Machine Learning Strategies for Medical Assessment

Chandrima Sinha Roy[1] and Dr. Tryambak Hiwarkar[2]
Research Scholar, Department of Computer Science[1]
Professor, Department of Computer Science[2]
Sardar Patel University, Bhopal, MP, India

Abstract: *Subfield of AI, is going to change the healthcare industry forever. However, it is not yet considered part of standard of care, especially when it comes to the treatment of specific patients. Whether or not data-driven methods are being used to support clinical making a Call (CDS). To date, there has been no comprehensive analysis of how research in machine learning and other data-driven techniques might effectively contribute to clinical care and what kinds of support they can bring to doctors. In this study, we investigate the potential contributions to clinical decision support systems of two data-driven fields: machine learning and data visualization. Here, we survey the research on three distinct CDS and how heuristic knowledge, machine learning, and visualization are now being used to analyse and improve them. Predictive modelling for alerts has been the subject of extensive study, although this technology is not yet integrated into CDS systems. Interactive visualizations and machine learning inferences are gaining popularity as a means to organize and review patient data, however these methods are still in the prototype stage and have not been implemented. We still lack CDS systems that could take use of prescriptive machine learning (e.g., individualized therapy suggestions). Possible explanations for the slow adoption of data-driven approaches in CDS are offered, along with suggestions for future study in this area. Clinical decision assistance; Visual analytics; Machine Learning.*

Keywords: Decision Tree Algorithm, Support Vector Machine, Random Forests

I. INTRODUCTION

Personalized, higher-quality, safer, and more efficient care is a goal of learning health systems. [1] The learning health systems pipeline involves the systematic collection of clinical data, the subsequent learning and generation of evidence, and the subsequent real-time feeding of that evidence back to physicians to aid in decision making. Delivering knowledge to the point of care is reliant on the widespread implementation of clinical decision support (CDS). However, many obstacles must be overcome before CDS can effectively support a learning health system's objectives. Some of the obstacles to efficient CDS implementation include a lack of patient-specificity, the use of overly simplistic CDS logic, a lack of generalizability, and the avoidance of human factor concerns. [2]

In order to address these issues, it may be useful to make use of recent advancements in machine learning and data visualisation, especially when used together. By creating new information from collected data, boosting patient specificity, aiding the discovery of complicated patterns, and increasing generalizability to diverse patients and conditions, machine learning (ML) technologies have the potential to enhance CDS tools. Clinical decision support (CDS) systems could benefit from the use of data and information visualisation (dataVis) approaches, such as static visualisations, interactive visualizations, and more advanced visual analytics, to help feed back information to doctors and increase transparency. In addition to enhancing CDS, machine learning and data visualization also have potential synergistic benefits (Figure 1). Therefore, there is compelling evidence in favor of putting more effort into combining machine learning with data visualization to facilitate the development of a learning health system.

Clinical decision assistance, machine learning, and data visualisation are all shown working together in Figure 1.

The purpose of this article is to provide a comprehensive summary of the current research on clinical decision assistance using machine learning and data visualisation techniques.

This analysis not only analyses successes and failures, but also highlights potential applications of machine learning and data visualisation in advancing CDS and the aims of a learning health system. Informaticians, designers, machine

DOI: 10.48175/IJARSCT-7690

learning professionals, and practitioners who are interested in learning how machine learning and visualisation synergy might tackle the problems plaguing clinical decision systems are the intended readers of this review.

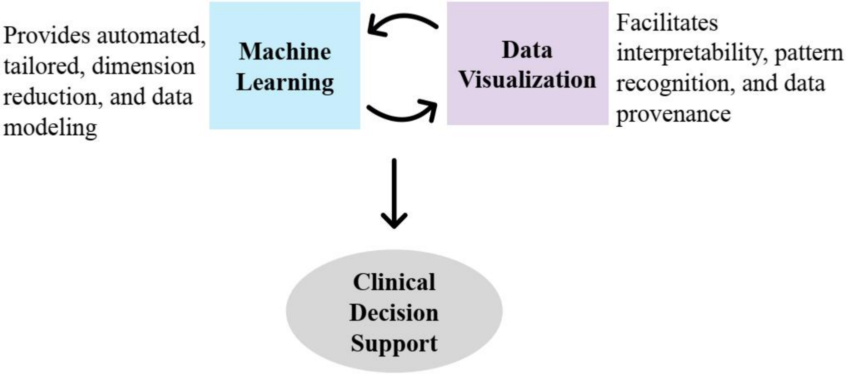
Provides automated, tailored, dimension reduction, and data modeling
Machine Learning
Data Visualization
Facilitates interpretability, pattern recognition, and data provenance
Clinical Decision Support

