Article No. 6

AI AND REINFORCEMENT LEARNING IN PRECISION MEDICINE

Shyamsree Nandi
Function Head, Data and Analytics (Healthcare & Life Sciences) Tech Mahindra Ltd

Abstract: Healthcare and Life Sciences have begun realising disruptive innovations using a combination of Big Data, AI and Machine Learning. One of the big motivations of this industry for the past couple of decades has been the need to personalise medicine and treatment. Whether we talk about "value-based care" or novel treatment approaches specific to an individual's genome pattern, the use case that is subtly underlined across all of these subjects is about measuring, predicting and self-learning models for treatment pathways and associated outcomes. The estimated market size of personalised medicine worldwide is projected at 2.77 trillion U.S. dollars by the year 2022 at a steady CAGR of about 12%.

Precision or Stratified Medicine holds great promise for humankind and has seen tremendous adoption rates due to the fact that it leads to tailored interventions and hence much better outcomes across different patient groups. That said, there is a significant amount of challenge in scientifically arriving at personalised treatment plans. Can Big Data and AI be the fundamental backbone of such methods? In this paper, we will explore the possibilities of using a sophisticated Reinforcement Learning technique viz. Markov Decision Process (MDP) on datasets residing in a Data Lake to guide stratified medicine. This solution helps in achieving a better response, higher safety margins and lower treatment costs.

MDP is uniquely suited to the parlance of medical science. At each *stage* of a disease progression the doctor has a set of *actions* to choose from and based on the action chosen, the patient is transitioned to another state. A sequence of such states may eventually lead to either a positive or negative outcome. At any given point in time, the desired outcome of the treatment/ intervention will be to maximise the *reward* or chances of achieving a positive outcome. Patient Data collected across different states of the disease can be consolidated, curated and standardised in a Big Data Lake. MDP can then be used across hundreds of genomic sequences to identify the actions that can lead to the *best* treatment outcome.

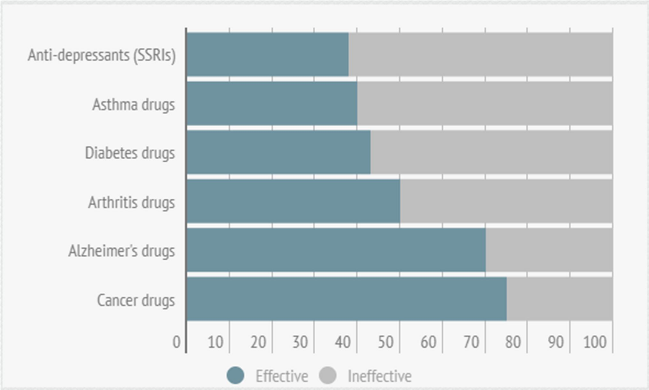
The fundamental gap that exists in care delivery today is the lack of a guided mechanism to assist doctors' decision of choosing appropriate treatment pathways. The methodology involves analysis of clinical datasets that provide necessary information to trace longitudinal view of diseases, followed by an architecture view of how to store and organise datasets, and disease-specific MDP modelling. Leveraging AI the models adapt over time leading to a more pragmatic guess about an intervention that is likely to result in the best possible outcome for a patient with specific clinical features and disease history.

Connecting all past events in a patient's medical journey with future outcomes is the uniqueness of MDP driven clinical decision support system and hence when coupled with the experience of a practitioner leads to a holistic analysis of expected outcomes ahead of time. Reinforcement Learning within AI thus offers an alleyway to maximise care delivery performance with far-reaching implications for the society.

Keywords:

Precision or Stratified Medicine – An Introduction

The goal of precision medicine is to devise novel; personalized treatment approaches for the subgroup of patients bearing unique intergroup characteristics but similar intragroup behaviour and maximising the positive responses to treatments within different subgroups. It is well known that treatment results may vary across different patients regardless of their disease type and despite giving them similar therapies purely attributable to their underlying genetic presentations and a complex combination of the response triggered by their body which is very individual to them. The aim of precision or stratified medicine is to firmly establish a deeper understanding of the disease not just at a universal level of the disease type itself but also at the level of disease progression and biomarkers specific to certain individuals.



1) Source: Association of the British Pharmaceutical Industry

Statistics show that a specific drug is never 100% equally effective across different patients for the same disease. The primary reason for this deviation can be attributed to the fact that diseases don't quite follow the same pathway in different people and hence the drugs are really treating "different diseases" even when given for the same condition. The graph on the left shows the average percentage of effectiveness of a particular drug in a class across the patient population. As can be seen, in the case of Asthma only 40% effectiveness is seen. Stratified medicine can help bridge or reduce this gap by bringing targeted treatment pathway to the right patients at the right state.

