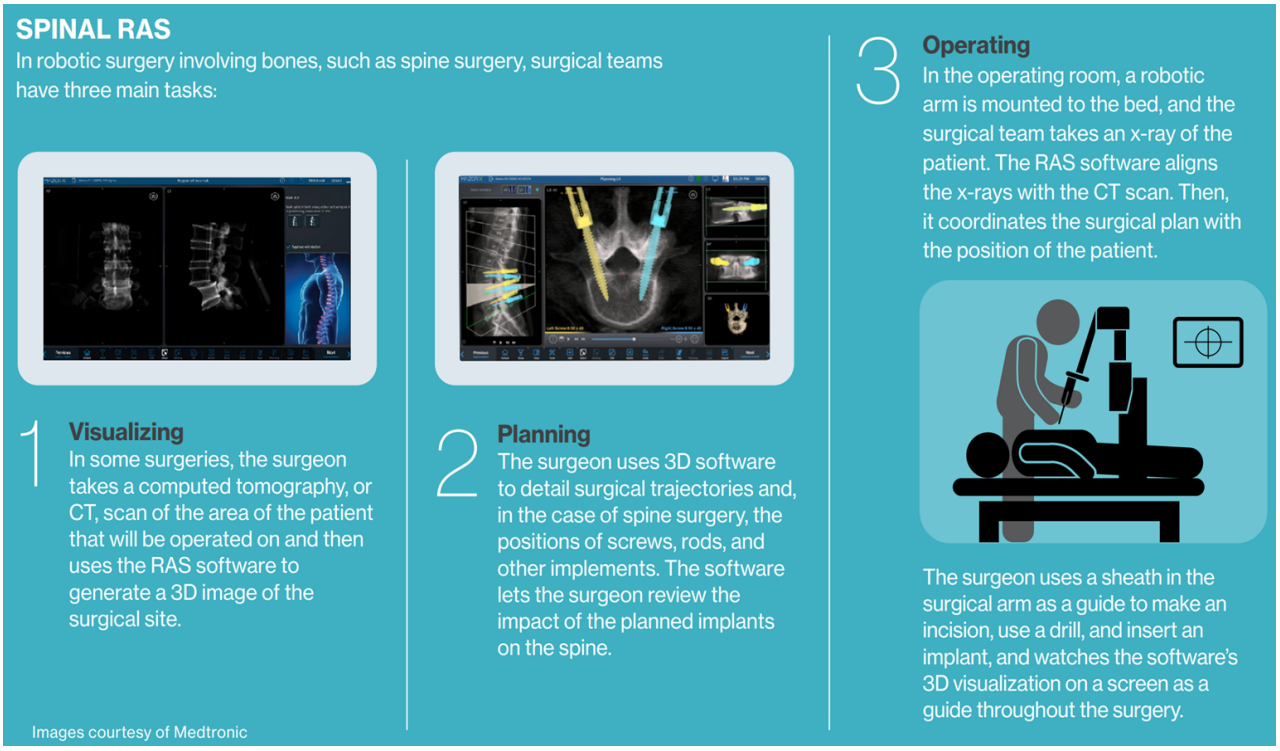

Figura 10

4.1. La navigazione intraoperativa

In alcuni casi è necessario unire il modello virtuale con delle immagini in tempo reale, definendo una vera e propria navigazione intraopertiva che evidenzia i pattern, le strutture e le variazioni anatomiche. L'esplorazione virtuale può rivelarsi un potente alleato per pianificare e simulare le procedure. Durante l'intervento il modello può essere sovrapposto con le immagini in tempo reale del paziente, ottenendo così una realtà virtuale in grado di permettere una visualizzazione trasparente interna completa anche in presenza degli organi o altri ostacoli (figura 11). Se inizialmente ci si limitava a strutture fisse ben in contrasto tra loro, come nella chirurgia maxillofacciale, nei tessuti flessibili le cose diventano molto più complesse a causa dei movimenti respiratori e delle deformazioni elastiche, in questo modo non possiamo fare riferimento ad immagini istantanee.

