APPLICATION OF MACHINE LEARNING TECHNIQUES IN EARLY DISEASE DETECTION: A REVIEW

Goyal A[1], Chadha N[1], Riggs J[2]
[1]ZS Associates, Gurgaon, India, [2]ZS Associates, Thousand Oaks, CA, USA

Background and Objectives

- Machine learning (ML) is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
- One of the key use cases for ML techniques is helping healthcare professionals with early diagnosis of diseases based on patient medical history, lab results etc. Being able to diagnose a disease early on can lead to better treatment planning, better outcomes & overall improvement in patient care

Objectives

- Our first objective was to summarize the results from peer-reviewed literature on the use of ML techniques in early detection/diagnosis of diseases
- Our second objective was to understand key benefits ML techniques have to offer against a traditional approach in this case use & also to identify key challenges that need to be addressed for faster adoption.

Methodology

We searched PubMed for articles related to ML and early disease detection based on table below.

Search strategy	▪ Search strategy: (machine learning[MeSH] OR artificial intelligence[MeSH]) AND ("early disease detection" OR "early diagnosis" OR "early screening" OR "early warning")
Search limits	▪ Articles published in last 5 years in English language
Inclusion Criteria	▪ For inclusion, articles must be focused on use of ML techniques for early disease detection/diagnosis using clinical or real-world datasets

Articles unrelated to early detection/diagnosis, or that did not use ML methods in outcomes research, such as studies of general context, were excluded.

Current Use in Early Diagnosis

- Out of 267 studies returned via pubmed search, 152 met the inclusion criteria and were further analyzed

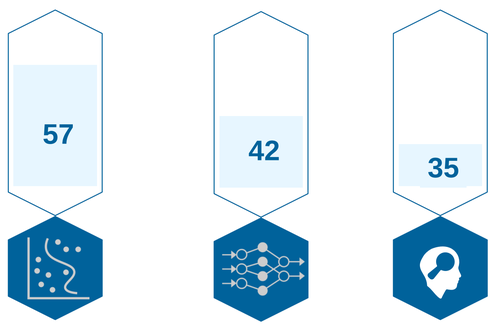
Publications by Study Type

Disease Area

Top disease areas with ML use in early diagnosis

- Neurological Disorders (n = 48)
- Cancer (n = 39)
- Cardiological Disorders (n = 16)
- Others - Diabetes, Kidney stone etc. (n = 49)

Algorithm Deployed



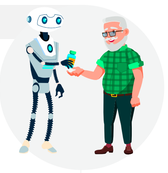
* Other methods include random forest, logistic regression, deep learning, naïve bayes and k nearest neighbors

These observations are in line with expectations for following two possible reasons:
- ✓ *Disease areas discussed above are amongst the leading cause of patient deaths worldwide [WHO Fact sheet – top 10 causes of death]*
- ✓ *Data generated as part of disease diagnosis (e.g. ECG / MRI scans) are well structured & can be analyzed easily through ML algorithms. Thus scope of improving early diagnosis through analysis is highest here*

Case Examples

Aguiar et al 2015

*A new methodology for the early detection of **epileptic seizure** based on the WiSARD weightless neural network model was developed using electroencephalogram data.*

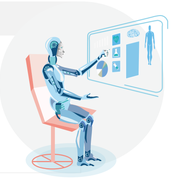


Farhan et al 2014

*An automated image processing based approach for the **identification of Alzheimer's disease** patients in early stage from MRI of the brain**.** An ensemble of classifiers was used to overcome the error caused by an independent base classifier.*

He et al 2018

*A proof of concept study was done utilizing artificial neural network models on functional connectome data to capture the individual variability inherent in the **developing brains** of pre-term infants. Objective was to enable early prediction of cognitive deficits in very preterm infants.*



Khaleghi et al 2015

*Early diagnosis of **bipolar disorder (BPD)** sub-types is quite challenging in adolescence. Multi-layer perceptron neural network was used to develop a classification model of adolescents with type I and type II of BPD using an electroencephalogram signal.*

Ghatwary et al 2019

*Study was done to enable **early esophageal adenocarcinoma** detection using deep learning object detection methods on high-definition white light endoscopy images. The object detection algorithms used include regional-based Convolutional Neural Network, Fast R-CNN, Faster R-CNN and Single-Shot Multibox Detector.*

