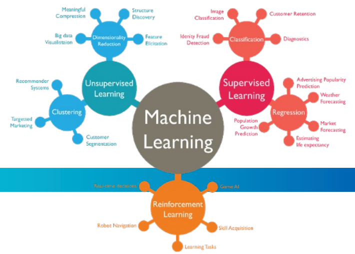
phuse·hybrid

Single Day Event

Anitha Cecelia, *Cognizant Technology Solutions*



Reshaping Clinical Trial Designing

using

Machine Learning (ML) Algorithms

Disclaimer

Content, Views and Opinions Expressed in this presentation are solely of the Author/s & Presenters and do not necessarily represent official policy or position or views of Cognizant.

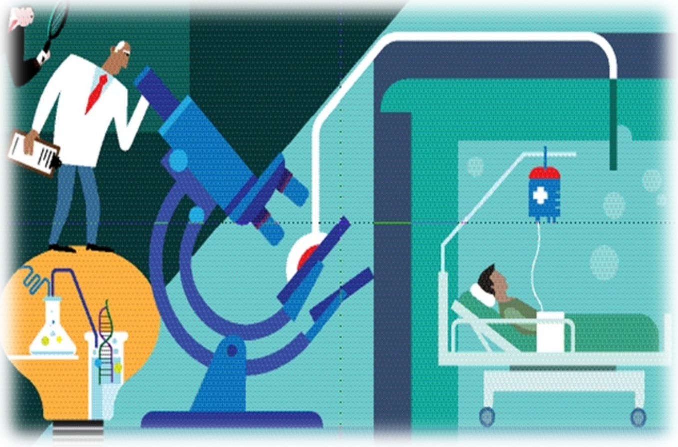
Agenda

- Introduction
- Basics of Machine Learning
- ML Algorithm in Rare Diseases
- ML Algorithm in Clinical Trial Designing
- Conclusion

Introduction

Clinical trials are research studies performed in people that are aimed at evaluating a medical, surgical, or behavioral intervention.

They are the primary way that researchers find out if a new treatment, like a new drug or diet or medical device is safe and effective in people or patients.



- One of the crucial element of success of a clinical trial is Designing and Execution.
- Another crucial challenge is the conduct of clinical trails for rare disease outcomes.

Introduction

Emerging machine learning technologies are beginning to transform medicine and healthcare and could also improve the diagnosis and treatment of rare diseases.

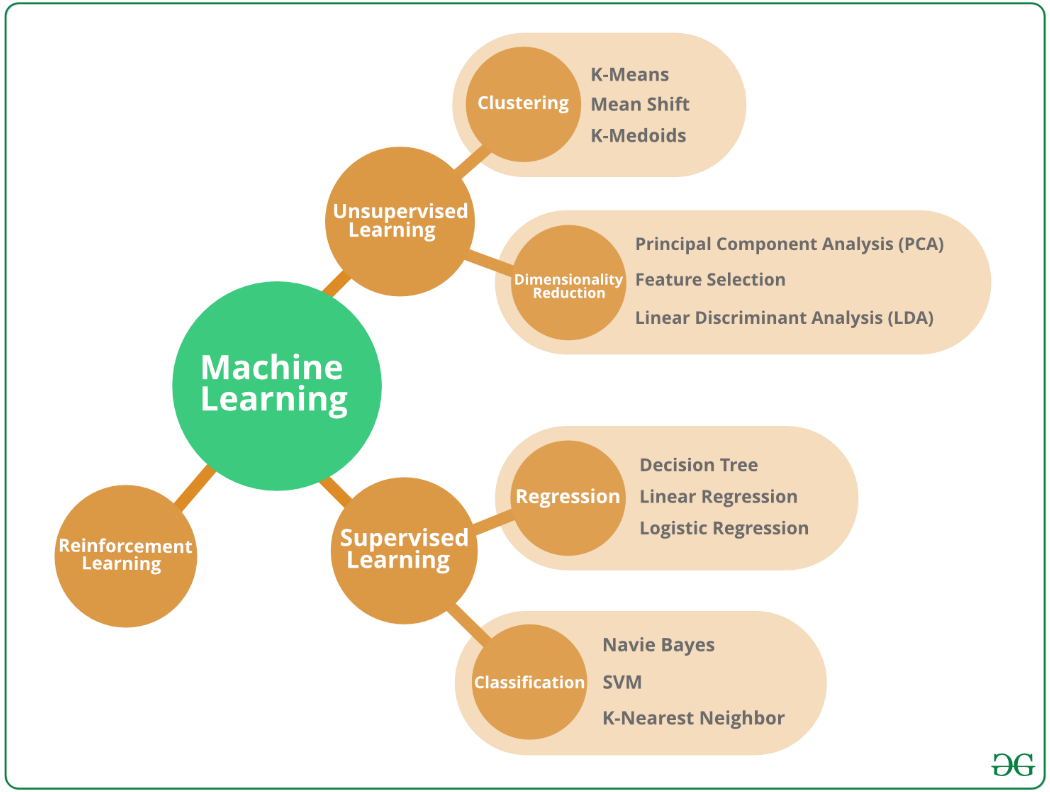
- Diseases that affect fewer than 5 patients per 10,000 are defined as rare diseases.
- For the past 10 years, studies have been conducted across the globe in application of machine learning in a rare disease context.
- Ensemble methods; support vector machines and artificial neural networks were the Machine Learning algorithms most applied across these studies.
- Advances in machine learning can significantly improve diagnosis, treatment and prognosis of rare disease patients.

