Module 2

AI, ML, Deep Learning and Data Mining Methods for Healthcare

2.0		**AI, ML, Deep Learning and Data Mining Methods for Healthcare**	10
	2.1	Knowledge discovery and Data Mining, ML, Multi classifier Decision Fusion, Ensemble Learning, Meta-Learning and other Abstract Methods.	
	2.2	Evolutionary Algorithms, Illustrative Medical Application-Multiagent Infectious Disease Propagation and Outbreak Prediction, Automated Amblyopia Screening System etc.	
	2.3	Computational Intelligence Techniques, Deep Learning, Unsupervised learning, dimensionality reduction algorithms.	

Arvin Agah - Medical Applications of Artificial Intelligence-CRC Press

Healthcare Ecosystem Fundamentals

NDHM Eco-system

Source: Practo

The 3 P's of Healthcare

- **Providers, Payors, Patients**
- **Providers-** Any person/ organization delivering care/ treatment
- **Payer-** Person/ Organization who "pays" for care services

Other Essential Stakeholders

- **Manufacturers-** Manufactures chemicals, devices, medications and consumables (gloves, syringes, etc.)
- **Group Purchasing Organizations (GPOs)-** Contract aggregator who reduces costs and provides value added services
- **Clinical Research Organizations (CROs)-** Conduct Clinical Trials for Drug Testing'; aid Drug Discovery Process
- **Pharmacy Benefit Managers (PBMs)-** Create approved drug lists with regulatory bodies and insurance firms for coverage, reducing medication costs

Global Healthcare Cloud Computing Market

21% of current HIT Spend is on Cloud Solutions

Cloud Market to reach **$64.7 billion** by 2025

79% of Healthcare Leaders are making cloud a strategic priority in 2022

Source:
https://blogs.perficient.com/2019/12/13/himss-survey-results-healthcare-and-the-cloud/

Why are Healthcare Organizations moving to Cloud?

Healthcare Challenges

- Increasing Consumerism of Healthcare- need for Stakeholder Engagement
- Imminent transition to Value Based Care
- Large Chronic Disease burden requiring long-term care
- Push for New Technology Adoption
- Need to comply with increasingly complex and evolving regulatory frameworks
- High cost of care

- Need for improved efficiency of care delivery
- Low Doctor:Patient requires improved collaboration
- Need for access to care in rural areas
- Increased Consumerism of Healthcare

Drivers of Cloud Adoption

Scalability

Flux in patient flow requires hospitals adaptable without having to purchase additional hardware

Ex: Sudden influx in patient inflow (like flu season)

Collaboration

Implementation of Value Based Care systems require seamless and user-friendly collaboration b/w stakeholders

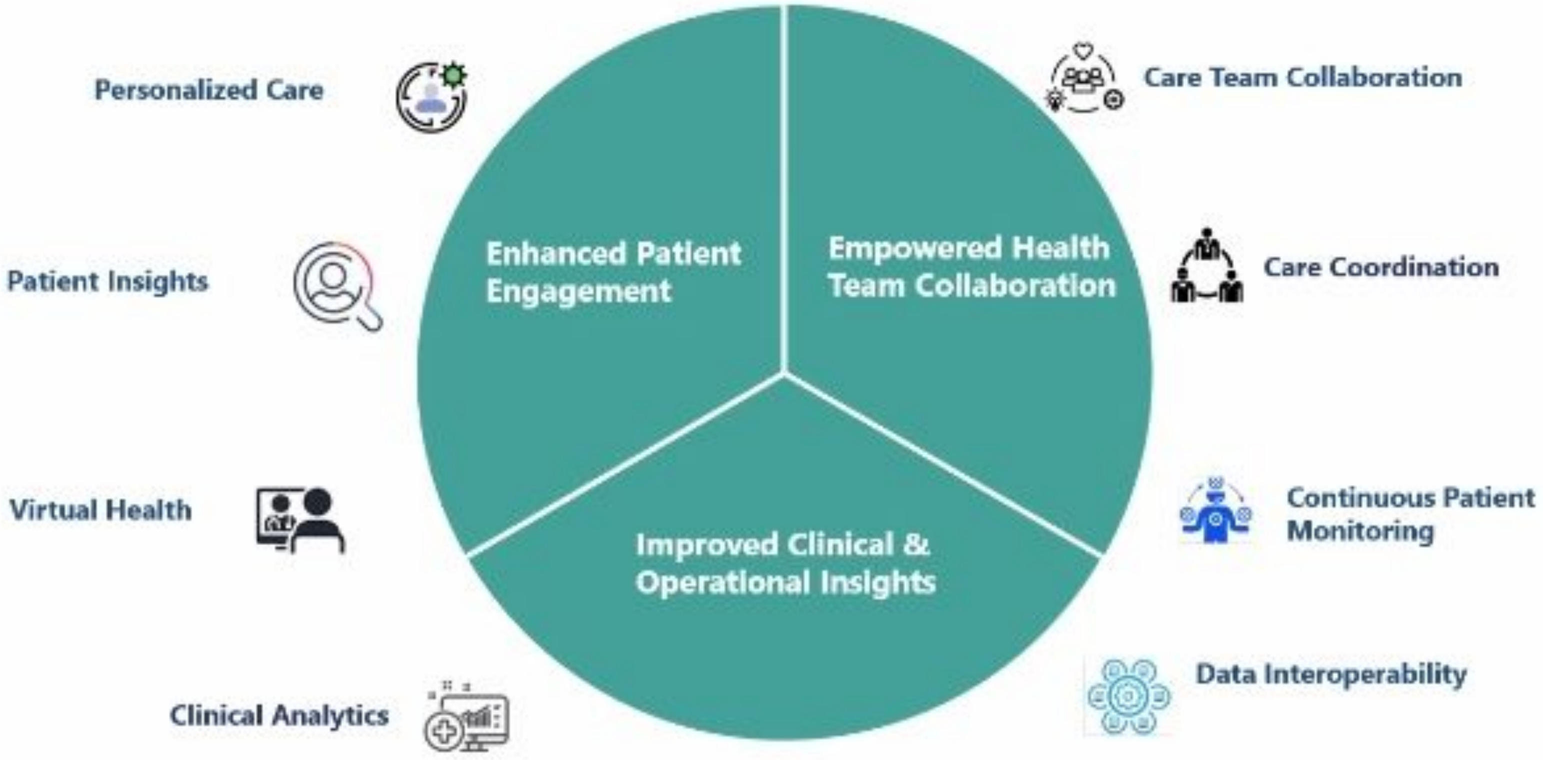
Security

- Security issues are managed and owned by vendors, reducing chances of such incidents
- Cloud systems include automation of backups and disaster recovery options

Driving Speed to Value

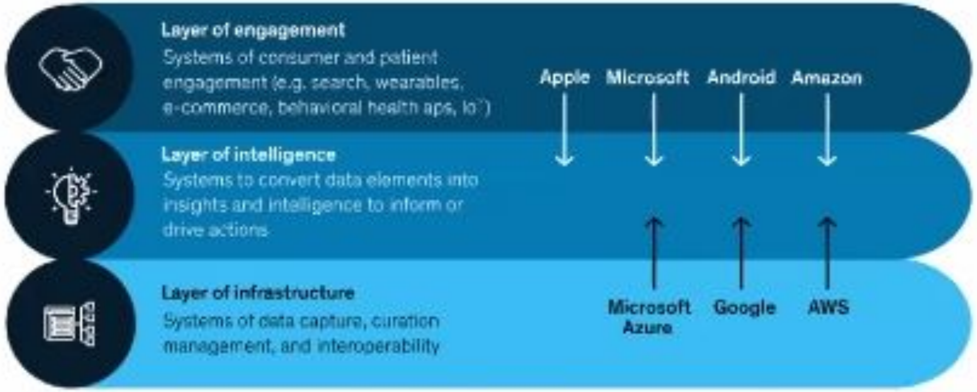
- Cloud provides a turnkey solution for building, releasing, and scaling digital healthcare products,
- Helps uncover negative space within the ecosystem - drive digital transformation across business functions and processes

Impact on Healthcare



Role of Cloud in Futuristic Healthcare Delivery

Technology giants are investing in capabilities across the layers of healthcare ecosystems.