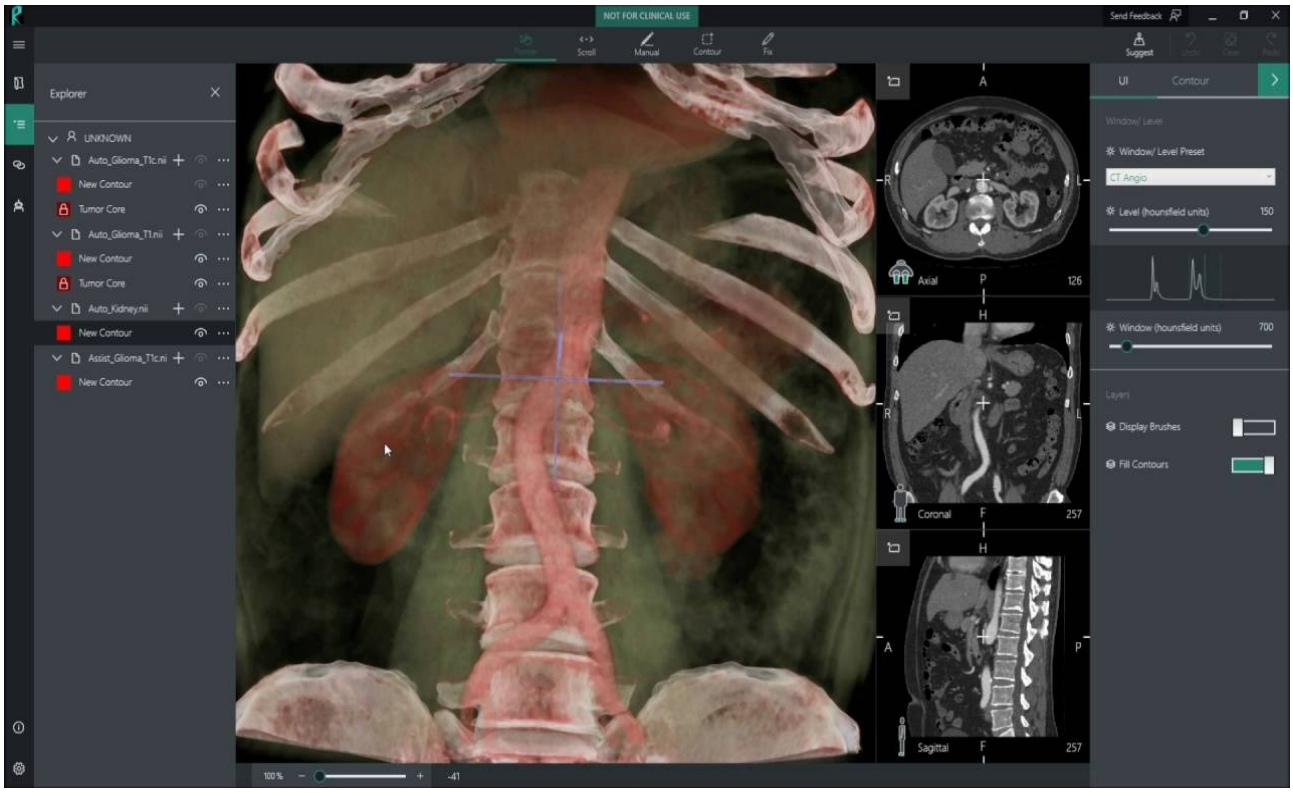

Figura 11

La miglior opzione consiste nel basarsi su un sistema rigido costruito tramite immagini 3D aggiornate costantemente in ogni momento dell'operazione, utilizzando *four-dimensional ultrasonography with ultrasound–CT image fusion*.[xi] Un'altra soluzione consiste nella predizione del movimento e delle deformazioni ottenute da modelli computazionali delle proprietà biomeccaniche degli organi o sistemi di guida elettromagnetici. "*Image guidance to complement robotically assisted procedures, through the concepts of augmented reality, represents the major revolution to increase safety and deal with difficulties associated with minimally invasive approaches*"[xii]

4.2. Computer vision surgery

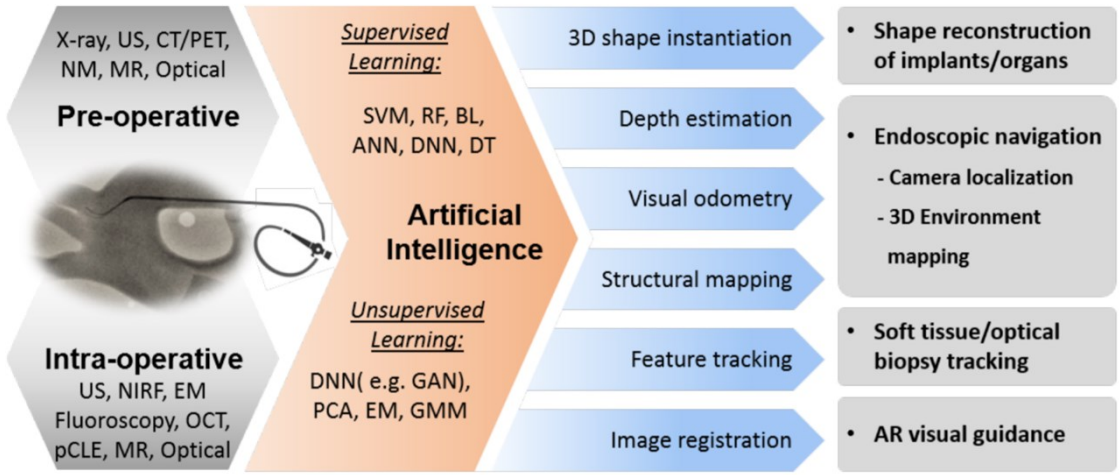

Figura 12

4.3. L'apprendimento automatico

Immagina di operare sotto lo sguardo di un chirurgo che ha assistito a migliaia di operazioni e ogni tipo di complicazione ed è pronto per fornire aiuto ad ogni momento. L'intelligenza artificiale punta a rendere questo realtà (figura 12). Anche se non possiamo fornire la vera vista ad un automa, possiamo, con il *machine learning*, far apprendere alle macchine tutti i *data streams* provenienti dalla sala operativa, come le posizioni degli strumenti, le informazioni dei macchinari, i tempi e le tempistiche di ogni fase, intere procedure, casistiche e molto altro. Il modo di comprendere la realtà delle macchine è sempre stato limitato alla comprensione procedurale, ma ora con il processamento di big data, enormi dataset e video le cose stanno cambiando.
Recenti frontiere del computer vision hanno utilizzato il *deep learning* per dare ai robot l'abilità di riconoscere immagini e oggetti. Questi processi permettono di comprendere procedure complesse tramite reti di connessioni multiple. La *Convolution Neural Network* (CNN) ha una struttura simile alla corteccia visuale umana e permette il riconoscimento degli oggetti come strumenti. Insieme ad una *long short term memory*, i computer riescono a contestualizzare ogni istante dell'operazione caricata a video. Con questi due strumenti è possibile far comprendere la chirurgia, ai robot, nel modo migliore. I ricercatori stanno tentando di far comprendere alle macchine come identificare le fasi dell'operazione e come riconoscere la gamma di strumenti da utilizzare.[xiii]

4.4. Identificazione automatica

Un punto focale di estrema importanza per i ricercatori IA consiste nel fornire algoritmi alle macchine per comprendere e saper differenziare le numerose fasi operative. Prima di tutto si è iniziato generando algoritmi per differenziare le fasi temporali in base determinati e precisi eventi o istanti temporali. Prima del *Computer Vision* era troppo laborioso utilizzare video dati per costruire modelli, oggi si usano data streams molto più adatti allo scopo come la presenza o l'accensione di determinati strumenti elettronici nel campo visivo. Alcune tecniche *Machine Learning* sono già state utilizzate in numerosi interventi dove le fasi sono state differenziate con successo grazie il numero limitato di strumenti utilizzati o con determinate situazioni che hanno reso ancora più semplice la differenziazione temporale.