One of the typical methodologies followed involves use of predictive biomarkers to segment patient subgroups that are likely to experience maximum effectiveness (best outcomes or least possible adverse events) from an intervention. Following is a brief summary of the steps followed:

- Marker Discovery and variable selection which will be measured
- Retrospective Data Analysis and designing of Stratum Discovery Studies
- Assay design and development considering assay accuracy and variability
- Strata Definition using multiple marker profiles and leveraging the existing clinical knowledge base
- Stratum Verification using both retrospective and prospective data

The above steps help define the subgroups, and eventually, new clinical studies can be launched with targeted objectives for these individual strata. If we now extrapolate this framework to a normal caregiving scenario, say in a typical hospital, it is possible to identify the stratum to which a patient belongs and using retrospective data gauge the clinical outcomes observed for specific interventions on that stratum.

This is the fundamental concept which can be exploited for tailoring treatment to achieve maximum effectiveness or reduce the propensity of adverse effects of treatment. In many cases, this may help control the progression of diseases, enhance the quality of life or delay the occurrence of catastrophic events, if not heal the patient completely.

Challenges of Targeted Therapy Identification

Although the benefits of this approach are many, there are significant challenges in implementing this framework in real life. One of the challenges stems from the fact that there are thousands of diseases, each of which is a complex group of many subtypes and hence targeting would mean considering many combinations of diseases for different patient subgroups. The sheer volume of data required for such analysis is extremely difficult to aggregate and computationally expensive, if not prohibitive. An approach that could be followed is to look at chronic diseases which account for 80% of healthcare costs and hence positively impact a larger set of the patient population.

Another challenge is that the datasets required for holistic analysis leading up to precision medicine, ideally, do not limit to structured data alone. They may include elements of unstructured or streaming datasets. For instance, glucometer datasets or prescriptions written by doctors may have great importance in effectively identifying states of Diabetes Type 2. This calls for a sophisticated system which can handle a variety of datasets and can also support extending these sources to newer ones more easily. Not just variety, there might also be the need to integrate high-velocity real-time data, for instance in the case of critical care where the decision needs to be taken on the spur of the moment.

A third challenge is an outcome of the fact that stratified medicine needs to consider the whole trajectory of a disease in coming up with targeted intervention and not just the defining states. Since every state is definitive of the subgroup to which a patient would belong, it is essential to capture the complete trail in arriving at clinical decisions. This leads to the complex behaviour of the statistical models that support this kind of analysis. With every disease state, the models become manifold more complex with several decision points that need additional data.

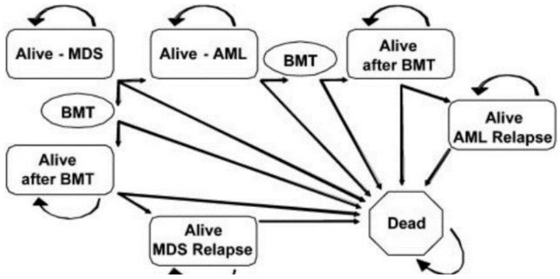
While these challenges are significant road blockers, modern day data architectures coupled with more computationally evolved Analytics tools and technologies have helped make Stratified Medicine a finite scope problem.

Understanding Disease States and Progression

Most of the diseases follow two unique transition steps. To begin with, the patient may belong to a healthy state and as the disease progresses they move across a predisease state to a Disease State.

Normal State
Predisease State
Disease State

The objective should, therefore, be to prolong the time taken to reach the "Disease State" or cure it completely by reversing back to the normal state.



2) Acute Myeloid Leukaemia Disease States

Consider figure (2) which shows different states that are part of acute myeloid leukemia (AML). The starting point is considered to be MDS (Myelodysplastic syndrome). These states specifically represent the outcomes of a potential BMT (Bone Marrow Transplant) as the intervention. The arrows represent state transitions from one state to another.

Ideally, the state transitions of disease would also involve probabilistic scores given that some states are more likely to follow a current state than others. This can be thought of as the probability of the edges.

We can further blow this diagram up if we included non BMT interventions too and observed the state transitions using those. Many modalities are possible such as chemotherapy and radiotherapy rather than BMT alone. There could also be a combination of modalities used which is quite common in the practical scenario.