McKinsey & Company

Sample Use Cases

Provider	Payer	Life Sciences
Virtual/ Home Health	Population Health	Pharmacovigilance
Remote Patient Monitoring	Member 360	Clinical Trial Cohort Analysis
Application Development- User Engagement	Multi Channel Communications	Disease Surveillance
Advanced Analytics Tools	FWA Analysis	Drug Target IDing
Health Data Lakes	Claims Analytics	Genomic Analysis
EHR on Cloud	Member Data Management	Patient Registry
IoMT	Real time eligibility verification	Biosensor Integrations

Healthcare Cloud Offerings

Cloud Migration

- Suitability Assessment
- Migration plan
- Migration
- Optimization

Cloud Essentials

- Hybrid/Multicloud foundation
- Cloud resource and cost governance
- Cloud IAM, Privacy and Security
- Cloud operations and monitoring

Cloud Engineering

- Cloud native application architecture
- DevOps maturity assessment & provisioning automation
- End user computing
- Edge computing/IoT solutions
- Data on cloud
- Analytics on cloud

SaaS Enablement

- Assessment
- Rollout plan & implementation
- Application support

EHR on Cloud

- EHR cloud readiness assessment
- Architecture & design
- Methodology

Key Partnerships

aws, Azure, Google Cloud, salesforce, Red Hat, RED HAT OPENSHIFT, snowflake, databricks, docker

Cloud Service Providers- Healthcare Offerings

	Google Cloud Platform	salesforce
Core Health IT	HIPAA & HITRUST compliant cloud architecture	HIPAA & HITRUST compliant cloud architecture
Applications on Cloud	Medical Imaging Suite, Cloud Migration of applications	CRM native Analytics Heroku: PaaS on app hosting & deployment
Clinical Systems Interoperability	Google's Cloud Healthcare API and data engine- mapping HL7v2 messages to FHIR across leading EHRs; Medical Device Connectors	Salesforce Health Cloud MuleSoft Anypoint Platform- Marketplace for APIs & API designer
Patient & Clinician Experience	• Rapid response Virtual Agent – COVID 19 • Visualization of HC services • Google Workspace for remote Provider Ops	• Experience Cloud: Connected data driven apps, sites & portals • CRM Platform to connect with patients on treatment plan and updates • Engages members with relevant data, resources, creates Patient 360 • Einstein GPT – generative AI for CRM • Patient Connect Platform- daily online patient survey
Medical Research	• FDA studies on Google MyCloud • Cloud Life Sciences API for genomic/biomedical data	Service Cloud: handles patient inquiries, provide support, and manage clinical trials/ research studies cases
Analytics & AI/ML	Plug & play dashboards for Disease prediction, Patient Experience, Health Data Analytics for Business Outcomes	• Einstein Discovery- Actionable AI driven insights • Scale Care Management- Patient Risk Stratification, Patient engagement, Utilization management analytics

Medical Coding with Amazon Comprehend

Amazon Comprehend

- A HIPAA-eligible natural language processing (NLP) service
- Uses machine learning pre-trained to understand and extract health data from medical text, such as prescriptions, procedures, or diagnoses.
- **Identify relationships among extracted health information and link to medical ontologies like ICD-10-CM, RxNorm, and SNOMED CT**
- Automate and lower the cost of processing and coding unstructured medical text with easy-to-use APIs
- **Implement patient data privacy solutions and identify protected health information (PHI) with a HIPAA–eligible service**

Process Flow – Extracting HC Data from Medical Text

Upskilling Roadmap

Foundations

- Cloud Computing Essentials: Cloud service models (IaaS, PaaS, SaaS), deployment models (public, private, hybrid, multi-cloud), and key cloud providers
- Network, storage and database essentials
- Visualization technologies used in Cloud (Containerization- Docker, Kubernetes; Hypervisors- VMware)
- Cloud Service Providers & their Offerings

Domain Knowledge

- Healthcare Ecosystem- Stakeholders & Processes
- Healthcare Terminology & Regulations (HIPAA, GDPR, etc.)
- Areas of Focus by stakeholder, their needs, goals, and challenges
- Technologies being used and their purpose
- Healthcare Applications and their use in Operations & Care Delivery
- Cloud-based solutions for healthcare
  - EHR systems & Care Applications
  - Areas of Patient data analytics
  - Telemedicine
  - Data Interoperability & Health Information exchange (HIE) platforms.

Cloud Architect

AWS Certified Cloud Practitioner	AWS Certified Solutions Architect- Associate	AWS Certified SysOps Administrator	AWS Certified Solutions Architect- Professional
Microsoft Certified: Azure Fundamentals	Microsoft Certified: Azure Administrator Associate	Microsoft Certified: Azure Solutions Architect Expert	Microsoft Certified: Azure DevOps Engineer Expert
Google Cloud Certified: Associate Cloud Engineer	Google Cloud Certified: Professional Cloud Architect	Google Cloud Certified: Professional Cloud Developer	
Salesforce Certified Administrator	Salesforce Certified Platform App Builder	Salesforce Certified System Architect	

Data Integration & Governance

AWS Certified Cloud Practitioner	Microsoft Certified: Azure Data Engineer Associate	Google Cloud Certified: Data Engineer	Salesforce Certified Data Architecture and Management Designer

Application Development on Cloud

AWS Certified Developer – Associate	Microsoft Certified: Azure Developer Associate	Google Cloud Certified - Associate Cloud Engineer	Salesforce Certified Platform Developer I & II

AI & Advanced Analytics

AWS Certified Machine Learning - Specialty	Microsoft Certified: Azure AI Engineer/ Data Scientist Associate	Google Cloud Certified: Professional Data/ML Engineer	Salesforce Certified Einstein Analytics and Discovery Consultant

AWS	Microsoft Azure
Google Cloud	Salesforce

2.1 Knowledge discovery and Data Mining, ML, Multi classifier Decision Fusion, Ensemble Learning, Meta-Learning and other Abstract Methods

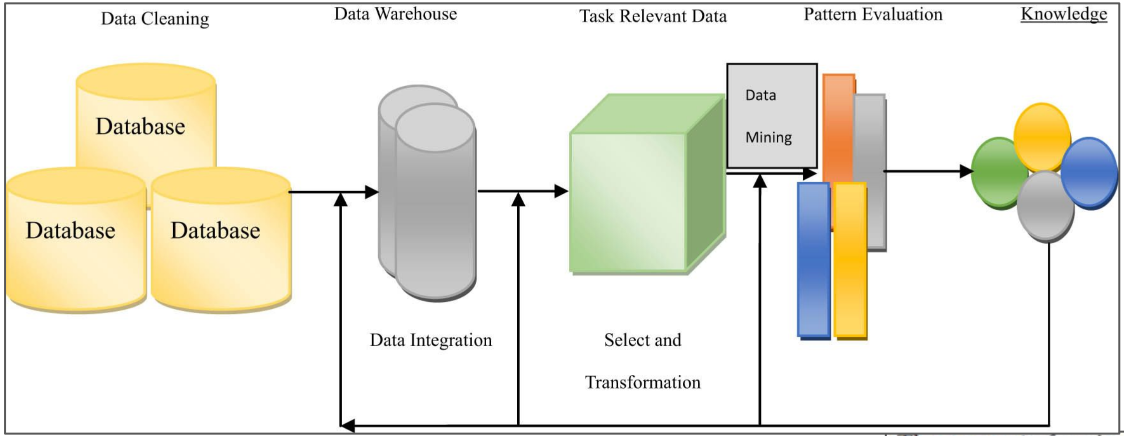
Knowledge Discovery and Data Mining

inherent prev: unknow info
to represent known.

3.2.1 Knowledge Discovery and Data Mining

The goal of data mining is to discover inherent and previously unknown information from data to represent as knowledge. In many cases, such methods are used to summarize and model very large data sets, capturing what salient (high support or occurring frequently) and interesting patterns reside in the data that may not have otherwise been discovered. Many of these methods were motivated by the growing demand to mine consumer transaction data to determine which items were purchased together, so that retail stores could better position their products within the store and track buying patterns. These methods therefore extend to online transactions and have therefore been heavily used by the Web community. As such, many algorithms leverage the transaction model for the core data structure and operating algorithmic assumptions. Two such methods are Association Rule Mining and Sequential Pattern Discovery.