5. Video processing

Processare dati video è un'impresa per i tradizionali sistemi di *machine learning*, 1 minuto di video contiene più di 25 volte dati di una tomografia computerizzata ad alta risoluzione[xiv]. Inizialmente gli sforzi dei ricercatori si sono rivolti verso tecniche per diminuire l'ammonto di informazioni delle immagini per imparare dati ridotti. Sono state le CNNs (figura 13) che hanno permesso di creare accurati e generalizzati algoritmi di intelligenza artificiale per riuscire a dividere i passi procedurali, come per esempio Endonet. Questa è stata migliorata a sua volta aggiungendo un semplice sistema di *machine learning* temporale per rendere consistente l'identificazione.[xv]

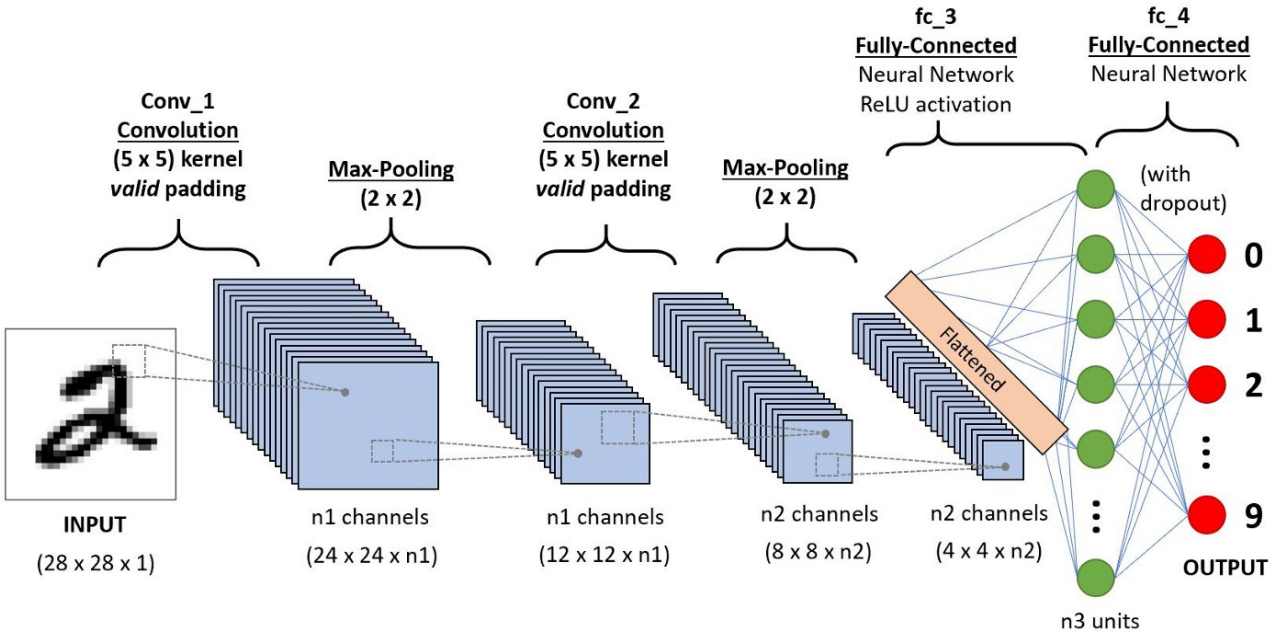

Figura 13

5.1. LSTM e selezione informazioni

Una milestone fondamentale è stata raggiunta implementando le LSTMs attraverso registrazioni degli eventi chirurgici passati. Le *long short-term memory* sono un tipo di rete neurale ricorrenti utilizzate nel *deep learning*. Le LSTM hanno una connessione di feedback, possono processare non solo singoli *data points*, ma intere sequenze di dati come video. Una comune LSTM è composta da una cella, un *input gate*, un *output gate* e un *forget gate*. La cella ricorda valori oltre intervalli di tempo arbitrario e i tre gate regolano il flusso di informazioni verso l'interno o l'esterno della cella (figura 14). Una unità LSTM è adatta per operare classificazione, processing, e fare previsioni basate su dati in serie temporale. Questi algoritmi possono raggiungere un'accuratezza del 80/90%. Un'altra miglioria è stata quella di preformare le macchine insegnandogli a saper selezionare i dati video indipendente da cui ottenere le informazioni di cui hanno bisogno riconoscendo i pattern di cui hanno bisogno. Teoricamente le RNNs classiche possono tenere traccia di dipendenze arbitrarie a lungo termine nelle sequenze di input. Il problema è di tipo computazionale: quando le alleniamo utilizzando la *back-propagation* i gradienti a lungo termine hanno una probabilità di tendere a zero o infinito. [xvi] Per questo hanno bisogno di essere ottimizzate con le LSTM . Sembrerebbe strano lasciare che le macchine imparino da dati che hanno autogenerato ma invece ha migliorato sostanzialmente le performance. La *detection affidabile* della presenza di uno strumento in un frammento video venne resa possibile da nuovi sistemi in grado di riconoscere il contorno degli strumenti e perfino tracciando le punte e le estremità di essi in tempo reale. L'incorporazione di una memoria componente diede la possibilità di rinforzare la performance visuale.[xvii]