At every state, a host of disease parameters may be captured which will potentially come from multiple systems such as Electronic Health Records, Lab Records and Images, Clinical Notes, Inpatient and Outpatient encounter records etc. The state itself is a representative outcome of all of the observed parameters for the patient at that state.

Markov Decision Process – Setting the Context

Consider a process which has multiple discrete states and a set of actions that determine the transition from one stage to another. MDP is a Reinforcement Learning technique within AI which can be used for modelling decision making where outcomes are partially random and partially under the control of a decision maker. It involves an agent which is interacting with an observable environment that responds continually by way of rewards to the actions of the agent and then presents new situations to the agent.

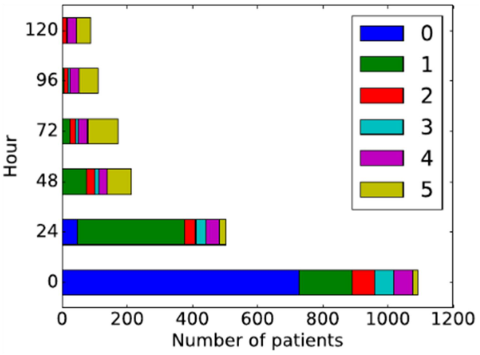
MDP assumes that the outcome of applying an action at a particular state depends only and only on the current state and not on the preceding states. Essentially MDP is memoryless. In the context of this paper let's consider a typical disease such as Lung Cancer. While it would be inappropriate to assume that the next state of Lung Cancer would depend only on the current state of the patient, it can be argued that the current step is an amalgam of all the previous disease events. Hence any new state based on an action taken, which could be a treatment like radiation would be dependent on the current state of the disease in the patient.

The reward provided at a specific state for a given action could be defined as the outcome of intervention with reference to the context of stratified medicine. Some examples of a positive outcome are the lapse of cancer or any disease, reduction in risk of morbidity or lower cost etc. MDP attempts to learn the decision process by leveraging these rewards. There could also be negative rewards such as the event of "death" which may be attributed to a terrible outcome. In the example provided as part of the previous section, AML relapse could be treated as a negative reward state with Death bearing the most negative reward.

MDP can either be deterministic or non-deterministic. For stratified medicine where we are trying to model disease trajectory and behaviour, there is a high amount of uncertainty involved. Additionally, many of the states may not be observable in the first place, i.e. they may not occur or could be too formidable to get to. So for the most part of modelling related to precision medicine, we can say that ask of the algorithms would be stochastic in nature.

The MDP Approach to Stratified Medicine

The dataset required for MDP acquires the form of tuples <*s, a, r*> which define the *current state* as *s,* an *action* at that state as *a,* which results in a *reward r* for moving into *next transition state s'* at a *treatment period t.* Based on retrospective data we can arrive at the state transition matrix which determines the relative probability of transitioning from one state to another. Consider the following example that summarises the disease states associated with coagulopathy in trauma. It is calculated simply by dividing the number of patients who move from one state to another out of the total number of patients within a specific time period. The observations are summarised in the graph on the left



3) Number of patients in each state across an observed set of hours

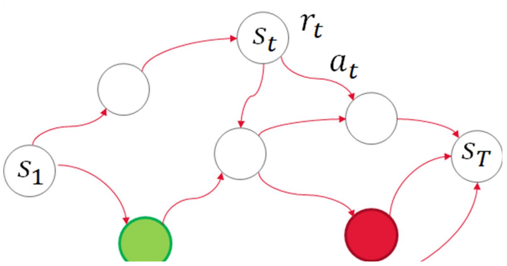
for coagulopathy in trauma.
Based on these observations following is the state transition matrix constructed:

	S0	S1	S2	S3	S4	S5
S0	0.156	0.768	0.017	0.013	0.046	0
S1	0	0.57	0.05	0.004	0.017	0.358
S2	0	0.121	0.757	0	0.07	0.052
S3	0	0.095	0.037	0.772	0.047	0.049
S4	0	0	0	0	1	0
S5	0	0	0	0	0.056	0.944

4) State Transition Probabilities for Coagulopathy in Trauma

Here S_0 through S_5 are disease states. If our objective is to achieve the best outcome for a patient, we will try to come up with an *optimal policy* that minimised the chances of fatality in any state. This optimal policy could be the "ideal treatment action" that we are choosing for a specific patient at a given state.

In stratified medicine, we can create such transition matrices for different diseases and then use Markov Decision Process to make appropriate clinical decisions at each of those stages. Following is a brief overview of how this can be mathematically accomplished.