Association Rule Mining



- knowledge in rule form.
- finds freq. itemsets.
- generates rules $X \to Y$
- Apriori Algo → deviation from independence
  → lift > 1.0 : X & Y coexist
- Application: (1) freq. diseases.
  (2) diagnosis & test.
  (3) Emergency dept
  (4) Hospital readmission Case Study

- produces knowledge in rule form
- It finds frequent sets of items (i.e., combinations of items that are purchased together in at least N database transactions) and generates from the frequent itemsets (such as {X, Y}) association rules of the form X → Y and/or Y → X.

The *support* of an itemset X is defined as

$$\text{Support}(X) = \frac{\text{number records with } X}{\text{number records in } D}$$

The *support* of a rule $(X \to Y)$ is defined as

$$\text{Support}(X \to Y) = \frac{\text{number records with } X \text{ and } Y}{\text{number records in } D}$$

The *confidence* of a rule $(X \to Y)$ is defined as

$$\text{Confidence}(X \to Y) = \frac{\text{number records with } X \text{ and } Y}{\text{number records with } X}$$

Apriori Algorithm Metrics Hierarchy

Lift

Strength of itemset association

Confidence

Likelihood of itemset co-occurrence

Support

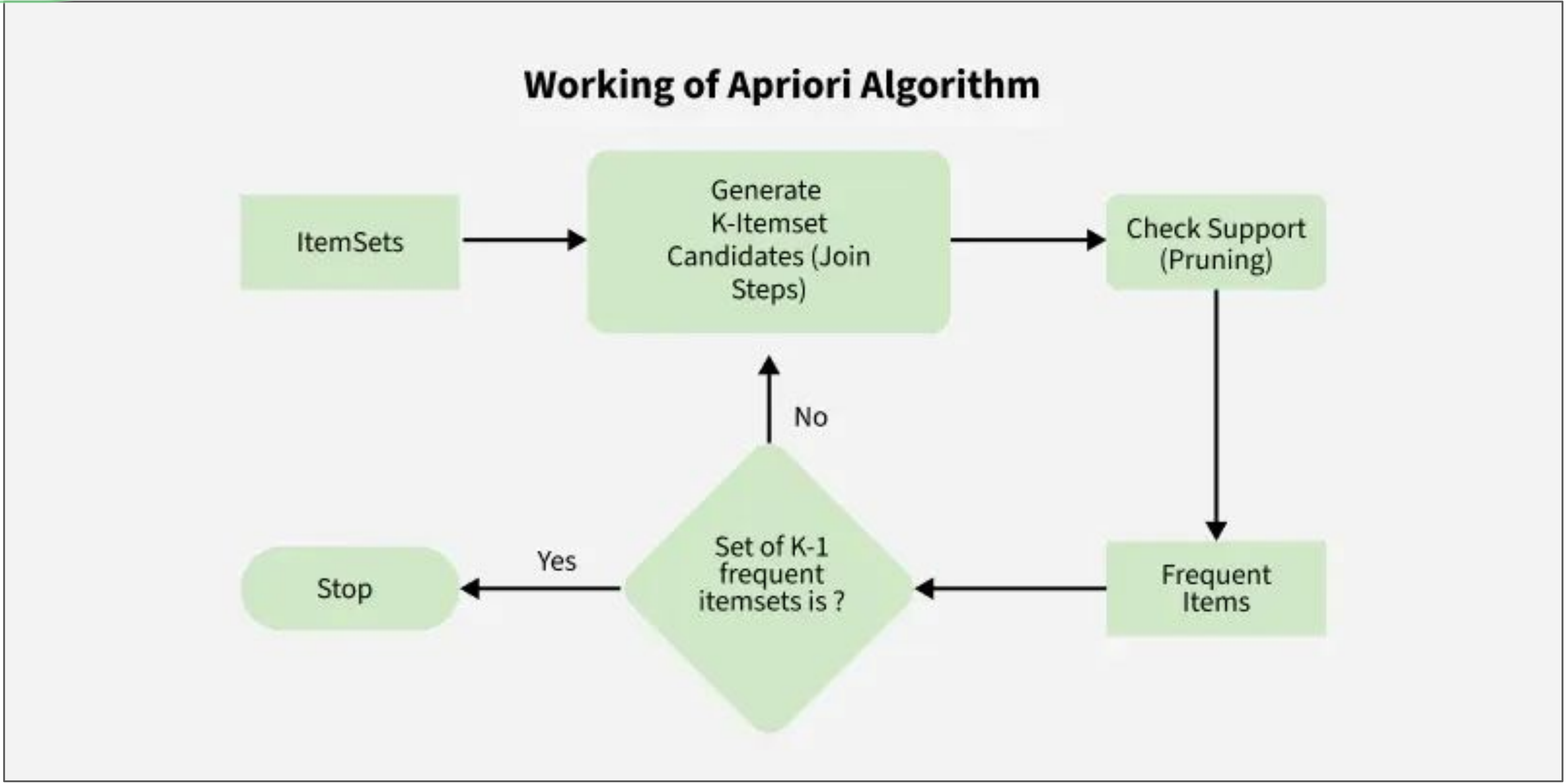
Frequency of itemset occurrence

Lift
strength of itemset association

confidence
Likelihood of itemset co-occurrence

Support
freq' of itemset occurrence

Apriori is the most widely used algorithm for finding frequent k-itemsets and association rules. It exploits the downward closure property, which states that if any k-itemset is frequent, all of its subsets must be frequent as well.



Association Rule Mining

The lift of a rule (X → Y) measures the deviation from independence of X and Y.

A lift greater than 1.0 indicates that transactions containing the antecedent (X) tend to contain the consequent (Y) more often than transactions that do not contain the antecedent (X).

The higher the lift, the more likely that the existence of X and Y together is not just a random occurrence but, rather, due to the relationship between them

$$\text{lift}(x \to y) = \frac{\text{confidence}(x \to y)}{\text{support}(y)}$$

$$\text{Lift}(X \to Y) = \frac{\text{Confidence}(X \to Y)}{\text{Support}(Y)}$$

$$L = \frac{C(x \to y)}{S}$$

Diagnosis, Prognosis,

Diagnosis, Prognosis, Prediction in medical field

A diagnosis identifies a specific illness or condition, a prognosis offers an informed prediction about how that condition is likely to progress.

A diagnosis is the identification of a disease, condition, or injury from its signs and symptoms. Your healthcare team may use your health history, a physical examination, tests like blood work, imaging like x-rays or a CT scan, and biopsies to help determine your diagnosis.

For example, if a patient comes to their doctor with symptoms like fever, a cough, and body aches, the doctor might run tests like a COVID-19 antigen or PCR test, a rapid influenza diagnostic test, and blood work. Depending on the results, the patient may be diagnosed with COVID-19, the flu, or another infection.

Based on the tests and other information gathered, the diagnosis may also be able to provide more details on your condition such as cancer type or the stage of the disease.

Diagnosis, Prognosis, Prediction in medical field

A **prognosis is a medical professional's educated prediction about the likely course and outcome of a disease or condition**.
While a diagnosis identifies a specific illness or condition, a prognosis offers an informed prediction about how that condition is likely to progress. It helps patients and their families anticipate outcomes, make treatment decisions, and plan for the future.

Prognosis is based on a combination of factors, including:

- The type and stage of the disease.
- How quickly the condition is likely to advance.
- Response to treatment options.
- The patient's overall health, age, and other medical conditions.

For example, your healthcare provider may talk about your prognosis in terms of a time range. They might say: "Based on survival statistics and your specific prognostic factors, you have a prognosis of three to five years to live."
An accurate assessment of prognosis differs from a **prediction** in that it requires stating a range of survival based upon relevant data.

Applications: ARM

ARM.

Due to the high patient volumes, **emergency departments** **(EDs)** are the main units of hospitals which may have **vast amount of the raw data of hospital's information system**. Besides, this **overcrowding increases the complexity in operational planning.**
Thus, **use of data mining becomes more of an issue in EDs to make effective decisions**

ARM is used to extract hidden patterns and relation between diagnosis and diagnostic test requirement in real-life medical data received from an emergency department. Apriori was used as an association rule mining algorithm.

ARM is one of the most important functions of data mining. It is a structured method of discovering all frequent patterns in a data set and forming noticeable rules among frequent patterns. In other words, it is a way of **discovering relations between items in big data.** In medical field, ARM is used to discover frequent diseases in specific areas.Particularly, in ED context, in addition to discover frequent diseases, relations between different types of diseases and diagnostic tests can also be highlighted by applying ARM to make rapid decisions and plan operations more efficiently.