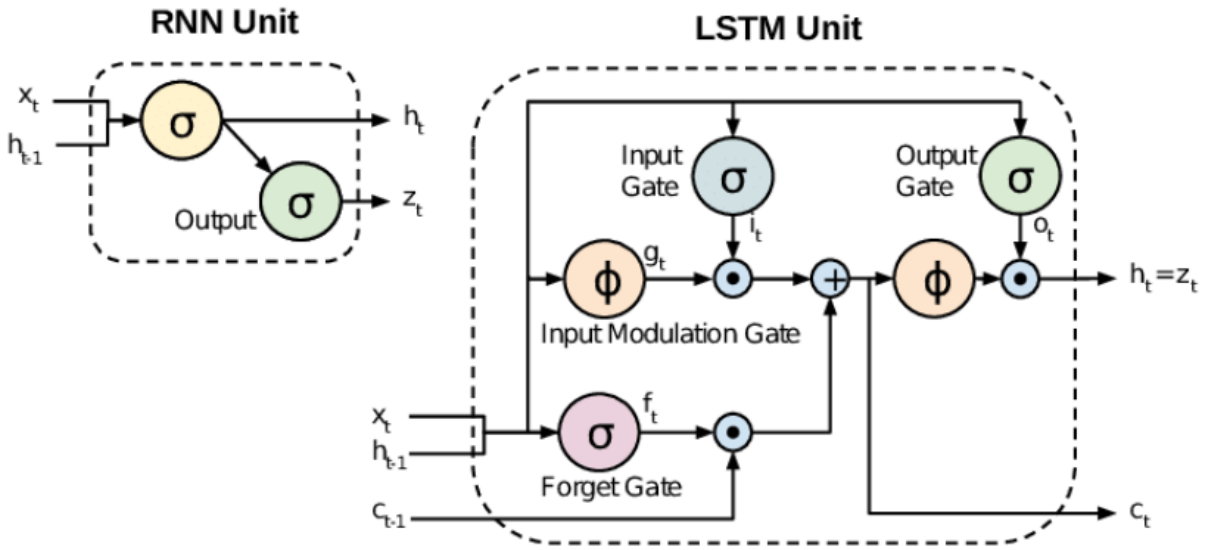

Figura 14

5.2. Recurrent Neural Network (RNN) vs LSTM

Le reti neurali ricorrenti sono state progettate per interpretare dati nel corso del tempo o informazioni sequenziali. Le RNN utilizzano dei *data points* per fare migliori predizioni. Queste lo fanno prendendo e utilizzando in input dati che poi vengono riutilizzati nei nodi precedenti o postcedenti nella sequenza per influenzare l'output.
Questo è molto importante come per esempio nell'estrazione di entità. Prendendo in considerazione l'esempio del correttore grafico, a livello più basso il sistema prende la parola che hai digitato come input. Utilizzando la parola come input, il sistema fa una predizione nel caso in cui lo spelling non fosse corretto. Se la parola non viene identificata nel database o se non è ben inserita contestualmente allora il sistema predice la miglior alternativa. La RNN non solo fa riferimento all'input inserito, ma effettua l'accesso alle funzioni di attivazione che sono scattate appena dopo la digitazione. Tutte le reti RNN hanno dei *feedback loops* nello strato ricorrente. Questo permette loro di mantenere la memoria, nel tempo. Può risultare complesso allenare una standard RNNs per risolvere problemi che richiedono dipendenze temporali a lungo termine. Questo perché il gradiente della funzione di perdita decade esponenzialmente con il tempo (chiamato il problema della scomparsa del gradiente).
Le reti LSTM sono una tipologia di RNN che utilizza unità speciali in supplemento a quelle standard. Le LSTM comprendono una cella di memoria che può immagazzinare e trattenere informazione in memoria per lunghi periodi di tempo. Un set di gates viene utilizzato per controllare l'uscita o l'entrata dell'informazione in memoria. Questo tipo di architettura permette di comprendere dipendenze a lungo termine.

5.3. Potenzialità e future direzioni

Una volta conferite capacità visuali e mnemoniche la prossima frontiera consiste nel definire una vera e propria capacità di comprendere la chirurgica. La nostra tecnologia odierna già riesce a finalizzare piccole rudimentali valutazioni come, per esempio, l'identificazione di eventi critici o evidenziare anatomie chiavi. Con i vantaggi della realtà virtuale simulata sta 'crescendo interesse verso le applicazioni di *surgical coaching* che provvedono validi feedback durante ogni tipo di procedura come, per esempio, il tempo impiegato o gli strumenti da utilizzare.[xviii] Il Computer Vision può inoltre evidenziare quali sono le zone sicure o pericolose di dissezione permettendo di seguire delle guide in realtà aumentata. [xix] Per poter essere in grado di servirsi del CV a scopo d'addestramento didattico/clinico i modelli di intelligenza artificiale hanno bisogno di quantità tali di informazioni che oggi non è semplice reperire. Parliamo di un migliaio di video per istituzione. Abbiamo la necessità di collezionare diversi dataset da diversi pazienti, patterns di pratica, e tecniche per minimizzare bias. Più otteniamo esempi di tecniche chirurgiche, varianti anatomiche e condizioni di pazienti meglio definiamo i nostri modelli e più accurati saranno. Abbiamo la necessità di migliorare la metrica per misurare la perfomance dei modelli oltre la semplice accuratezza. [xx] Non possiamo permetterci nemmeno l'errore dell'1% nel riconoscimento di strutture critiche. Bisogna essere in grado di allenare i modelli per localizzare l'anatomia e utilizzarla nelle loro decisioni. In questo modo l'intellligenza artificiale oltre a divenire la definitiva guida per operazioni chirurgiche sarà anche in grado di guidarci tra i casi e le complicazioni nel modo più sicuro e prevedibile possibile.