5) Markov Decision Process showing states and rewards

Suppose the treatment period t $\in \{1, 2, .., T\}$, i.e. it is one of the time periods from 1..T. If each of the states is associated with a reward r_t, then the total reward or the expected utility at the end of treatment will be the sum of all of the rewards which have been received across the intermediate disease states. In the schematic depicted in figure (5) above the green state is associated with a positive reward such as in cases where the disease may be recovering and where the patient's condition is significantly improving. On the other hand, the red state may be associated with a negative reward which might be cases such as cancer recurrence. The idea is to maximise the overall utility of the whole treatment process.
This expected utility can be represented as which needs to be maximised:

$$E\left[\sum r_t(s_t, a_t,s_{t+1}) + r_{T+1}(s_{T+1})\ \right]$$

Using Bellman's recursive equations, we can solve this problem by:

$$V_t(s) = \sum_{s' \in S} P_t(s'|s,a)\left(r_t(s,a,s') + V_{t+1}(s')\right) \qquad for\ t = 1, 2, .., T$$

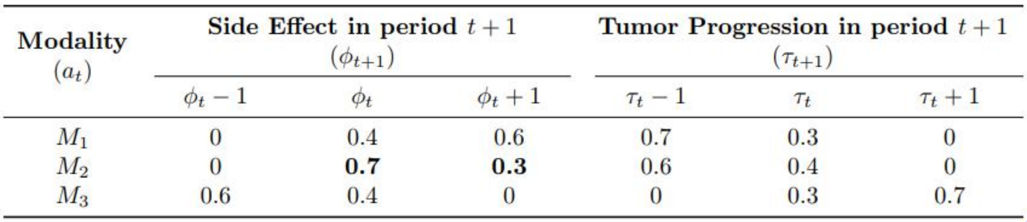
This value V_t is the basis that needs to be considered when any intervention is being planned for a patient at a certain state. The probabilities mentioned P_t in the above equation are the state transition probabilities that we earlier saw in figure (4) above.

The rewards r_t can be intelligently decided using the clinical domain knowledge. For instance, we can set a high positive reward for a state transition from metastatic cancer to a cancer state that is showing remission. Reward or Patient Utility itself could be a function of many parameters. In the case of cancer, we can potentially define it as a combination of tumour shrinkage and the side effects of the treatment. This reward function may be individual to each disease type and can be defined statistically. It would typically be a linear equation of multiple clinical parameters each of which in some way defines the positive impact of a treatment on a patient. The reward function may be representative of different tradeoffs such as efficacy vs side effects anticipated from the treatment or the improvement in the quality of life.

Optimal Treatment Policy

The Optimal Treatment Policy is closely related to disease state transition matrix. The reason for this is that if there is a favourable state, then the optimal treatment policy will try to include it in the maximum proportion of treatment pathways. It is directly related to the fact that the reward function is maximised and the utility is highest when these favourable states are included.

Let's try to visualise this impact using the example below of side effect vs Tumor progression. The objective is to minimise the side effect and reduce the tumour as much as possible.



Modality (a_t)	Side Effect in period $t+1$ (ϕ_{t+1})			Tumor Progression in period $t+1$ (τ_{t+1})		
	$\phi_t - 1$	ϕ_t	$\phi_t + 1$	$\tau_t - 1$	τ_t	$\tau_t + 1$
M_1	0	0.4	0.6	0.7	0.3	0
M_2	0	**0.7**	**0.3**	0.6	0.4	0
M_3	0.6	0.4	0	0	0.3	0.7