Association Rules Mining for Hospital Readmission: A Case Study

› Stud Health Technol Inform. 2001;84(Pt 2):1399-403.

Mining association rules from clinical databases: an intelligent diagnostic process in healthcare

S Stilou [1], P D Bamidis, N Maglaveras, C Pappas

Affiliations + expand

PMID: 11604957

Abstract

Data mining is the process of discovering interesting knowledge, such as patterns, associations, changes, anomalies and significant structures, from large amounts of data stored in databases, data warehouses, or other information repositories. Mining Associations is one of the techniques involved in the process mentioned above and used in this paper. Association is the discovery of association relationships or correlations among a set of items. The algorithm that was implemented is a basic algorithm for mining association rules, known as a priori. In Healthcare, association rules are considered to be quite useful as they offer the possibility to conduct intelligent diagnosis and extract invaluable information and build important knowledge bases quickly and automatically. The problem of identifying new, unexpected and interesting patterns in medical databases in general, and diabetic data repositories in specific, is considered in this paper. We have applied the a priori algorithm to a database containing records of diabetic patients and attempted to extract association rules from the stored real parameters. The results indicate that the methodology followed may be of good value to the diagnostic procedure, especially when large data volumes are involved. The followed process and the implemented system offer an efficient and effective tool in the management of diabetes. Their clinical relevance and utility await the results of prospective clinical studies currently under investigation.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC61314/

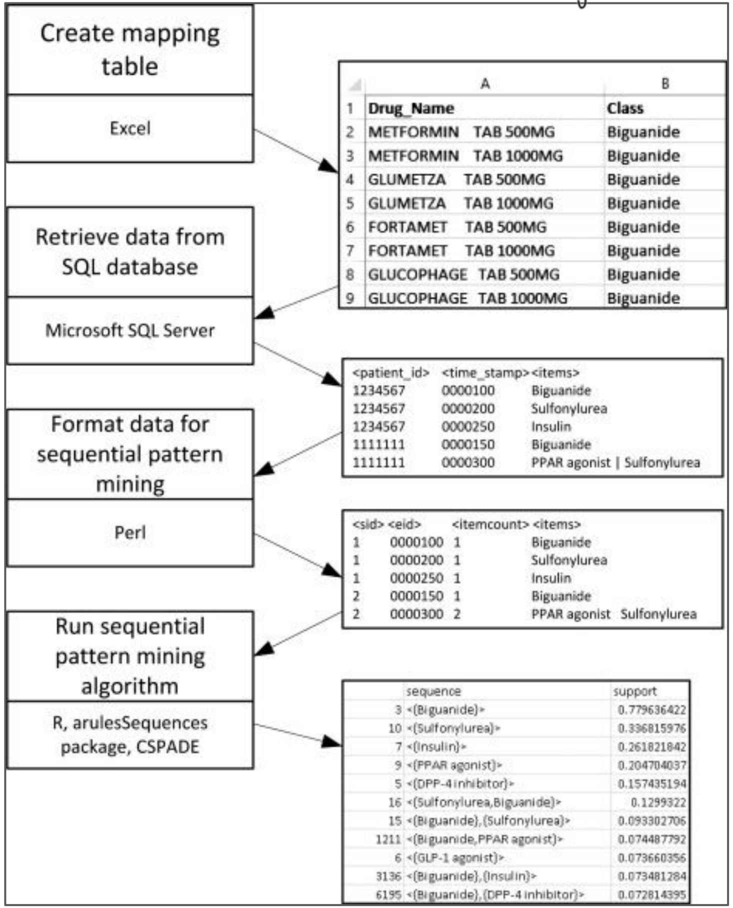
Sequential Pattern Discovery

Trad. ass. rule mining

mining salient

Traditional association rule mining was designed for **mining salient patterns within a set of transactions, ignoring the temporal order of the transactions.** Thus, association rule mining becomes inefficient for temporal and sequential applications that require ordered matching rather than simple subset testing

The use of sequential pattern mining to predict next prescribed medications.



Retrieving from SQL DB

Sequential Pattern Discovery

Basic of Sequential Pattern (GSP) Mining:

- **Sequence:** A sequence is formally defined as the ordered set of items {s1, s2, s3, …, sn}. As the name suggests, it is the sequence of items occurring together. It can be considered as a transaction or purchased items together in a basket.
- **Subsequence:** The subset of the sequence is called a subsequence. Suppose {a, b, g, q, y, e, c} is a sequence. The subsequence of this can be {a, b, c} or {y, e}. Observe that the subsequence is not necessarily consecutive items of the sequence. From the sequences of databases, subsequences are found from which the generalized sequence patterns are found at the end.
- **Sequence pattern:** A sub-sequence is called a pattern when it is found in multiple sequences. The goal of the GSP algorithm is to mine the sequence patterns from the large database. The database consists of the sequences. When a subsequence has a frequency equal to more than the "support" value. For example: the pattern <a, b> is a sequence pattern mined from sequences {b, x, c, a}, {a, b, q}, and {a, u, b}.

Medical data analysis: GSP mining can be used to **analyze medical data, such as patient records, and identify patterns in the data that are indicative of certain health conditions or trends.**

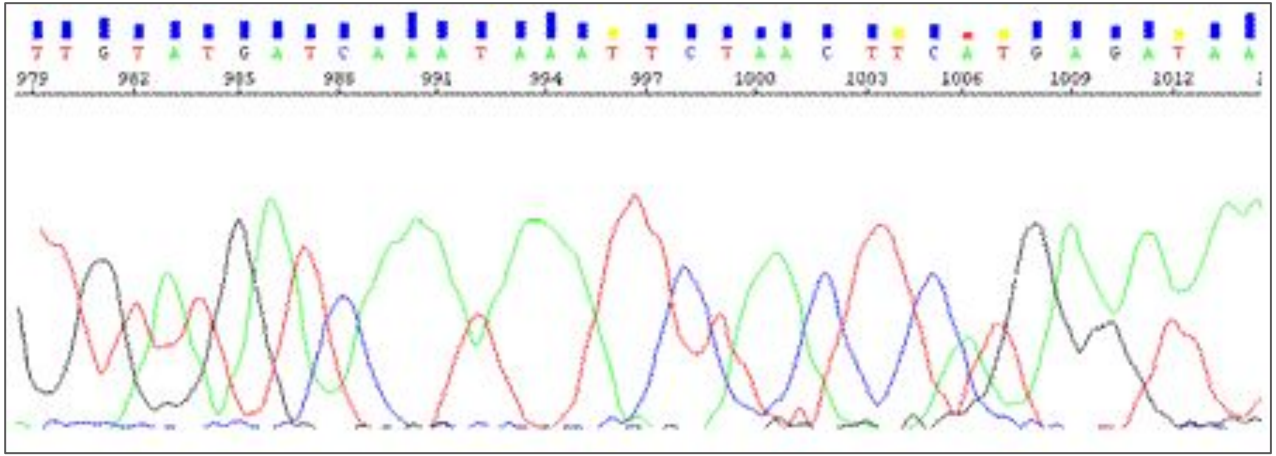
Sequential Pattern Discovery

Temporal ordered sets of items/events
Accounts for temporal order.
App: (1) next pres. med.
(2) patient records.
(3) DNA sequence analysis.
(4) motifs & tandem repeats in DNA
(5) infectious disease
(6) disease evolution

In DNA sequence analysis, approximate patterns become helpful because DNA sequences can include **(symbol) insertions, deletions, and mutations.** Such diverse requirements can be considered as constraint relaxation or application.

algo. → GSP
SPADE
Prefix Span.

[SPADE](), [PREFIX SPAN]()



Sequential Pattern Discovery

A sequential pattern as a temporally ordered set of items or events.

As with association rule mining(ARM), **combinatorial explosion and the number of data passes affect speed and memory requirements** of these algorithms.