What is ML?

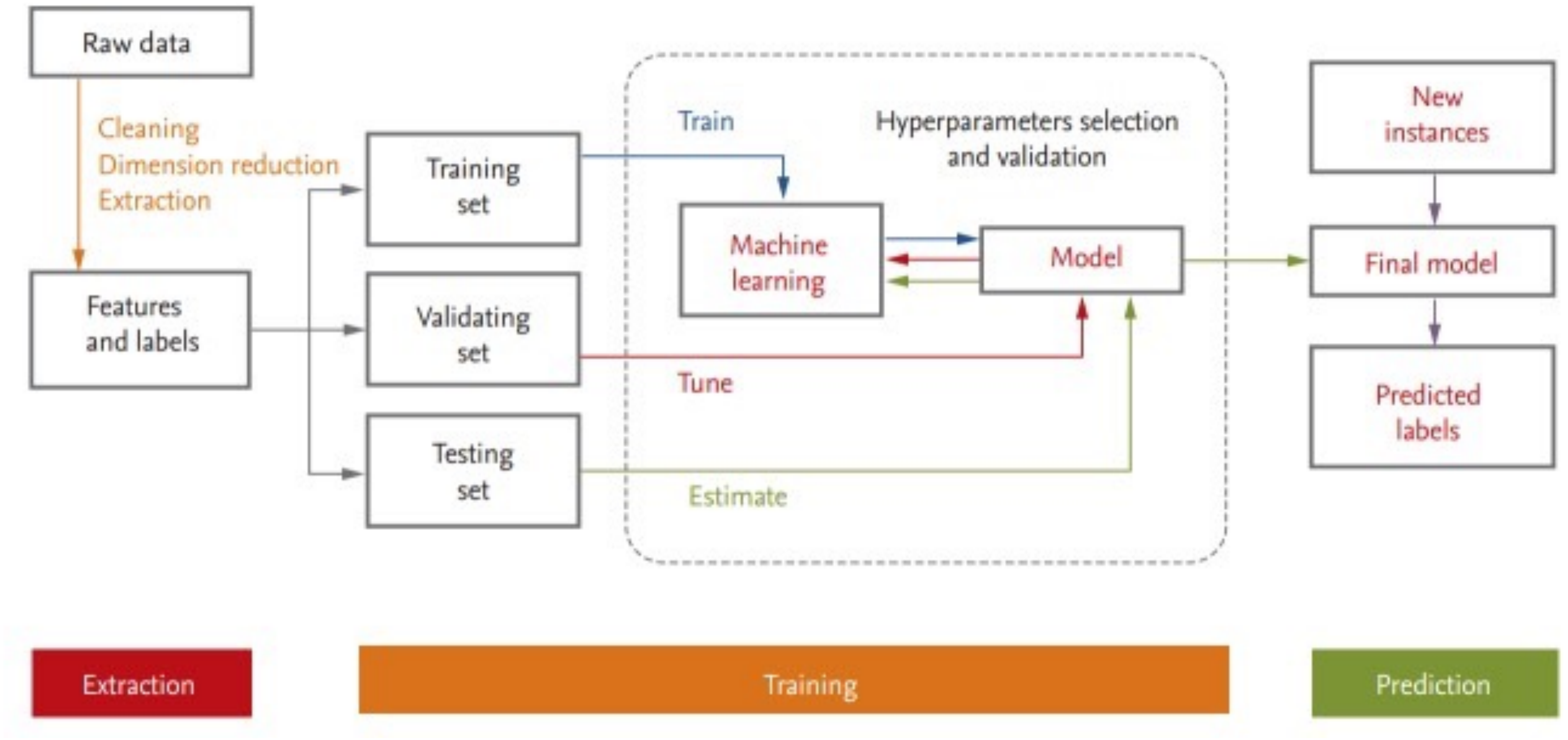
Machine Learning (ML) is a mathematical model where a computer/machine learns from the past experiences (input data) and makes future predictions.