5.4. Machine Learning

L'apprendimento automatico è una categoria dell'informatica che appartiene all'intelligenza artificiale. Questo ramo comprende numerose modalità e tecniche per essere realizzato. Il *machine learning* si pone come obiettivo di poter realizzare scelte basate su un approccio induttivo in modo tale da poter finalizzare decisioni su esempi o problemi mai incontrati prima. Attraverso un addestramento o propriamente detto learning si forniscono più dati possibili in input e in output affinché la macchina ne possa estrarre un metodo sempre valido. Gli esempi di addestramento si assume appartengano ad una certa distribuzione di probabilità. La macchina così ha il compito di costruire un modello generale sullo spazio delle occorrenze per poter produrre previsioni sufficientemente accurate su nuovi casi.[xxi] I compiti dell'apprendimento vengono classificati in :

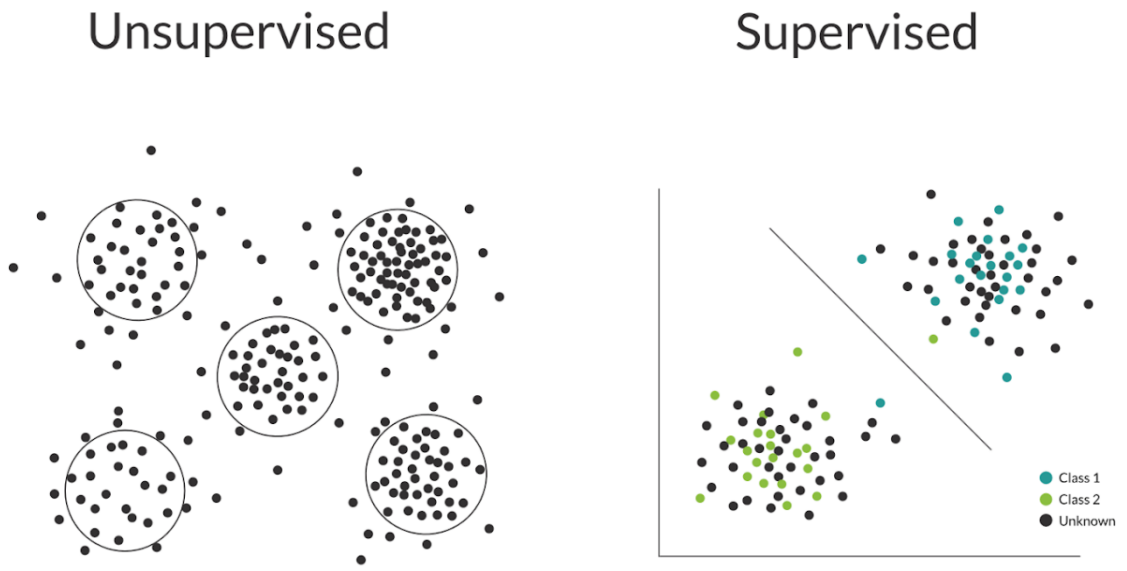

Figura 15

- apprendimento supervisionato (dati in input e output per estrarre regola associativa)

- apprendimento non supervisionato (il modello ha lo scopo di trovare struttura input)

- apprendimento per rinforzo (insegnante che indica raggiungimento obiettivo)

Un'altra categorizzazione può essere fatto in base il tipo di output desiderato:

- classificazione (gli input sono divisi in classi)

- regressione (output e modello continui ma sempre problema supervisionato)

- *clustering* (divisione in gruppi senza sapere i gruppi, non supervisionato)

5.5. Reti Neurali e deep learning

Il *deep learning* è un metodo specifico di *machine learning* che incorpora reti neurali in strati successivi per imparare dai dati in modo iterativo. Il *deep learning* è particolarmente utile quando si cerca di apprendere i modelli da dati non strutturati. Le reti neurali complesse di *deep learning* sono progettate per emulare il funzionamento del cervello umano, così i computer possono essere addestrati ad occuparsi di astrazioni e problemi mal definiti. Il bambino di cinque anni in media può riconoscere facilmente la differenza tra il volto del sorvegliante e quello del suo insegnante per l'attraversamento. Al contrario, il computer deve fare un sacco di lavoro per capire chi è chi. Le reti neurali e il *deep learning* sono spesso utilizzati nelle applicazioni per il riconoscimento delle immagini, di linguaggio e di visione artificiale.

6. Computer vision

Il *computer vision* è un campo dell'intelligenza artificiale che permette a computer e a sistemi di ottenere informazioni dettagliate tramite immagini digitali, video e altri input visivi. La visione computazionale permette ai robot di vedere, osservare e comprendere. Per funzionare hanno bisogno di enormi quantità di dati che vengono analizzate finché non ne vengono estratte determinate funzioni. Il ML utilizza modelli algoritmici per imparare a riconoscere dati in base il contesto. Gli algoritmi permettono alla macchina di imparare da sola senza che qualcuno stia a programmarla. Già dal 1950 gli scienziati hanno tentato di costruire computer che possano processare dati di tipo visuali. Con il *computer vision* vi sono stati enormi passi in avanti fino ai giorni d'oggi. Nel 2012 con la nascita di Alexnet i ricercatori hanno creato un algoritmo che superava qualsiasi tipo di algoritmo di riconoscimento.

6.1. AlexNet

Alexnet vinse il *computer vision contest* del 2012 con un'accuratezza del 85%. Nel core di Alexnet c'è una rete neurale convoluzionale o CNNs. La rete convoluzionale è un tipo di rete neurale utilizzata nel computer vision per riconoscere oggetti e immagini pattern. Questa è formata di diversi strati di neuroni artificiali. I neuroni artificiali sono funzioni matematiche che calcolano il peso somma di multipli input e output, calcolando un valore di attivazione. Il comportamento di ogni neurone è definito dal suo peso. Quando questo è alimentato dai valori dei pixel, il neurone artificiale ne estrae le caratteristiche. Quando inserisci un'immagine all'interno di una convet, ognuno dei suoi strati genera numerose mappe di attivazione. Le mappe di attivazione evidenziano le caratteristiche più rilevanti di ogni immagine. Ogni strato di neuroni prende un percorso di pixels come input, moltiplica il valore cromatico per peso, li somma e li processa tramite la funzione di attivazione. Il primo strato solitamente determina gli aspetti e linee, i contorni fondamentali come orizzontale, verticale e gli angoli. L'output del primo layer viene dato in input al secondo layer il quale ne estrare altre caratteristiche più avanzate come angoli e spigoli. Più a fondo ti muovi nella rete convoluzionale più iniziano a definirsi curve e poi veri e propri oggetti, facce e forme complesse. L'operazione di moltiplicare i valori dei pixels dal peso sommandoli è chiamato convoluzione.