Challenges are **pattern summarization** (i.e., providing an expressive, compact result set), the ability to naturally **handle and analyze temporal data** (e.g., activity sequences), and achieving **acceptable performance for long sequences and reduced support (low frequency)**

Potential medical applications of sequential pattern discovery include, but are not limited to, the following: • **DNA sequence alignment**

- **Motifs and tandem repeats in DNA sequences**
- **Modeling clinical visit patterns**
- **Analyzing infectious disease and micro-level disease observations**
- **Studying the evolution of a disease and agent/host interaction patterns**
- **Analysis of multi-omics experiment data involving temporal treatments**

Predicts the probability of an event by fitting the logistic regression curve to the data

Machine Learning

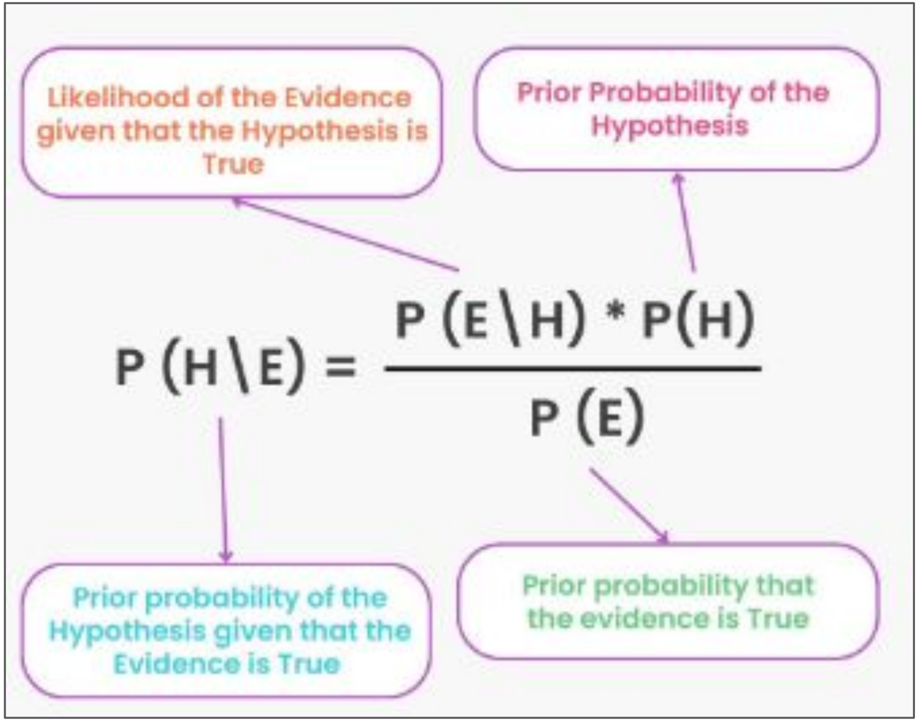
Logistic regression is a statistical model that predicts the probability of an event occurring by fitting a logistic curve to the data. Several predictor variables are typically used to describe the relationship between risk factors and outcomes.

Bayesian networks and the **naïve Bayesian classifier** is known for their simplicity and clear probabilistic semantics Know

Naïve Bayes prediction typically takes place using the following process, where *C* represents a random variable indicating an instance's class, *A* represents a vector of random variables indicating the attribute values that are observed, *c* represents the label of a class, and *a* represents a vector of observed attribute values: 1. Use Bayes' rule to compute the probability of each class given *a* 2. Predict the most probable class

$$p(C=c \mid A=a) = \frac{p(C=c)\prod p(A_i = a_i \mid C=c)}{Z}$$

Machine Learning



[Health Prediction and Medical Diagnosis using Naive Bayes]

Machine Learning

k-Nearest Neighbor and K*: These methods are categorically termed **instance-based learning** or example-based learning and generally use some form of similarity measure to determine "distance" or "match" between instances.

K* utilizes entropy (motivated by information theory) as a distance measure. This allows for a more **consistent approach to handle missing values and symbolic and real attribute values.** The K* distance method's general approach is similar to work done in comparing DNA sequences, where the distance between two instances is computed as a measure of the sum of all possible transforming paths (represented by transformation probabilities). The selected number of instances is termed the "sphere of influence," which dictates how the instances are weighed. During classification, the category corresponding to the highest probability is selected.

https://www.naturalspublishing.com/download.asp?ArtcID=21872

Machine Learning

Decision Tree (Pruned and Unpruned)

Overfitting is reduced by estimating the error rate of each subtree and replacing it with a leaf node if its estimated error is smaller, a process termed "pruning". Finally, the decision tree can be transformed into a set of rules by traversing the tree paths from the leaf nodes.

The resulting set of rules can then be simplified in a variety of ways (and some rules completely removed) to arrive at a final set of rules for classification.

Random forests have their roots in work involving **feature subset selection, random split selection, and ensemble classifiers.**

A random forest is constructed by selecting a random feature vector that is independent of the previous chosen feature vectors (but identically distributed) and growing a tree-structured classifier using this vector and the training set. This is done T times to create a random forest of T decision trees, each of which independently votes on a class. The most popular class from all trees is used to label the test instance

A Random Forest based predictor for medical data classification using feature ranking

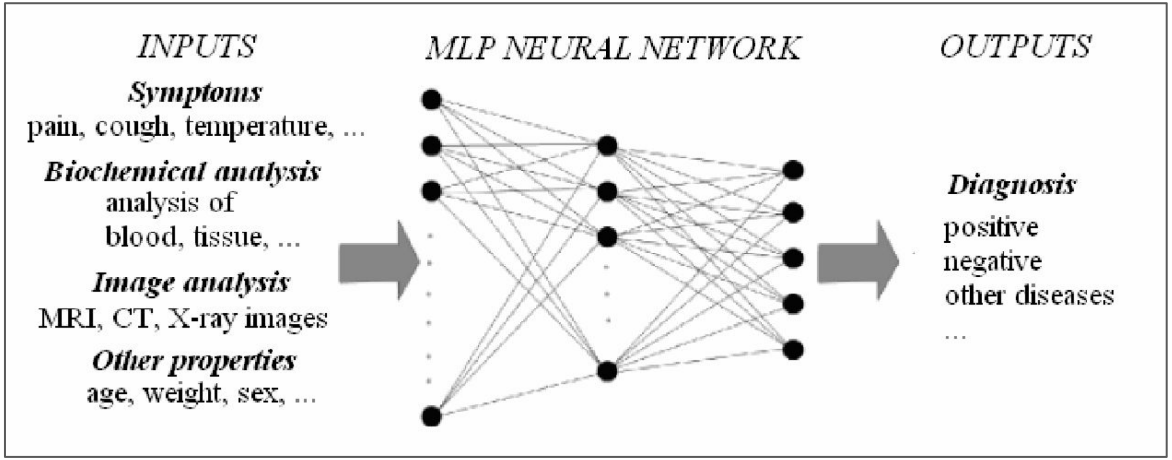
Machine Learning

Another rule-learning method is the **decision table classifier**, with a default rule mapping to the majority class.

A decision table has two main components: a schema (set of features) and a body (set of instances with feature values and class labels—the training set).

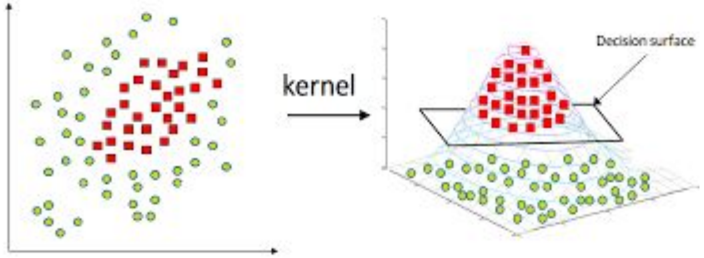
An optimal feature set is determined by transforming the problem into a state space (feature subset) search, using best-first search with the heuristic being k-fold cross-validation to estimate the future prediction accuracy. Incremental cross-validation can be used to provide a speedup for algorithms that support incremental addition and deletion of rules

Artificial Neural Network:ANN



Machine Learning:

Support vector machines are kernel-based learning machines that make no assumptions on the data distribution.
Support vector machines are backed by strong mathematical rigor, are robust to noise, and are currently popular in the literature.
They are designed for binary (two-class) problems but can be extended for multiclass classification data sets.
The support vector Machine’s goal is to find separators in the form of a hyperplane.
Various kernels have been developed for this method, one of which is the popular polynomial kernel.
[SVM application]



Machine Learning:Multi-Classifier Decision Fusion:

classifier fusion, mixture of experts, committees, ensembles, teams, collective recognition, composite systems

Each machine learning paradigm has its **own strengths and weaknesses stemming from its underlying theory, mathematical underpinning, and assumptions about the data and/ or decision space.**

Therefore, the accuracy of each algorithm is dependent upon how well those assumptions apply to the input data. Different algorithms achieve **different levels of accuracy on the same data,** but specific classifiers can develop levels of expertise on **portions of the decision space.** These aspects provide support for investigating decision fusion and meta-learning approaches.