Figure 1: Process Structure of ML & Visualization

II. LITERATURE WORK

This analysis draws on a search of published works on both CDS and the methods used in CDS applications. Publication pool was generated by searching PubMed for "clinical decision support," "machine learning," and "visualisation." Machine learning for healthcare conference papers (Machine Learning for Healthcare, NeurIPS Machine Learning 4 Health Workshops) were also assembled, as were papers from the IEEE Visualization conference. Our search was primarily confined to works published between 2010 and 2019, although we also included seminal works published as far back as 1959. Pearl growth was used to find more publications till we had exhausted the topic. Our search was limited to articles that focused on clinician-facing clinical decision support, either patient-specific or cohort-level, that made use of EHR data gleaned through clinical record. The total number of papers considered for this analysis comes to 244. The papers were divided into three broad categories of CDS, although the insights gained from analysing the state of the art in each of these categories should be applicable to additional CDS. In this analysis, we use the terms "Infobutton," "Content Aggregation and Organization," and "Alert" to refer to the three distinct kinds of CDS defined by Musen and colleagues [3].

Info buttons are a form of CDS designed to expedite the process by which doctors find and access external resources, including as scholarly articles and practiseguidelines that are useful in the treatment of individual patients. Infobuttons facilitate staying up to date and well informed as medical evidence is continuously developed and updated and as physicians have less time at the point of service.

2.1 Content Aggregation and Organization

Clinical decision support (CDS) is used to restructure and present information at the patient or cohort level in a way that improves readability, pattern identification, and clinical decision making. As it stands, current EHR systems store vast volumes of data even for individual patients, making the processes of information gathering and synthesis mentally taxing and time-consuming. The goal of CAO CDS is to standardise and consolidate patient data so that it may be used more effectively in decision making.

The following portion of the paper compiles and summarizes prior work on each CDS type (Infobuttons, CAO CDS, and alert CDS), as well as the machine learning and data visualisation techniques employed. The various CDS-related literature is organized and defined according to the methodologies they employ (Figure 2). We examine the various deployments of each CDS type by examining: heuristics, which are expert-curated rules or knowledge-based sources like ontologies; (2) machine learning, which are data-driven and learning-based methods for knowledge development; (3) data visualisation, which is the sophisticated visual representation of data and information through static or interactive graphs, diagrams, or pictures; and (4) any combination of the aforementioned.