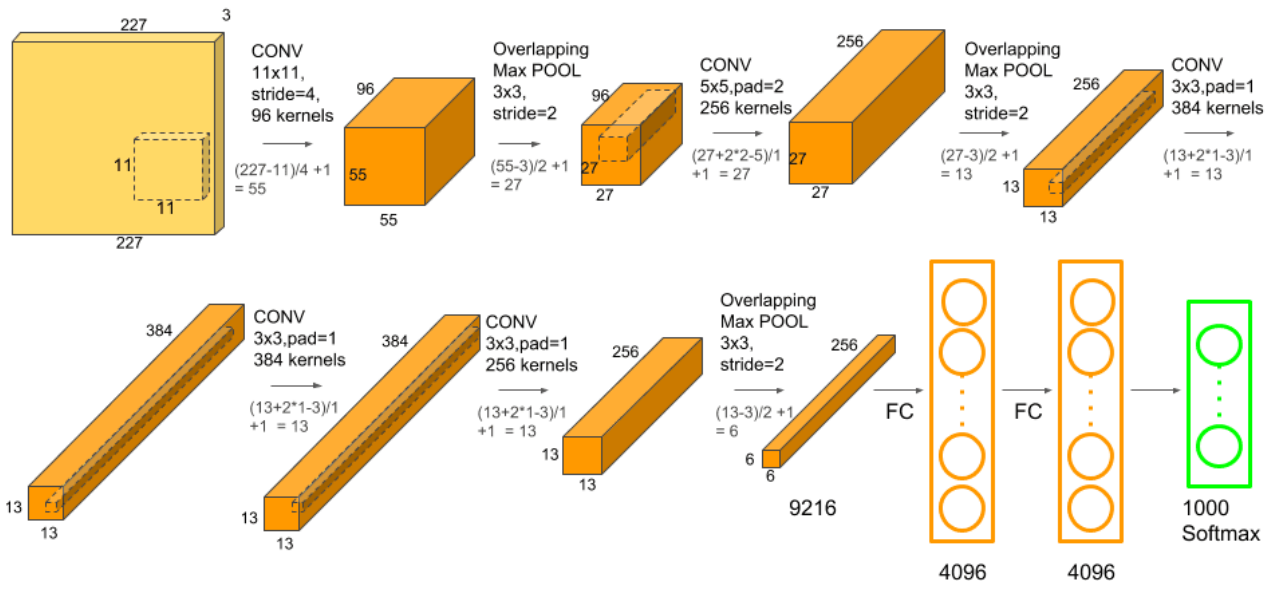

Figura 16

6.2. Convnet (CNN)

Una CNN a differenza di una RNN è composta da componenti come i fogli di convoluzione e altri. Il layer finale della CNNs è di tipo classificazione e svolge un ruolo fondamentale, sceglie la classe dell'output una volta che l'input è stato completamente filtrato. Il layer di classificazione emette un punteggio di confidenza tra 0 e 1 che specifica quanto un'immagine si avvicina allo standard della classe (figura 17). Una delle più grandi imprese nello sviluppo delle CNNs è aggiustare il peso dei neuroni individuali per estrarre i giusti aspetti dalle immagini. All'inizio la rete fu ideata con pesi random. Durante il training, gli sviluppatori si occupano di alimentare la rete neurale con enormi data set di immagini annotate con la classe corrispondente. La ConvNet processa ogni immagine con i suoi valori randomici e poi compara i suoi output con le etichette pre-fornite con l'output, come nell'apprendimento supervisionato. Se l'output non corrisponde alla classe indicata dall'etichetta la rete commette un piccolo aggiustamento al peso dei neuroni per migliorare il modello di classificazione. Ogni volta che eseguo un training dataset la rete aggiusta sempre più nel dettaglio il peso dei suoi neuroni finché ad un certo punto raggiunge il limite di convergenza. Solitamente i modelli dopo essere stati allenati li si può testare su un campione di dati. Se non c'è sufficiente varietà di dati ci troviamo in *overfitting* e abbiamo la necessità di ampliare il nostro spettro di dati. Il successo delle rete convoluzionali è dovuto al fatto la possibilità di potersi allenare su enormi quantità di dataset, e poi utilizzare modelli pre-allenati con piccoli aggiustamenti più specifici.

Nonostante la loro potenza e la complessità, le reti convoluzionali alla fine sono macchine che riconoscono patterns. Possono sfruttare enormi risorse di calcolo per scovare dettagli minuscoli e poco rilevanti che non possono essere notati dall'occhio dell'uomo. Quando si parla del riconoscimento di contenuti le cose diventano molto più complesse. Le reti neurali non hanno la possibilità di astrarre concetti o idee.[xxii]