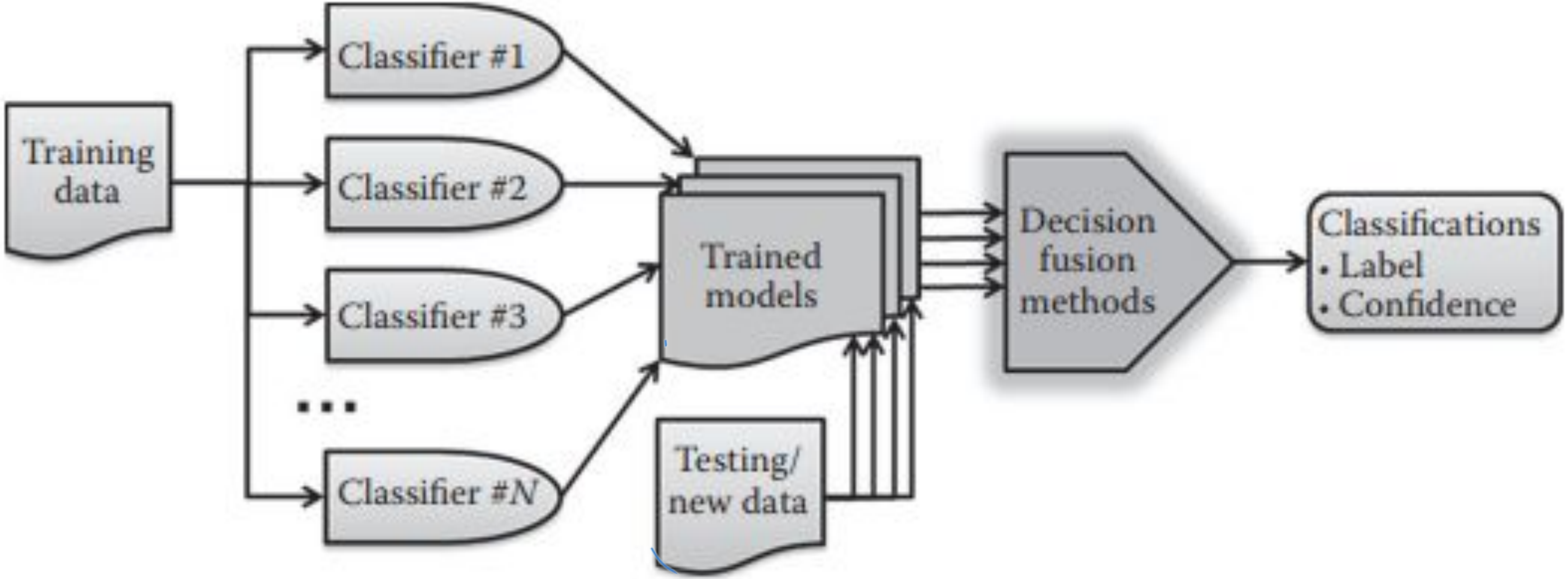
Since **different learning methods typically converge to different solutions, combining decisions from different learning paradigms can be leveraged for improved accuracy**

When developing a multi-classifier system, its members can be a **mixture of weak (i.e., high-error-rate) and strong (i.e., low-error-rate) classifiers.**

Weak classifiers are typically simple to create, at the expense of their accuracy on complex data sets. **Strong classifiers** are typically time consuming and expensive to create, as their parameters are fine-tuned and tweaked for optimal performance.

Combining weak/strong or homogeneous/heterogeneous classifiers offers the benefit of encompassing different levels of expertise and knowledge bases

Machine Learning:Multi-Classifier Decision Fusion:



Machine Learning:Multi-Classifier Decision Fusion:

classifier fusion, mixture of experts, committees, ensembles, teams, collective recognition, composite systems

If models are uncorrelated, they likely will misclassify different instances, and combining them better enables the system to correctly classify more instances. **A significant improvement over a single classifier can only happen if the individual classifier theories are substantially different.** It is desired to obtain a balance between high performance and complementarity in a team where decisions are combined. **If one learner does not predict correctly, the other learners should be able to do so. Diverse models** are therefore more likely to make errors in different ways. Methods to accomplish this typically involve introducing diversity in terms of **learning paradigms, feature subsets, or training sets. As modern data volumes grow in size, attempts at distributed processing and parallel machine learning are gaining popularity**

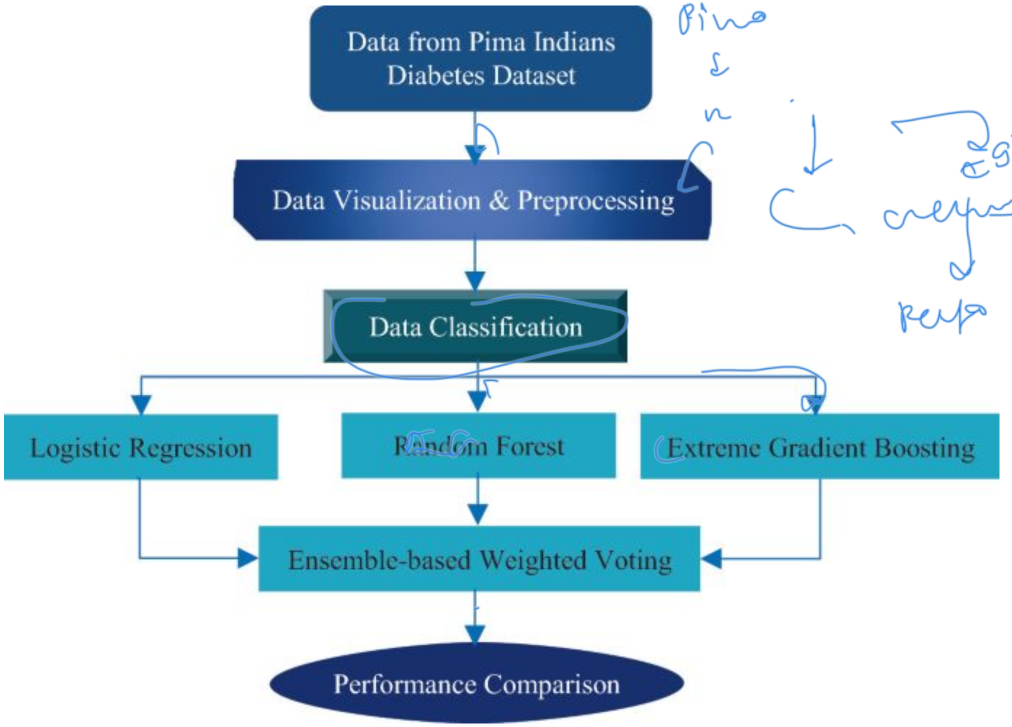
Pure voting methods are among the simplest ways to combine decisions from multiple classifiers. Examples of pure voting methods are majority vote, maximum confidence, average confidence, and product of confidences.

Accuracy-driven voting methods are those that take into account the accuracy, confidence, or probability of each learner offering its vote. This moves a step beyond simple voting, where votes from classifiers that have exhibited strong classification accuracies are weighted more, and weak classifier votes are either discarded or weakly weighted.

Machine Learning:Multi-Classifier Decision Fusion:

classifier fusion, mixture of experts, committees, ensembles, teams, collective recognition, composite systems

A popular accuracy-driven voting method is weighted voting, where each classifier is assigned a voting weight (usually proportional to its operational accuracy), and the class with the highest weighted vote is used for classification.