DOI: 10.48175/IJARSCT-7690

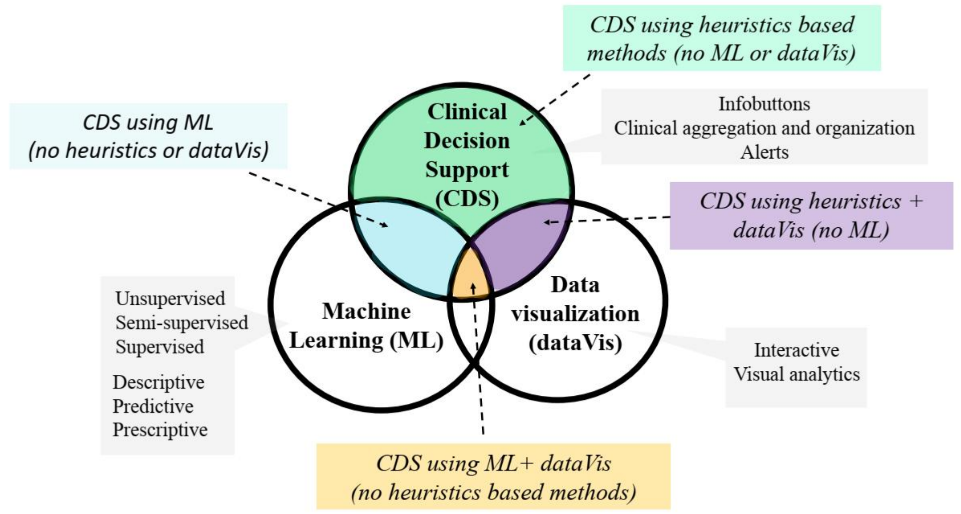
CDS using heuristics based methods (no ML or dataVis)
CDS using ML (no heuristics or dataVis)
Clinical Decision Support (CDS)
Infobuttons
Clinical aggregation and organization
Alerts
CDS using heuristics + dataVis (no ML)
Unsupervised
Semi-supervised
Supervised
Descriptive
Predictive
Prescriptive
Machine Learning (ML)
Data visualization (dataVis)
Interactive
Visual analytics
CDS using ML+ dataVis (no heuristics based methods)

Figure 2. Venn diagram showcasing the intersections between clinical decision support,

Data visualisation with machine learning. In this context, we mean by "heuristics-based approaches" rules that are either curated by experts or that draw on knowledge sources like ontologies. One use of machine learning is in the field of medicine, with data-driven clinical procedures currently under development. Static, interactive, and cutting-edge visual analytics on clinical data are all examples of visualisation approaches.

III. BACKGROUND DETAILS

One of the most successful methods for identifying disorders affecting the human body from the inside out is medical imaging. Although radiologists and physicians' diagnoses are essential to the process of recognizing abnormalities in collected images, the expansion of medical knowledge has made it impossible for them to keep track of all the possible diagnoses of different diseases. Machine learning in medical imaging can help radiologists of all levels make more accurate diagnoses of difficult cases.

A survey of the relevant literature reveals that scientists are also investigating the potential of machine learning algorithms in fields like protein-function prediction and gene expression [15], [16]. Machine learning approaches to protein function prediction, in contrast to sequence- and structure-based approaches, do not necessitate prior knowledge of homology or homology-derived parameters. As a result, there has been an increase in studies aimed at creating reliable machine learning methods for predicting protein function in the context of illness diagnosis.

In this study, we investigate the use of machine learning methods to the development of CAD systems. The numerous medical classification methods are briefly discussed in Section II. Decision trees, Support vector machines, Random forests, Evolutionary Algorithms, and Swarm Intelligence are all reviewed in detail in Section III. The conclusion highlights the need for further research in this field.

3.1 Decision Tree Algorithm

There are many different kinds of categorization algorithms, and one of them is the decision tree. The method employs a "divide and conquer" technique [17] to build the tree. Attributes connect the sets of instances together. The "nodes" of a decision tree stand for tests on attribute values, while the "leaves" stand for the "class" of instances that meet the criteria. The answer is either "true" or "false." The nodes along the route from the root to the leaf can be used as preconditions for a rule to predict the class at the leaf node. Branches need to be cut back to reduce redundant assumptions and steps.

Decision tree research for medical classification and disease diagnosis has garnered interest. Due to its simplicity and interpretability, decision tree data categorization techniques are popular [2]. Breast, liver, brain, and dermatologic disorders have been diagnosed with it.

Decision tree classifies breast cancer . Each research compares it to ANN, logistic regression, Bayesian network, KNN, and case-based reasoning. Case-based fuzzy decision tree diagnoses breast cancer with 99.5% accuracy. A three-classifier breast cancer detection decision support tool. BDT outperformed SDT with 98.83% accuracy and decision tree forest with 97.07% accuracy.

Yeh et al. [23] found that decision tree model was best for cerebrovascular illness with 99.59% accuracy.Luk et al. [24] proposed a CART model to distinguish Hepatocellular Carcinoma (HCC) from non-malignant liver tissue. HCC is the deadliest cancer because it's not detected till late. Decision tree methods were effectively employed for building classification model based on hidden pattern in the challenge dataset. brink et al. [21] proposed a hybrid model case-based reasoning and fuzzy decision tree (CBFDT) for liver illness with the maximum accuracy of 81.6%. In medical data, imbalance and cost sensitivity limit decision trees. Inconsistent data bothers them.