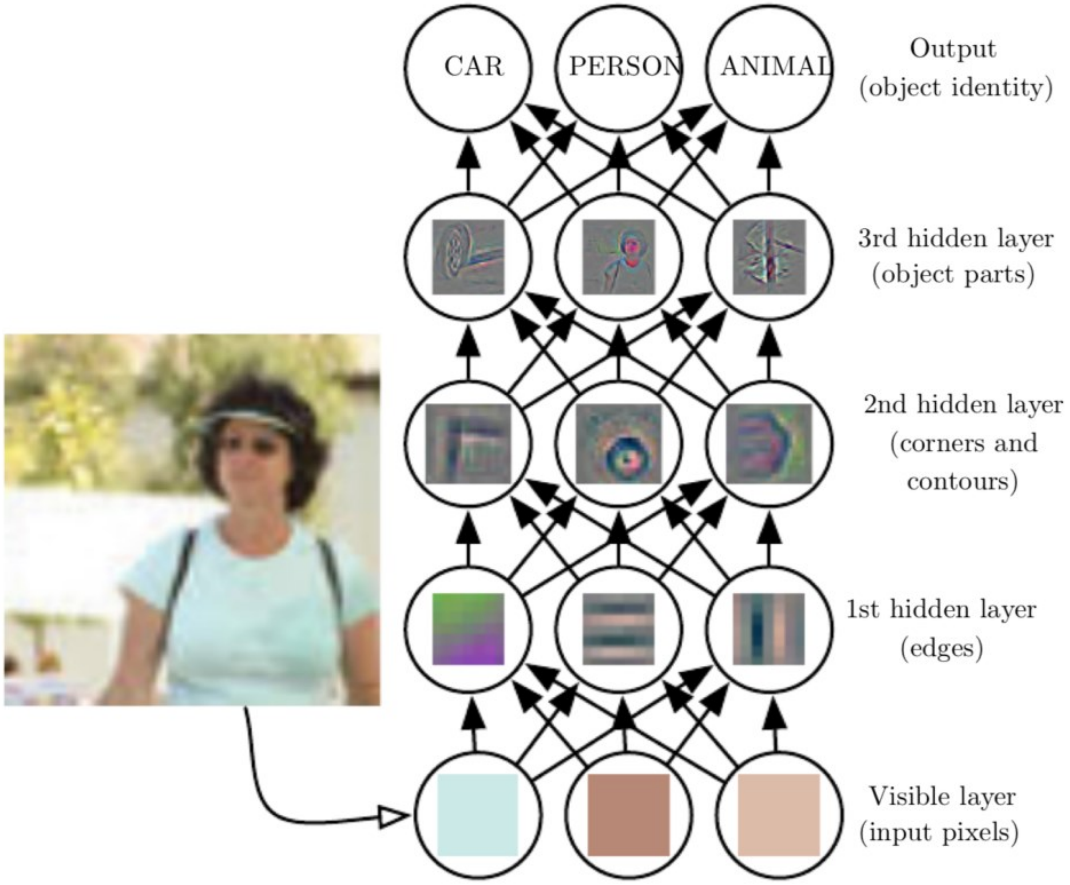

Figura 17

7. Conclusione

La *Intuitive Surgical* ha messo in commercio apparecchiature che possono essere considerate alla stregua delle più moderne tecnologie mediche. Sono 111 in Italia (22 solo in Lombardia) e circa 5.000 nel mondo le piattaforme chirurgiche da Vinci, in molti ambiti considerate oggi più efficaci e sicure della chirurgia tradizionale. Sebbene il robot da Vinci in Italia possa sembrare ancora una novità, dal 2018 sono stati effettuati oltre 20.000 interventi con l'uso di questa tecnologia, il 18% in più del 2017. Anche se robot come il Da Vinci continuano ad appartenere a delle fasce di prezzo alte, la sua diffusione nei centri ospedalieri più avanzati è una realtà affermata da tempo. A lungo termine, il suo costo viene ripagato rendendo non più necessarie tutte quelle procedure di recupero, cure antibiotiche e molte altre possibili conseguenze postoperatorie che si avrebbero avuto con la chirurgia tradizionale. Dobbiamo aspettarci nel tempo anche un'ottimizzazione a livello di interfaccia e software, in modo che il suo utilizzo non debba essere preceduto da lunghi corsi di training da parte degli utenti. Al giorno d'oggi ci sono numerosissime operazioni che non vengono più effettuate se non tramite questi tipi di macchinari. La robotica ci permettere di arrivare alle forme più delicate e complesse di intervento come la chirurgia urologica, toracica, ginecologica, laringoiatrica che altrimenti sarebbero impraticabili o risulterebbero eccessivamente invasive, traumatiche, pericolose. Con l'avvento dell'era dei big data diverrà sempre più facile sfruttare uno strumento potente come lo sono i dati, i video e le informazioni per le applicazioni di intelligenza artificiale nella robotica. Potremo lavorare tra simulazioni e controlli intraoperativi di qualsiasi tipo in modo da affrontare con sicurezza anche le eventualità più delicate. L'automatizzazione delle forme più complicate di chirurgia non sarà più fantascienza. Stiamo facendo passi in avanti nello sviluppo di reti neurali convoluzionali per il riconoscimento di oggetti e molto altro ancora. Presto arriveremo a dare più libertà d'azione alle macchine, in questo modo sarà possibile ridurre notevolmente i tempi d'esecuzione, operare ancora con più precisione e immediatezza. Potremo disporre di una guida nella durata di tutto l'intervento, anche a scopo didattico e avere un riferimento per potersi consultare nelle fasi più critiche. Sarà possibile ottenere dei dispositivi autonomi in grado di poter risolvere e gestire situazioni di elevata urgenza e tempestiva disponibilità anche in assenza di chirurghi o di una vera e propria equipe medica.