A machine learning algorithm is the method by which the computer system conducts its task by programmatically predicting output values from given input data.

ML Algorithms



Workflow of Machine Learning



Rare Diseases

Table 1 Data extracted from the studies

Variable	Categories	Definition	Example(s)
Rare disease	All rare diseases described at least once in the studies (studies investigating more than one rare disease were categorized as "Diverse")	Orphanet disorder name	Cystic fibrosis, Sickle cell anemia, Gaucher disease
Disease group	All disease groups of the 381 specific diseases included in the search as well as disease groups of other diseases identified in the studies	Orphanet disease group as defined by the preferential parent in the classification hierarchy	Rare neurologic disease, Rare respiratory disease, Rare endocrine disease
Publication year	Years from 2010 to 2019	Year of the publication date of the article	
Country of study	All countries that published at least one article	Country of institution of senior (i.e. last) author of the study	
Medical application	Diagnosis	Studies aiming to correctly diagnose patients	Classification of cases and controls or different disease subtypes, Identification of biomarkers, Deep phenotyping, Decision support
	Treatment	Studies aiming to improve treatment or develop new therapies	Detection of therapeutic targets, Identification of binding proteins

Ophthalmology

Macular telangiectasia (Mac Tel), Type 1 is a rare eye disorder predominantly in male.

- Ellipsoid zone (EZ) defect areas were measured on spectral domain OCT images of each eye
- The fully automatic segmentation algorithm was used to assess EZ defect areas and was able to reliably reproduce the statistically significant primary outcome measure of the clinical trial.

Retinal Lesion Detection With Deep Learning Using Image Patches,

- Retinal lesions are tumors that form in the eyes.
- Regionally trained convolutional neural networks can generate lesion-specific probability maps able to detect and distinguish between subtle pathologic lesions.

Alzheimer's Disease

Alzheimer's disease (AD) is an adult-onset cognitive disorder (AOCD) which represents the sixth leading cause of mortality and the third most common disease after cardiovascular diseases

- Alzheimer's Disease Neuroimaging Initiative (ADNI) was used to train ML classifiers that can differentiate between individuals who had declining cognitive function and individuals with stable cognitive function.
- Model-based Reinforcement learning enhances the existing knowledge of the neural substrates. Ensemble leanings, helps to improved classification and prediction performance in cognitive outcomes.

Alzheimer’s Disease

Alzheimer’s disease (AD) is an adult-onset cognitive disorder (AOCD) which represents the sixth leading cause of mortality and the third most common disease after cardiovascular diseases

- Three classifiers (Random Forest (RF), Naive Bayes (NB), and Gradient boosting) are simple and efficient method of dementia subject identification.
- Linear discriminant analysis (LDA) and SVM achieved high AD detection accuracy and Kernel density estimations helped extract texture information from the MR images.
- Shallow and deep neural network architectures helped assess caregiver burden

Rheumatic Disorders

Rheumatic diseases affect your joints tendons, ligaments, bones, and muscles. Among them are many types of arthritis, a term used for conditions that affect your joints.