3.2 Support Vector Machine

Statistics-based learning systems, on which SVM algorithms are based, are widely employed for classification tasks. The SVM method finds the best border, or hyperplane, between two sets in a vector space without regard to the training vectors' probability distribution. The boundary that is farthest from the vectors that are closest to the border in both sets is located by this hyperplane. As the name implies, "supporting vectors" are those that are positioned in close proximity to the hyperplane. There may be no separating hyperplane if the space is not linearly separable. The issue is addressed by employing the kernel function. When applied to a set of data, the kernel function performs an analysis of the data's relationships and then partitions the space into numerous subsets.

For their great generalization performance, SVM algorithms have been advocated as a reliable statistical learning method for classification. An intuitive description of what a support vector machine (SVM) does is to discover the hyperplane in which the greatest number of points from a particular class lie on the same plane. The optimal separating hyperplane (OSH) is a hyperplane that helps reduce the possibility of incorrectly categorising test data.

Numerous studies have been conducted on SVM for medical diagnosis of breast cancer utilising Wisconsin breast cancer diagnostic (WBCD) data, and most of them have reported good classification accuracies [4]–[5]. For example, Padila and Lopez et al [6] employed least square SVM and achieved a precision of 98.53 percent. Additionally, a support vector machine (SVM) model using grid search and feature selection was developed for identifying breast cancer. There are two challenges that arise while working with support vector machines: deciding on the best possible input feature subset for SVM and determining the most effective kernel parameters. A good predictive and less computationally costly model can be achieved through feature selection, which restricts the number of input features in a classifier. The accuracy of an SVM classification can be enhanced through careful setup of the model parameters in addition to feature selection. The gamma (c) of the radial basis function (RBF) kernel is one example of a kernel function parameter that needs optimising. The grid search method is utilised to look for the best SVM parameters, and a modified version of F-score is used to determine the most relevant features. Using a training/test split of 80%/20%, the proposed model achieved a remarkable 99.51% accuracy. It is used a combination of artificial neural networks, support vector machines, and fuzzy logic to make breast cancer diagnoses. Improved MRI picture quality was achieved via application of a type-II fuzzy algorithm in the proposed method. After that, we used pulse-coupled neural networks for segmentation to pull out the relevant areas. These regions were used to extract wavelet features, which were then employed by SVM to make a final diagnosis and discriminate between distinct regions of interest to identify whether they represent cancer or not. The outcomes demonstrated that the suggested hybrid model using SVM provided more accuracy than other machine learning techniques like decision tree, neural network, etc.

Hybrid approaches involving SVM have been widely and successfully used in the medical diagnosis of a wide range of diseases, including diabetes, liver, and heart disease (Genetic + Fuzzy + SVM; prostate cancer (Artificial Neural Networks [ANN] + SVM; pain identification (ANFIS] + SVM;and Alzheimer's disease (nonnegative matrix factorization (NMF) + SVM; However, the SVM-based classification model has a practical limitation due to its black-box character. Using SVM rule extraction techniques or a hybrid-SVM model in conjunction with other, more interpretable models is one approach to fixing this problem.

DOI: 10.48175/IJARSCT-7690

3.3 Random Forests

The random forest technique is highly effective in data classification, and it can process massive datasets with minimal loss of information. It is a model-building technique based on ensemble learning, wherein many decision trees are built at training time, and the modal class is then produced. A random forest is a set of predictor trees where the predictions for each tree are based on the values of a random vector picked independently from the same distribution as the other trees in the set. The underlying idea is that several "weak learners" can combine their efforts to create a single "strong learner."

For their great generalisation performance, SVM algorithms have been advocated as a reliable statistical learning method for classification. An intuitive description of what a support vector machine (SVM) does is to discover the hyperplane in which the greatest number of points from a particular class lie on the same plane. The optimal separating hyperplane (OSH) is a hyperplane that helps reduce the possibility of incorrectly categorising test data.