8. Bibliografia

[i] H. Ashrafian1,2, O. Clancy3 , V. Grover3 and A. Darzi1,2,"The evolution of robotic surgery: surgical and anaesthetic aspects" 1 Institute of Global Health Innovation, Imperial College London, UK, 2 Department of Surgery and Cancer, Imperial College London, UK and 3 Department of Anaesthesia and Critical Care Medicine, Royal Marsden Hospital, London, UK Journal of Anaesthesia, 119 (S1): i72–i84 (2017)"

[ii]. Kwoh YS, Hou J, Jonckheere EA, Hayati S. A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery. IEEE Trans Biomed Eng 1988; 35: 153–6

[iii]Benabid AL, Hoffmann D, Lavallee S, et al. Is there any future for robots in neurosurgery? Adv Tech Stand Neurosurg 1991; 18: 3–45

[iv]Watanabe G. Are you ready to take off as a robo-surgeon? Surg Today 2010; 40: 491–3

[v]Marescaux J. [Code name: "Lindbergh operation"]. Ann Chir 2002; 127: 2–4

Marescaux J, Leroy J, Rubino F, et al. Transcontinental robotassisted remote telesurgery: feasibility and potential applications. Ann Surg 2002; 235: 487–92

[vi] Lum MJ, Rosen J, King H, Friedman DC, Lendvay TS, Wright AS et al. Teleoperation in surgical robotics – network latency effects on surgical performance. Conf Proc IEEE Eng Med Biol Soc 2009; 2009: 6860–6863. Rayman R, Croome K, Galbraith N, McClure R, Morady R, Peterson S et al. Long-distance robotic telesurgery: a feasibility study for care in remote environments. Int J Med Robot 2006; 2: 216–224.

[vii]Ashrafian H, Darzi A, Athanasiou T. A novel modification of the Turing test for artificial intelligence and robotics in healthcare. Int J Med Robot 2015; 11: 38–43

Ashrafian H. Intelligent robots must uphold human rights. Nature 2015; 519: 391

[viii]M. Diana and J. Marescaux Research Institute Against Cancer of the Digestive System (IRCAD), European Institute of TeleSurgery (EITS) and International Institute for Image-Guided Surgery (IHU Strasbourg), Strasbourg, France Correspondence to: Professor J. Marescaux, IRCAD–Hôpitaux Universitaires de Strasbourg, 1 Place de l'Hôpital, F-67091 Strasbourg, France "Robotic Surgery"

[ix]https://www.intuitive.com/

Jonathan Pitocchi INTEGRAZIONE DELLE IMMAGINI ECOTOMOGRAFICHE NEI SISTEMI DI CHIRURGIA ROBOTICA: STATO DELL'ARTE E SVILUPPI FUTURI

[x]Nicolau S, Soler L, Mutter D, Marescaux J. Augmented reality in laparoscopic surgical oncology. Surg Oncol 2011; 20: 189–201.

D'Agostino J, Diana M, Soler L, Vix M, Marescaux J. Three-dimensional virtual neck exploration prior to parathyroidectomy. N Engl J Med 2012; 367: 1072–1073

[xi]Mauri G, Cova L, De Beni S, Ierace T, Tondolo T, Cerri A et al. Real-time US–CT/MRI image fusion for guidance of thermal ablation of liver tumors undetectable with US: results in 295 cases. Cardiovasc Intervent Radiol 2014; [Epub ahead of print].

Shekhar R, Dandekar O, Bhat V, Philip M, Lei P, Godinez C et al. Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. Surg Endosc 2010; 24: 1976–1985

[xii]M. Diana and J. Marescaux Research Institute Against Cancer of the Digestive System (IRCAD), European Institute of TeleSurgery (EITS) and International Institute for Image-Guided Surgery (IHU Strasbourg), Strasbourg, France Correspondence to: Professor J. Marescaux, IRCAD–Hôpitaux Universitaires de Strasbourg, 1 Place de l'Hôpital, F-67091 Strasbourg, France "Robotic Surgery"

[xiii] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inform Process Syst. 2012:1097e1105.

Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436e444.

[xiv]Natarajan P, Frenzel JC, Smaltz DH. Demystifying Big Data and Machine Learning for Healthcare. Boca Raton (FL): CRC Press, Taylor & Francis Group; 2017.

[xv]Ward TM, Hashimoto DA, Ban Y, et al. Automated operative phase identification in peroral endoscopic myotomy [e-pub ahead of print]. Surg Endosc. 2020. https://doi.org/10.1007/s00464-020-07833-9. Accessed September 20, 2020. 17. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. Surg Endosc. 2020;34: 4924e4931.

[xvi] Jin Y, Dou Q, Chen H, et al. SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network. IEEE Trans Med Imaging. Zisimopoulos O, Flouty E, Luengo I, et al. DeepPhase: Surgical phase recognition in CATARACTS videos. In: Frangi AF, Schnabel JA, Davatzikos C, Lopez Alberola C, Fichtinger G, eds. Medical Image Computing and Computer Assisted Intervention e MICCAI 2018. New York: Springer International Publishing; Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann

Surg.Kitaguchi D, Takeshita N, Matsuzaki H, et al. Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: Experimental research. Int J Surg. Twinanda AP, Mutter D, Marescaux J, de Mathelin M Padoy N. Single- and multi-task architectures for surgical workflow challenge. Jin Y, Li H, Dou Q, et al. Multi-task recurrent convolutional network with correlation loss for surgical video analysis. Med Image Anal. Ban Y, Rosman G, Ward T, et al. Aggregating long-term context for learning surgical workflows; 2020. http://arxiv.org/abs/2009

[xvii]. Yu T, Mutter D, Marescaux J, Padoy N. Learning from a tiny dataset of manual annotations: a teacher/student approach for surgical phase recognition; 2019. http://icube-publis.unistra.fr

[xviii] Malpani A, Vedula SS, Lin HC, Hager GD, Taylor RH. Effect of real-time virtual reality-based teaching cues on learning needle passing for robot-assisted minimally invasive surgery: a randomized controlled trial. Int J Comput Assist Radiol Surg. 2020;15:1187e1194.

[xix] Ban Y, Rosman G, Ward T, et al. Aggregating long-term context for learning surgical workflows; 2020. http://arxiv.org/abs/2009.00681. Accessed September 3, 2020.

Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann Surg. 2019;270:414e421.

[xx] Gordon L, Grantcharov T, Rudzicz F. Explainable artificial intelligence for safe intraoperative decision support. JAMA Surg. 2019;154:1064e1065.

[xxi] https://www.ibm.com/it-it/analytics/machine-learning

[xxii] https://bdtechtalks.com/2020/01/06/convolutional-neural-networks-cnn-convnets/