- Ensemble Method was used for Classification of Rheumatoid Arthritis from Electronic Image Sensor.
- Regularization method found in Least Absolute Shrinkage and Selection operator (LASSO), Elastic Net, and Ridge regression was used to perform variable selection which transform the counts of tender and/or swollen joints and the erythrocyte sedimentation rate (ESR) into a single formulated feature, Disease Activity Score (DAS28)-ESR, which better assesses disease activity in patients with RA.
- Random survival forest helped in identifying important risk factors for survival in patient with systolic heart failure.
- Phenomapping for novel classification of heart failure with preserved ejection fraction.

Some more examples

- Black-box machine learning algorithms give us a sense of the predictive capabilities of our scientific theories.
- Hidden Markov models have been well validated as an effective temporal pattern recognition tool.
- Risk-based monitoring (RBM) systems started utilizing machine-learning algorithms to quickly analyze data in real time and provide insights for decision makers.
- Support Vector Machine Algorithm (SVM) was used to classify patients with CAD from healthy individuals. The model identified regional wall motion abnormality and ejection fraction (echocardiographic features), age, and hypertension as risk factors.
- Convolutional neural networks (CNN) were shown to classify images of skin cancer on par with dermatologists and could enable lifesaving and fast diagnoses, even outside the hospital via installation of apps on mobile devices.

Clinical Trial Designing

Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemp Clin Trials Commun. 2018 Aug 7;11:156-164.

This paper documents a table with the list of factors associated with problems or challenges when preparing for or executing a clinical trial, along with the opportunities for artificial intelligence to help alleviate these issues.

Table 1

A list of factors associated with problems or challenges when preparing for or executing a clinical trial, along with the opportunities for artificial intelligence to help alleviate these issues. Abbreviation: NLP = natural language processing.

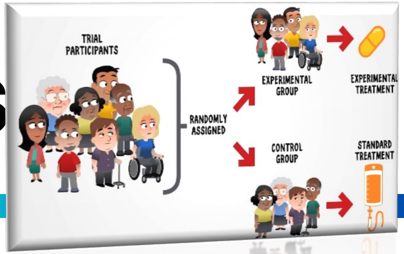
Factor	Opportunity	Role for Artificial Intelligence
Poor study design	More complete literature review	NLP of available literature, finding similar trials, trials addressing similar issues, or trials addressing different issues utilizing similar techniques, summarized for the study designer
	Appropriate endpoints	NLP of available literature, showing endpoints/measures used in other similar studies

Use of ML in Clinical Trial designing

Factor	Opportunity
Poor study design	More complete literature review
	Appropriate endpoints
	Inappropriate eligibility criteria
	Appropriate statistical analysis
	Determination of appropriate sample size
	Reducing likelihood of amendments
	Inconsistencies in protocol

Factor	Opportunity
Ineffective site selection	Effective measurement of trade-offs for each site
Poor recruitment	Improved use of funds
	Ensuring appropriate eligibility criteria
	Facilitating locating eligible patients
	Enrolling patients who are likely to complete the trial
Patient burden/safety	Minimize travel and wait times

Identification of Subjects



Once the specific cohort has been selected, natural language processing (NLP) has shown promise in identification of patients matching the desired phenotype, which is otherwise a labor-intensive process.

- **Phenotyping algorithms** applied to data enable investigators to identify large cohorts for clinical and genomic research. They helps to identify individuals (e.g., hypertensive) and controls and provide a portable module for implementation at other sites.

- A **cross-modal inference learning model algorithm** jointly encodes enrollment criteria (text) and patient records (tabular data) into a shared latent space, matching patients to trials using Electronic Health Record (EHR) data in a significantly more efficient manner than other machine learning approaches.

Data Stratification

Empirical Risk Minimization **(ERM) algorithm** is that which defines a family of learning algorithms and is used to give theoretical bounds on the performance of learning algorithms.