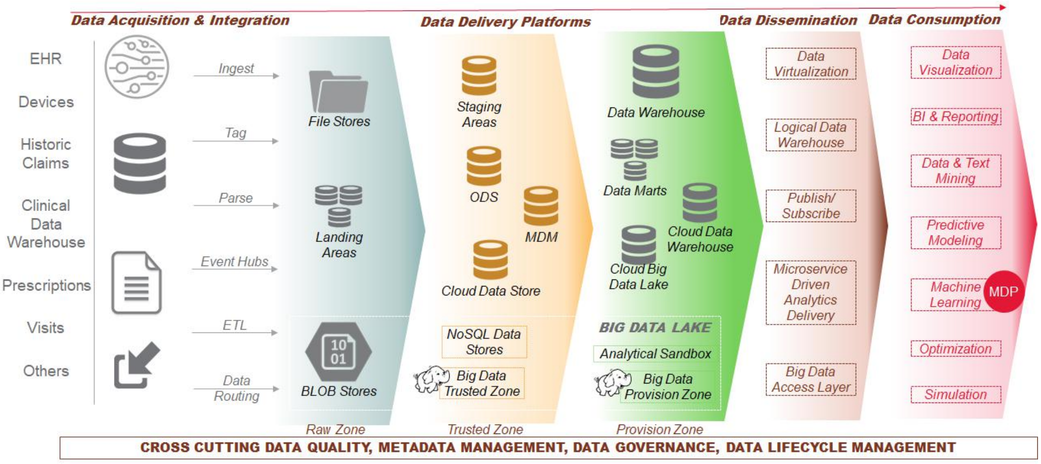
Modality shown above is different options that are available. For M_1 we see that there is an increased probability that at time t+1 there will be an increase in a side effect, but for M_2 we see that there is a decreased probability that at t+1 side effect will increase. So M_2 is more favourable in this case. The reward function can be devised in a manner that will make M_2 more risk-free and hence it will more often feature in the Optimal Treatment Policy.

MDP's value and policy iteration approaches could be used to devise more appropriate policies that are optimum for the treatment of a specific disease in a specific patient. Optimality can be measured in multiple ways. For instance, measures like QALY (Quality Adjusted Life Years) which are computed to be accumulated during the remaining life of a patient. Historic data can be used to further strengthen the understanding of Optimality in the context of treatment outcomes. Different treatment options can be evaluated by observing their outcomes from previous cases. The results of this analysis can be used to further improve the optimality conditions in future. Similar results could also be achieved through simulation.

Another aspect to consider is partial observability. For the most part, we assumed that the state s is known, but in clinical context sometimes it may be challenging to identify the state s precisely. As per the model discussed so far the action a is known only when s is also observable. Sometimes for specific disease, we cannot quite objectively identify a state s or the state differentiation is hazy. One example could be the in-between range of cancer that is about to be metastatic but still cannot be clinically proved to be so. In such cases finding the optimal treatment, the policy becomes tedious. For this, we use the POMDP (Partially Observable MDP) approach. In this case, adaptive policies are created which prescribe a choice of treatment modalities at each state for each time period.

High-Level Reference Architecture

Now that we saw a brief overview of how MDP can prove to be invaluable for devising targeted and well-founded treatment pathways, we can look at the underlying technical architecture that can support such an implementation. Given the sheer nature of the volumes, variety and velocity involved, a Big Data Lake can be a great way of aggregating rich information across different data sources.



Above is a high-level reference architecture that sums up the data journey and shows how it can be implemented in an enterprise.

The reference architecture consists of the following key components:

- Data Acquisition and Integration Layer
- Data Delivery Platforms
- Data Dissemination Layer
- Data Consumption Layer
- Data Management Layer

Data Acquisition and Integration Layer

The reference architecture requires several integration components to ingest and acquire data from clinical data sources typically used in hospital settings. Some of these source systems include Electronic Medical Records, Imaging Systems, Lab and other sources of information including clinical notes, device data etc. This layer focusses on the following functionality:

- ETL and ELT data integration patterns
- Usage of EAI services, big data ingestion or middleware in more real-time, low latency data integration scenarios
- Optimised design to subjectively handle small and large datasets, real-time and batch-oriented datasets

Data Delivery Platforms

This layer consists of the core data stores, marts, sandboxes, data lakes and warehouses which are used for onward delivery of data to run the MDP processes and other predictive models. The logical layers in the Data Delivery Platform are:

- Raw Data Zone: Data in the raw zone will be typically stored in a native form which will encapsulate the complexity on the data source side and will help in making the downstream architecture agnostic of the data sources
- Refined Data Zone: This zone will leverage some of the staging schemas from the existing landscape which can be reused, but the purpose of this zone will be to prepare data as per enterprise rules and not for specific business functions. This zone will be

used for standard operational reporting. This zone may involve more nimble updates and are more transactional.

- For Purpose Zone: This zone will be used by most of the data access components and will consolidate data more analytically across disease states. Most of the MDP processes would leverage data from this zone.

Data Dissemination Layer

Given that the datasets required for this analysis are enormous and have individual structures which may need to be harmonised for ease of analysis, the Data Dissemination Layer can be used to standardise and create disease-specific models that can be used by MDP processes. This will include the creation of semantic models which can integrate physically different datasets into logical structures suitable for analysis

Data Consumption Layer

This is the layer where all of the ML algorithms would reside and operate on data which has been aggregated in the preceding layers. The MDPs can be used as microservices where inputs can be disease state matrices, their transition probabilities and the reward functions. In a typical clinical setting, say in an Operation Theater where the doctor is trying to evaluate treatment pathways, the MDP models could be invoked through these modular services to produce the optimal treatment policy.