Numerous studies have been conducted on SVM for medical diagnosis of breast cancer utilising Wisconsin breast cancer diagnostic (WBCD) data, and most of them have reported good classification accuracies [25]–[30]. For example, Polat and Gunes [26] employed least square SVM and achieved a precision of 98.53 percent. Additionally, a support vector machine (SVM) model using grid search and feature selection was developed [27, 28] for identifying breast cancer. There are two challenges that arise while working with support vector machines: deciding on the best possible input feature subset for SVM and determining the most effective kernel parameters. A good predictive and less computationally costly model can be achieved through feature selection, which restricts the number of input features in a classifier. The accuracy of an SVM classification can be enhanced through careful setup of the model parameters in addition to feature selection. The gamma (c) of the radial basis function (RBF) kernel is one example of a kernel function parameter that needs optimising. The grid search method is utilised to look for the best SVM parameters, and a modified version of F-score is used to determine the most relevant features. Using a training/test split of 80%/20%, the proposed model achieved a remarkable 99.51% accuracy. In their study, Hassanien and Kim [31] used a combination of artificial neural networks, support vector machines, and fuzzy logic to make breast cancer diagnoses. Improved MRI picture quality was achieved via application of a type-II fuzzy algorithm in the proposed method. After that, we used pulse-coupled neural networks for segmentation to pull out the relevant areas. These regions were used to extract wavelet features, which were then employed by SVM to make a final diagnosis and discriminate between distinct regions of interest to identify whether they represent cancer or not. The outcomes demonstrated that the suggested hybrid model using SVM provided more accuracy than other machine learning techniques like decision tree, neural network, etc.

Hybrid approaches involving SVM have been widely and successfully used in the medical diagnosis of a wide range of diseases, including diabetes, liver, and heart disease (Genetic + Fuzzy + SVM prostate cancer (Artificial Neural Networks [ANN] + SVM; [32]; pain identification (ANFIS] + SVM; and Alzheimer's disease (nonnegative matrix factorization (NMF) + SVM. However, the SVM-based classification model has a practical limitation due to its black-box character. Using SVM rule extraction techniques or a hybrid-SVM model in conjunction with other, more interpretable models is one approach to fixing this problem.

3.4 Evolutionary Algorithms

Breast cancer [4], prostate cancer [5], brain tumour [3], colon cancer, and heart disease are just few of the diseases for which genetic algorithms (GA) are utilised in designing the decision support system. When creating computer-aided diagnosis for specific organs, evolutionary algorithms (EAs) are often integrated with other classification algorithms to build hybrid systems. By combining GA, apriori, and a decision tree, cruz et al. [43] created a highly accurate model for diagnosing hypertension. Cancers such as leukemia, lymphoma, and colon cancer were all detected using a hybrid fuzzy + GA method . Some researchers believe that asymptomatic carotid stenosis is a major risk factor for stroke.

Evolutionary methods look for the best interpretation (disease presence/absence) of the mined data by determining the parameter values of the knowledge representation established by the designer. In order to feed the important boundary features of the brain tumor region to ANFIS [40], a GA efficiently searches for them. To improve lung sound prediction while decreasing processing load and time, a genetic algorithm (GA) looks for optimal structure and training parameters of neural network. A unique neuro-genetic algorithm has been implemented into a system for analyzing digital mammograms. First, features are extracted, and then the GA chooses the most important ones to feed into a neural network. We are quite pleased with the results that this method has produced. Models for the diagnosis of diseases, such as those for thalassemia, chest discomfort, and lung abnormalities, are developed using genetic programming in conjunction with other classification algorithms.