- It is a natural choice for a learning algorithm that helps to determine a good classification and regression learning function from a bad one.

Ground Truth(GT) supports single and multi-class semantic segmentation labeling jobs.

- Ground truth selection for AI-based training along with scientific and clinical validation is very important for CVD stratification to avoid risk of bias.

Imputation of missing data

A novel deep learning architecture, Multi-directional Recurrent Neural Network interpolates within data streams and imputes across data streams.

- The power of this approach was demonstrated using five real-world medical datasets. Improved estimation of missing measurements were compared with state-of-the-art benchmarks (including Spline and Cubic Interpolations, MICE, MissForest, matrix completion, and several RNN methods).

- This method includes the temporal aspects of the data hence reducing the number of assumptions made on missingness of data.

Dose Allocation

Safe Efficacy Exploration Dose Allocation (SEEDA) is an algorithm that aims at maximizing the cumulative efficacies while satisfying the toxicity safety constraint with high probability.

- An extended SEEDA-Plateau algorithm that is tailored for the increase-then-plateau efficacy behavior of molecularly targeted agents (MTA).

- Using both synthetic and real-world datasets, SEEDA outperforms state-of-the-art clinical trial designs by finding the optimal dose with higher success rate and fewer patients.

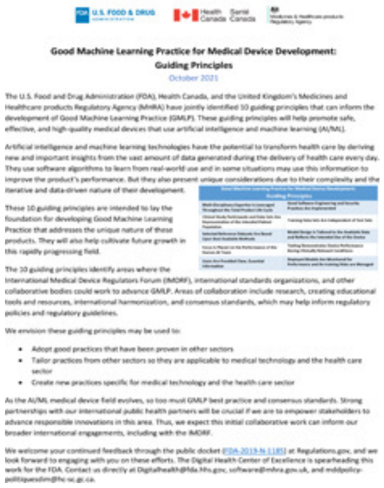
Conclusion

- Current machine learning algorithms are used for the diagnosis and prognosis of rare diseases, which is a good step forward.
- ML can also play an important role in improving the treatment of rare diseases, and future studies could focus more on this aspect, for example by using machine learning to accelerate drug development.
- ML allows users to process large quantities of data and make smart inferences and guide in the designing and execution of clinical trials.

Conclusion

- The insights from the available data can be used to automate parts of the system leading to a faster and a more efficient clinical trial system.
- Automation allows the ML predictions to be fed back into the system and specific actions to be taken, reducing the need for human intervention, and improving quality and speed of clinical trial execution.
- ML and automation can be applied across every stage of the trial process.
- Machine Learning or other Technology based analysis cab deliver critical insight to clinicians at the point of decision making and is not necessarily a replacement for the conventional clinical trial methodology.

Good Machine Learning Practice for Medical Device Development: Guiding Principles


Good Machine Learning Practice for Medical Device Development:
Guiding Principles
October 2021

The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP). These guiding principles will help promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning (AI/ML).

Artificial intelligence and machine learning technologies have the potential to transform health care by deriving new and important insights from the vast amount of data generated during the delivery of health care every day. They use software algorithms to learn from real-world use and in some situations may use this information to improve the product's performance. But they also present unique considerations due to their complexity and the iterative and data-driven nature of their development.

These 10 guiding principles are intended to lay the foundation for developing Good Machine Learning Practice that addresses the unique nature of these products. They will also help cultivate future growth in this rapidly progressing field.

The 10 guiding principles identify areas where the International Medical Device Regulators Forum (IMDRF), international standards organizations, and other collaborative bodies could work to advance GMLP. Areas of collaboration include research, creating educational tools and resources, international harmonization, and consensus standards, which may help inform regulatory policies and regulatory guidelines.

We envision these guiding principles may be used to:

- Adopt good practices that have been proven in other sectors
- Tailor practices from other sectors so they are applicable to medical technology and the health care sector
- Create new practices specific for medical technology and the health care sector

As the AI/ML medical device field evolves, so too must GMLP best practice and consensus standards. Strong partnerships with our international public health partners will be crucial if we are to empower stakeholders to advance responsible innovations in this area. Thus, we expect this initial collaborative work can inform our broader international engagements, including with the IMDRF.