Data Management Layer

This layer would focus on ensuring the health of the overall data ecosystem presented above. It would include components that would manage master data and reference data such as the disease codes or procedure codes, patient demographic information etc. It would also include metadata management components that can help deal with diverse glossaries and also for audit and traceability from a risk and compliance perspective. It also deals with governance and lifecycle management of data given the sensitivity of the information involved in this analysis process. One of the important aspects is of Data Quality which needs to be highly accurate for running this type of analysis. So this layer would involve prepackaged data quality processes which would constantly monitor and manage data quality.

Applicability to Indian Geography

Whilst the idea of implementing MDP for treatment effectiveness and clinical decision making has far-reaching benefits for patients at large and for the healthcare delivery processes, it will be only as much success as the data provided to it. From the perspective of the Indian Healthcare Industry, it is important to realise that EMRs are yet to be firmly implemented across all care provider settings. Often times we may have disjoint information about an individual patient's journey across disease states. This may lead to potential errors in judging the effectiveness of the personalised intervention. Also, one single hospital or provider may not have all of the data for the patients as they tend to move to other providers from time to time for a second or secondary opinion.

That said, we can mitigate the challenge by leveraging the benefit of diverse data that Indian Healthcare data sources are likely to provide. The anticipated diversity and high volume data can be used to undercut the negative effects of incomplete datasets. The other approach is also to look at large hospital chains and develop the models in those settings which tend to be more regulated and to look at a select set of diseases, to begin with. Most of the large hospital chains in India do keep details of clinical records for patients. The bulk of this information is used purely for operational purposes today, but we could tap into the clinical datasets and also a couple that with partnership from diagnostics and clinics for better understanding of patient journeys.

Conclusion

This paper demonstrates how MDP which is a Reinforcement Learning technique in AI can be used to drive more accurate and sequential decision making. This process helps in weighing the merits of individual disease trajectories defined by a set of states and how these can be personalised because of responses or outcomes measured for specific patients. This paper also outlines the architecture which can enable MDP driven analysis at a massive scale.

The paper also outlines how Bellman's equation can be used to leverage complex reward functions that can help in measuring tradeoffs such as efficacy vs side effects for patient subgroups. Since patient characteristics differ and their responses to treatment modalities differ too, the outcome is highly specific and can be best handled with a more targeted treatment pathway underpinned by Stratified Medicine. We also saw how state transition matrices could be used to identify Optimal Treatment Policies.

Markov Decision Process thus holds great promise for more effective treatment, personalised for patients and hence better outcomes over traditional techniques. AI and Reinforcement Learning can be used to better assist doctor's judgment with a guided and pragmatic view of expected clinical outcomes.

References

- A decision analysis of allogeneic bone marrow transplantation for the Myelodysplastic syndromes http://www.bloodjournal.org/content/bloodjournal/104/2/579.full.pdf?sso-checked=true by The American Society of Hematology
- Identification of disease states associated with coagulopathy in trauma https://www.researchgate.net/figure/State-transition-matrix-These-are-the-probabilities-of-moving-from-one-state-to-another_tbl2_308547757 by BMC Medical Informatics and Decision Making
- Computational Reinforcement Learning http://web.mit.edu/course/other/i2course/www/vision_and_learning/reinforcement_learning_slides.pdf by University of Massachusetts Amherst
- Machine Learning by Tom M. Mitchell
- From Data to Optimal Decision Making: A Data-driven, Probabilistic Machine Learning Approach to Decision Support for Patients with Sepsis https://www.ncbi.nlm.nih.gov/pubmed/25710907 by Department of Computer Science and Genome Center, University of California

Acronyms and Abbreviations

#	Acronym	Description
1	AI	Artificial Intelligence

2	AML	Acute Myeloid Leukemia
3	BMT	Bone Marrow Transplant
4	EAI	Enterprise Application Integration
5	EHR	Electronic Health Record
6	ELT	Extract Load Transform
7	ETL	Extract Transform Load
8	MDP	Markov Decision Process
9	MDP	Myelodysplastic syndrome
10	ML	Machine Learning
11	QALY	Quality Adjusted Life Years
12	POMDP	Partially Observable Markov Decision Process