Benefits & Challenges

Key Benefits

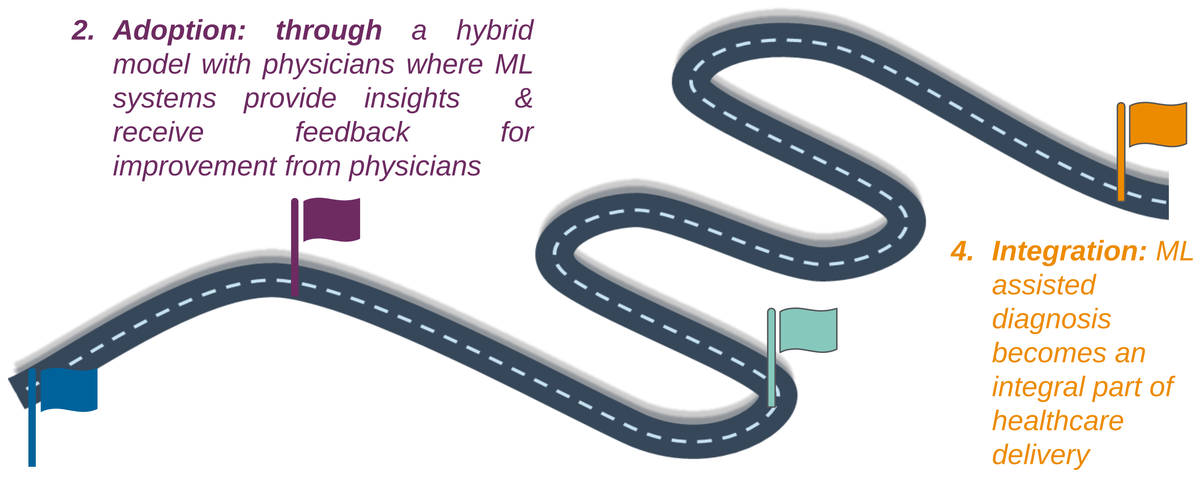
- There is an upwards trend in the publication of medical knowledge & healthcare data availability over time. ML techniques provide a great platform to integrate new developments into healthcare decision making for patients at a faster pace
- Often in comparison studies, ML algorithms have outperformed human counterparts in terms of accuracy [Karakülah et al, Esteva et al, Gulshan et al].
- Working with unstructured datasets (social media feed, physician notes etc.) is easier because of NLP techniques

Challenges

- One of the key dependencies for ML algorithms is training data availability. This limits use of these techniques for new or rare disease detection scenarios where data is scarce and / or of poor quality for a model to be trained upon.
- There is an active concern around patient data privacy & how data should be made available for studies in a more secure & anonymized format.
- There is limited regulatory guidance available on use of ML/AI based systems to influence patient care. This limitation prevents a lot of stakeholders from investment in further development of these ideas.

Machine Learning in Early Diagnosis (Possible Future Roadmap)

2. ***Adoption: through** a hybrid model with physicians where ML systems provide insights & receive feedback for improvement from physicians*



4. ***Integration:** ML assisted diagnosis becomes an integral part of healthcare delivery*

1. ***Current State:** Medical community is becoming aware about use of ML in early diagnosis & its benefits. We see a lot of proof of concept studies being conducted and compared against traditional methods*

3. ***Data Consolidation:** through payers, patients, hospitals, government & other stakeholders coming together to share data and creating a more integrated ecosystem for data capture & analysis. Technologies like blockchain can play a pivotal role here*

References

1. WHO Fact sheet: The top 10 causes of death 2018.
2. de Aguiar K, França FMG, Barbosa VC, Teixeira CAD (2015) Early detection of epilepsy seizures based on a weightless neural network. Conf Proc IEEE Eng Med Biol Soc. 4470-4.
3. Farhan, S., Fahiem, M. A., & Tauseef, H. (2014). An ensemble-of-classifiers based approach for early diagnosis of Alzheimer's disease: classification using structural features of brain images. Computational and mathematical methods in medicine, 862307.
4. He L, Li H, Holland SK, Yuan W, Altaye M, Parikh NA. (2018) Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework. Neuroimage Clin. 18:290–297.
5. Khaleghi A, Sheikhani A, Mohammadi MR, Nasrabadi AM, Vand SR, Zarafshan H, et al. (2015) EEG classification of adolescents with type I and type II of bipolar disorder. Australas Phys Eng Sci Med. 38:551–559.
6. Ghatwary N, Zolgharni M, Ye X. (2019) Early esophageal adenocarcinoma detection using deep learning methods. Int J Comput Assist Radiol Surg. 14(4):611–621.
7. Karakülah G , Dicle O , Koşaner O , et al . (2014) Computer based extraction of phenoptypic features of human congenital anomalies from the digital literature with natural language processing techniques. Stud Health Technol Inform 205:570-4.
8. Esteva A , Kuprel B , Novoa RA , et al . (2017) Dermatologist-level classification of skin Cancer with deep neural networks. Nature 542:115–8.
9. Gulshan V , Peng L , Coram M , et al . (2016) Development and Validation of a Deep Learning Algorithm for detection of Diabetic Retinopathy in retinal fundus photographs. JAMA 316:2402–10.