We welcome your continued feedback through the public docket (FDA-2019-N-1185) at Regulations.gov, and we look forward to engaging with you on these efforts. The Digital Health Center of Excellence is spearheading this work for the FDA. Contact us directly at Digitalhealth@fda.hhs.gov, software@mhra.gov.uk, and mddpolicy-politiquesdim@hc-sc.gc.ca.

The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP).

These guiding principles will help promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning (AI/ML).

References

- Top 10 Algorithms every Machine Learning Engineer should know
- Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. 2017 Jan;24(1):162-171.
- Electronic health records-based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc. 2013 Dec;20(e2):e226-31.
- Good Machine Learning Practice for Medical Device Development: Guiding Principles | FDA; Last accessed on 13 Sep 2022.

References

- Machine learning for clinical trials in the era of COVID-19. Statistics in Biopharmaceutical Research 2020, Vol. 12, No. 4, 506–517
- Estimating Missing Data in Temporal Data Streams Using Multi-Directional Recurrent Neural Networks. IEEE Transactions on Biomedical Engineering, Vol. 66, no. 5, pp. 1477-1490, May 2019.
- A Logical Approach for Empirical Risk Minimization in Machine Learning for Data Stratification. Research Journal of Mathematics and Computer Science, 2017; 1:3
- Understanding Machine Learning: From Theory to Algorithms; 2014 by Shai Shalev-Shwartz and Shai Ben-David

References

- Improved Alzheimer's Disease Detection by MRI Using Multimodal Machine Learning Algorithms. Diagnostics (Basel). 2021 Nov; 11(11): 2103.

- The use of machine learning in rare diseases: a scoping review. Orphanet Journal of Rare Diseases (2020) 15:145

- Skin Cancer classification using convolutional neural networks: systematic review. J Med Internet Res. 2018;20.

- All Pictures in the presentation are sourced from Google Images.

Thank You

The Global Healthcare
Data Science Community

Contact Channels
UK +44 1843 609600
office@phuse.global
phuse.global

Social Media
@phusetwitta
/phusebook
/phusetube
/company/phuse

Additional Slides

Artificial Intelligence	Machine Learning	Deep Learning
AI stands for Artificial Intelligence, and is basically the study/process which enables machines to mimic human behaviour through particular algorithm.	ML stands for Machine Learning, and is the study that uses statistical methods enabling machines to improve with experience.	DL stands for Deep Learning, and is the study that makes use of Neural Networks(similar to neurons present in human brain) to imitate functionality just like a human brain.

Additional Slides

Artificial Intelligence	Machine Learning	Deep Learning
AI is the broader family consisting of ML and DL as it's components.	ML is the subset of AI.	DL is the subset of ML.
AI is a computer algorithm which exhibits intelligence through decision making.	ML is an AI algorithm which allows system to learn from data.	DL is a ML algorithm that uses deep(more than one layer) neural networks to analyze data and provide output accordingly.

Artificial Intelligence	Machine Learning	Deep Learning
Three broad categories/types Of AI are: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI)	Three broad categories/types Of ML are: Supervised Learning, Unsupervised Learning and Reinforcement Learning	DL can be considered as neural networks with a large number of parameters layers lying in one of the four fundamental network architectures: Unsupervised Pre-trained Networks, Convolutional Neural Networks, Recurrent Neural Networks and Recursive Neural Networks

Additional Slides

Artificial Intelligence	Machine Learning	Deep Learning
The efficiency Of AI is basically the efficiency provided by ML and DL respectively.	Less efficient than DL as it can't work for longer dimensions or higher amount of data.	More powerful than ML as it can easily work for larger sets of data.
Examples of AI applications include: Google's AI-Powered Predictions, Ridesharing Apps Like Uber and Lyft, Commercial Flights Use an AI Autopilot, etc.	Examples of ML applications include: Virtual Personal Assistants: Siri, Alexa, Google, etc., Email Spam and Malware Filtering.	Examples of DL applications include: Sentiment based news aggregation, Image analysis and caption generation, etc.