Machine Learning: Multi-Classifier Decision Fusion:

Data manipulation methods, which manipulate the training and testing data to attempt to achieve optimal classification accuracy, are by far the most popular means to train and combine multiple classifiers

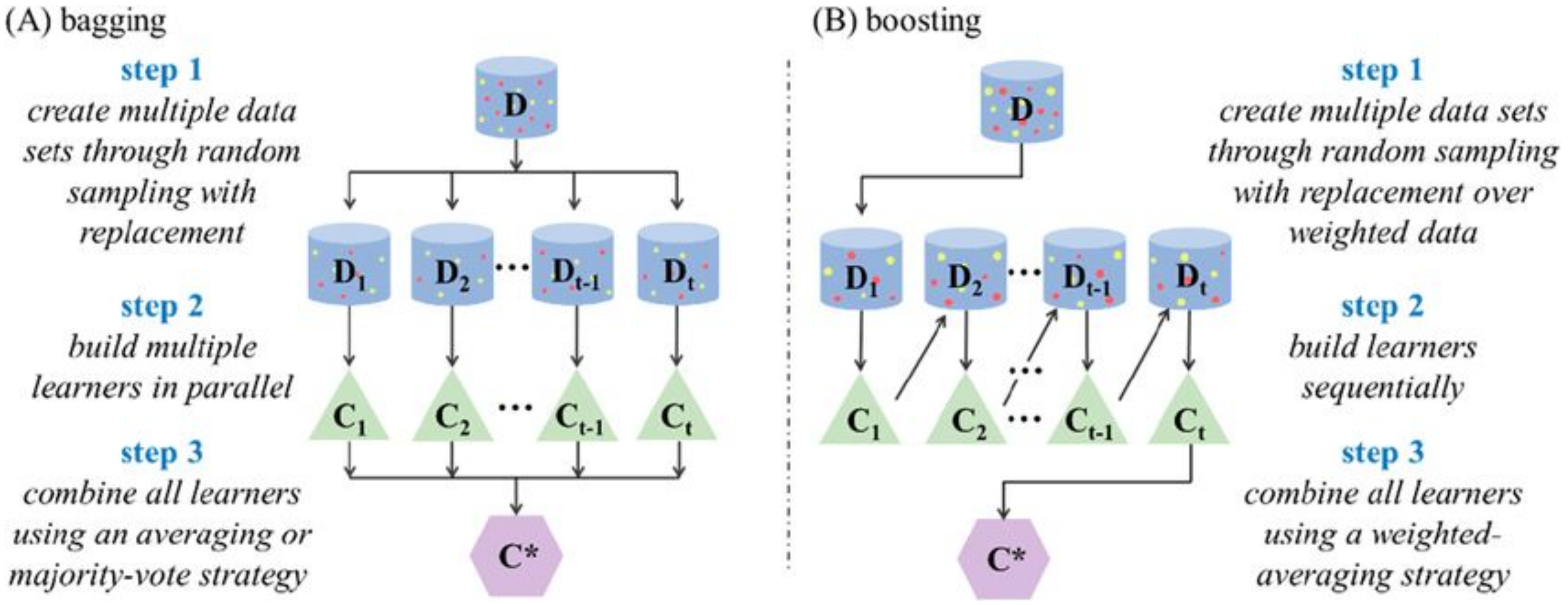
- **Input decimation:** Classifier correlation can be reduced by purposefully **withholding some parts of each pattern** (i.e., only using a subset of the features for certain classifiers). Feature inputs can be “pruned” by measuring how each affects the classifier output. Those features that have the least effect on the output can be selected for removal without compromising overall classifier performance. One example is to have one classifier per class, where each classifier only uses the features with high correlation to that class.

Machine Learning: Multi-Classifier Decision Fusion:

- **Boosting** adaptively changes the training set distribution based on performance of previous classifiers, attempting to reduce both bias and variance. Based on results of previous classifiers, diverse training samples are collected such that instances that were incorrectly predicted play a more important role in training (i.e., further learning focuses on difficult examples). This method relies on multiple learning iterations. At each iteration, instances incorrectly classified are given greater weight in the next iteration. Thus, the classifier in each iteration is forced to concentrate on instances it was unable to correctly classify in previous iterations. All classifiers are then combined after all iterations have been processed, or a threshold is met.
- **Bagging** creates a family of classifiers by training on stochastically different portions of the training set. N training “bags” are initially created, each obtained by taking a training set of size S and sampling the training set S times with replacement. Some instances could occur multiple times, while others may not appear at all. Each bag is then used to train a classifier, and classifiers are then combined using an equal weight for each.

S times. bags.

Machine Learning: Multi-Classifier Decision Fusion:



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8574072/

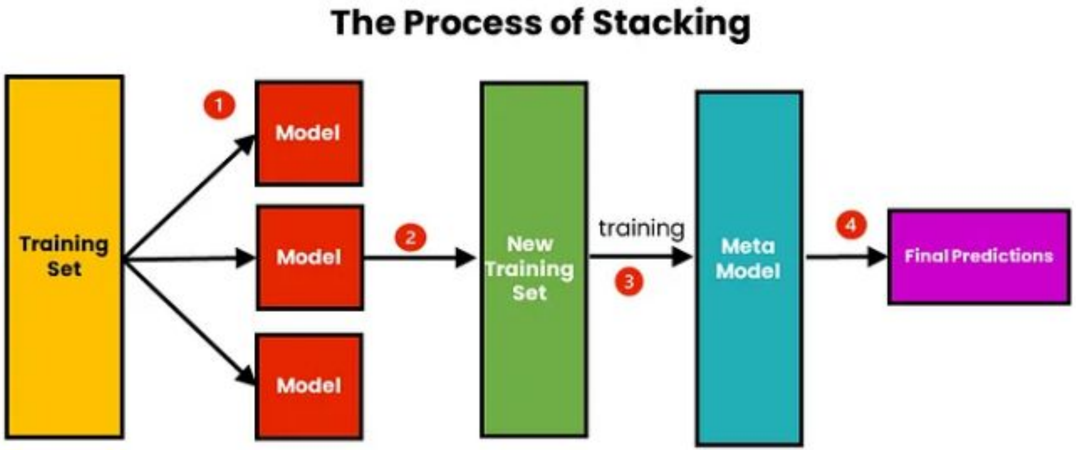
Ensemble Learning, Meta-Learning, and Other Abstract Methods

One of the most active areas in supervised learning is constructing successful ensembles of classifiers.

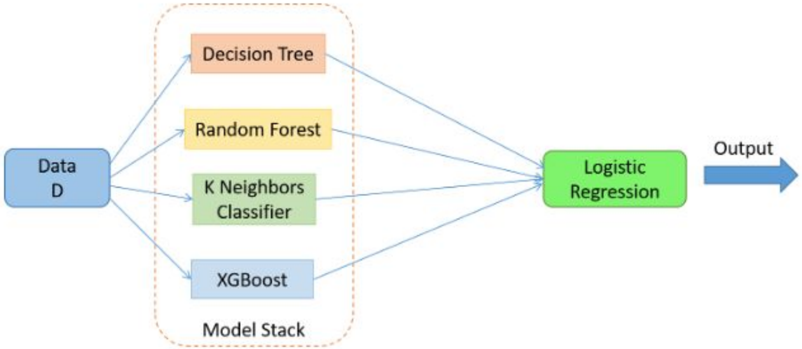
Ensemble methods construct a set of classifiers and then classify new data points by taking a (weighted) vote of their predictions . Ensembles are well established as a method for obtaining highly accurate classifiers by combining less accurate ones. They can often outperform any single classifier. However, to be more accurate than a single classifier, the ensemble must be composed of both accurate and diverse classifiers

Similarly, meta-learning uses a machine learning algorithm to model the decision patterns of a set of classifiers to construct a model that could yield increased accuracy. Figure 3.3 illustrates the meta-learning process for a multi-classifier system, where decisions are collected from a set of classifiers, used to create a model of the decision pattern of those classifiers, and that model is tested to arrive at a single classification for each instance. **The goal of the meta-classifier is to learn the mapping of classifier decisions that provides the most correct labels based on the training data. The meta-classifier will then label input instances based on the decisions from the underlying individual classifiers. This process is also known as "stacking."**

Stacking



Stacking for Machine Learning



Stacking, also known as "***Stacked Generalization,***" is a machine learning ensemble strategy that integrates many models to improve the model's overall performance. The primary idea of stacking is to feed the predictions of numerous base models into a higher-level model known as the meta-model or blender, which then combines them to get the final forecast.

https://colab.research.google.com/drive/1AVerQCYaK9NPLD5MztTzJqRtwOHxe8jG#scrollTo=kr3oUGJ1MKDh

Evolutionary Algorithm

Evolutionary algorithms are a class of stochastic search and optimization methods that mimic natural biological evolution. The common underlying premise is that given a population of individuals, the environmental pressure causes natural selection, which further causes a rise in the quality of the population

Given a quality function to be maximized (i.e., **the fitness function)**, a random set of candidate solutions can be evolved that maximize or improve upon previous fitness.