3.5 Swarm Intelligence

In order to optimise feature selection in classifier systems for medical diagnosis, swarm intelligence (SI) algorithms like PSO, ACO, and others are frequently utilised as pre-processing techniques. This improves classification accuracy while significantly reducing required processing resources. For gene expression data classification challenges, we employ an improved binary particle swarm optimization (IBPSO) to choose features (genes) and a K-nearest neighbour (K-NN) as an evaluator. The proposed technique achieved the maximum classification accuracy in 9 out of 11 test problems including gene expression data. Ovarian cancer detection using mass spectral data and an ACO as the feature selection method has shown great classification accuracy. The wavelet coefficients were chosen using an ACO. PSO/ACO is used to construct hybrid classification models for the detection of various diseases. Additionally, hybrid support vector machine (SVM) and PSO (PSO) models for breast cancer diagnosis have been developed. With an average accuracy of roughly 97.4% for breast cancer, the CBRPSO has been proven to surpass the other techniques.

With an average accuracy of 76.8%, the CBRPSO model is also applied to liver problems. High-accuracy hybrid models using PSO in conjunction with other classification algorithms have been proposed for a number of applications, including the diagnosis of coronary artery disease [2], leukaemia, and MR brain image classification [1]. Accordingly, SI algorithms have been utilised as optimization strategies in numerous domains, including function optimization, ANN training, fuzzy system control, and medical diagnosis.

IV. CONCLUSION

The integration of computer-based systems in healthcare can be aided by the successful application of machine learning algorithms in medical diagnosis. Especially in countries with a high mortality rate and a low number of doctors per capita, such as India, where there is only one doctor for every 1700 people, machine learning approaches in medical diagnostics can aid physicians in making accurate diagnoses and treating patients quickly. While technology can't take the place of a doctor's training and experience, it can free them up to focus on more complex procedures. Increasingly, doctors must rely on imaging technology to determine the nature of a patient's illness. Due to the growing difficulty in deciphering contemporary medical images, machine learning algorithms used to this field can greatly aid in medical diagnosis. They can increase the diagnosis accuracy, sensitivity, and specificity of interns and other less-experienced doctors by facilitating more consistent evaluation of medical pictures.

Machine learning approaches also play a crucial role in predicting protein functions. Since machine learning methods are both effective and cheap, they might be used to analyse massive amounts of complex biological data. The results of this study will benefit not only medical practitioners in terms of disease diagnosis, but also health planners in terms of disease diagnosis and prevention on a societal level.

REFERENCES

[1]. H. Brink and J. W. Richards, Real-World Machine Learning, 7th ed. Manning Pulications, 2013.

[2]. N. Esfandiari, M. R. Babavalian, A.-M. E. Moghadam, and V. K. Tabar, “Knowledge discovery in medicine: Current issue and future trend,” Expert Syst. Appl., vol. 41, no. 9, pp. 4434–4463, Jul. 2014

[3]. J. Nahar, T. Imam, K. S. Tickle, A. B. M. Shawkat Ali, and Y.-P. P. Chen, “Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer,” Expert Syst. Appl., vol. 39, no. 16, pp. 12371–12377, Nov. 2012.

[4]. A. T. Azar and S. M. El-Metwally, “Decision tree classifiers for automated medical diagnosis,” Neural Comput. Appl., vol. 23, no. 7–8, pp. 2387–2403, Dec. 2013.

[5]. S. N. Deepa and B. A. Devi, “A Survey on Artificial Intelligence Approaches for Medical Image Classification,” Indian J. Sci. Technol., vol. 4, no. 11, pp. 1583–1595, Nov. 2011.

DOI: 10.48175/IJARSCT-7690

[6]. P. Padilla, M. López, J. M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, and Alzheimer‟s Disease Neuroimaging Initiative, “NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer‟s disease,” IEEE Trans. Med. Imaging, vol. 31, no. 2, pp. 207–216, Feb. 2012.

[7]. D. Delen, G. Walker, and A. Kadam, “Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods,” ArtifIntell Med, vol. 34, no. 2, pp. 113–127, Jun. 2005.

[8]. N. Cruz-Ramírez, H. G. Acosta-Mesa, H. Carrillo-Calvet, L. Alonso Nava-Fernández, and R. E. Barrientos-Martínez, “Diagnosis of breast cancer using Bayesian networks: A case study,” Comput. Biol. Med., vol. 37, no. 11, pp. 1553–1564, Nov. 2007