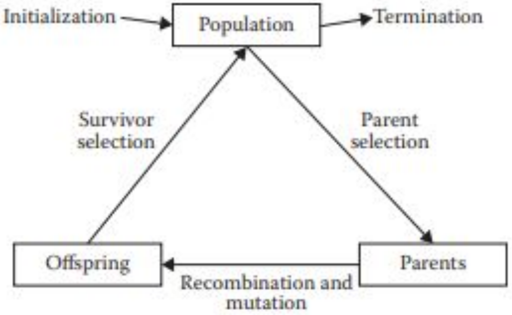
This fitness measure ensures that the better (more fit) candidates are selected to seed the next generation as baseline candidates, further evolved by applying reproduction operators such as recombination and mutation to them.

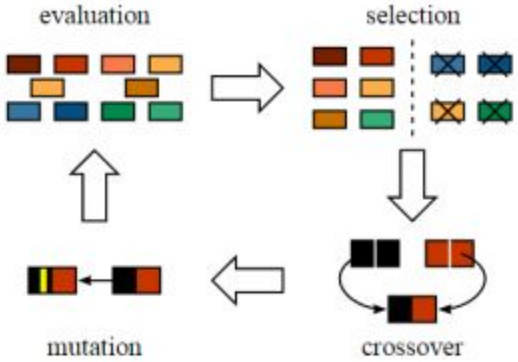
Recombination is an operator that applies to two or more selected candidates (the parents) and results in one or more new candidates (the children). **Mutation** is applied to one candidate and results in one new candidate compete—based on their fitness—with other candidates for inclusion in the next generation. This process, called a **generation,** is iterated until a termination condition is met. The termination condition can be the creation of a solution with sufficient quality, a time limit, or some computational limit

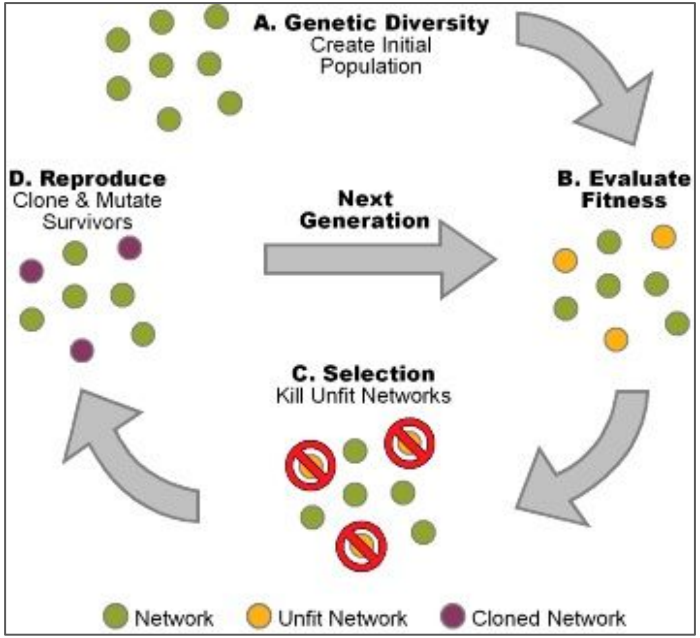
Evolutionary Algorithm

The general pseudocode for an evolutionary algorithm can be summarized by the following:

```
BEGIN
        INITIALIZE Population with random Candidate solutions
        EVALUATE each Candidate;
REPEAT WHILE (TERMINATION CONDITION is not satisfied)
        SELECT Parents;
        RECOMBINE pairs of Parents;
        MUTATE resulting Offspring;
        EVALUATE new Candidates;
        SELECT Individuals for the next Generation;
END
```







[The Applications of Genetic Algorithms in Medicine]

AI/ML in Healthcare – One Page Summary

1. Knowledge Discovery & Data Mining (KDD) + ARM

- ARM Purpose: Find frequent itemsets → Rules (X → Y).
- Apriori Algorithm: Uses Downward Closure Property (if superset is frequent → all subsets frequent).
- Lift: >1 = strong relation (non-random).
- Healthcare Apps: Frequent diseases, hidden diagnostic patterns, ED planning, readmission cases.

🔑 Mnemonic: “AALH” → Apriori, Association, Lift, Healthcare.

---

2. Sequential Pattern Discovery

- Definition: Ordered patterns over time.
- Algorithms: GSP, SPADE, PREFIX-SPAN.
- Challenges: Memory, temporal handling, long sequences.
- Healthcare Apps:
- Med prescription prediction,
- DNA sequences (mutations, insertions),
- Clinical visit modeling,
- Disease progression patterns.

🔑 Mnemonic: “GSP-DNA” → GSP, SPADE, PrefixSpan → DNA, Next Meds, Appointments.

---

3. Machine Learning Algorithms

- Logistic Regression (LR): Probabilities, risk prediction.
- Naïve Bayes (NB): Simple, probabilistic.
- kNN / K*: Instance-based, similarity (K* = entropy).
- Decision Tree (DT): Rule-based, prunable.
- Random Forest (RF): Ensemble of trees + feature ranking.
- Decision Table: Rule + default majority class.
- ANN: Neurons → hidden patterns.
- SVM: Hyperplanes, robust, binary → multiclass.

🔑 Mnemonic: “LNKDRAS” → LR, Naïve Bayes, kNN, Decision Tree, Random Forest, ANN, SVM.

---

4. Ensemble Learning & Decision Fusion

- Why: Combines models → ↑ Accuracy (needs Accuracy + Diversity).
- Methods:
- Voting (majority, weighted, product).
- Bagging (resampling).
- Boosting (focus on difficult cases).
- Input Decimation (reduce correlation).
- Stacking (meta-model).

🔑 Mnemonic: “VBBSI” → Voting, Bagging, Boosting, Stacking, Input decimation.

---

5. Evolutionary Algorithms (EA)

- Steps: Initialize → Select → Recombine/Mutate → Next Gen → Terminate.
- Apps: Genetic Algorithms for disease spread prediction, amblyopia screening.

🔑 Mnemonic: “ISRTT” → Init, Select, Recombine, Terminate.

Medical Context

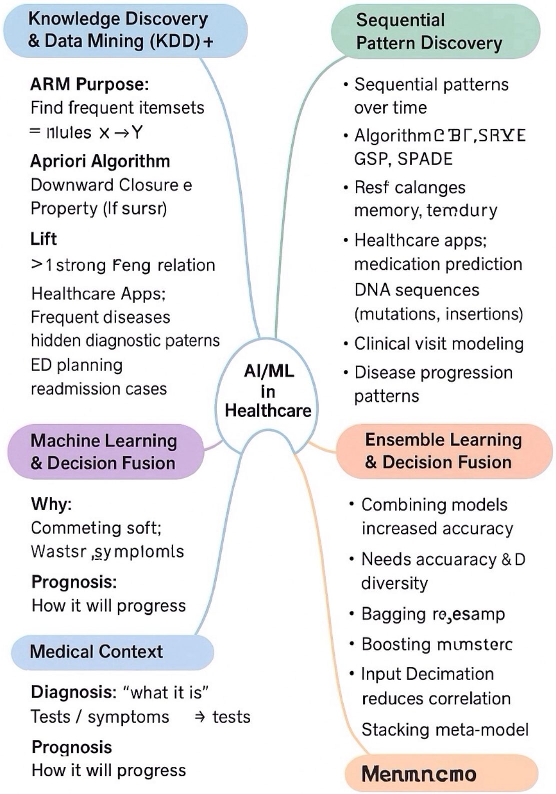
- Diagnosis: What it is → Tests, symptoms.
- Prognosis: How it will progress → Future outcomes, survival range

Master Mnemonic to Recall Entire PDF

👉 “A Smart Lovely Nurse Reads Daily Stories & Evolves Daily Patient Prognosis.”

- A → ARM (Apriori, Lift, Healthcare)
- Smart → Sequential Patterns (GSP, DNA)
- Lovely → Logistic Regression & ML algorithms (LNKDRAS)
- Nurse → Naïve Bayes, kNN, Decision Tree
- Reads → Random Forest & Rule-based classifiers
- Daily Stories → Decision Fusion & Stacking (VBBSI)
- Evolves → Evolutionary Algorithms (ISRTT)
- Daily Patient Prognosis → Diagnosis vs Prognosis

AI/ML in Healthcare – One Page Summary



A Smart Lovely Nurse Reads Daily Stories & Evolves Daily Patient Prognosis