Application of Machine Learning in Medical

Sweta Kumari Barnwal, Assistant Professor

Department of Computer Science, Arka Jain University, Jamshedpur, Jharkhand, India
Email Id- sweta.b@arkajainuniversity.ac.in

ABSTRACT: The article offers a machine learning-based overview of the evolution of intelligent data analysis in medicine, including a historical perspective, a state-of-the-art viewpoint, and an outlook on potential future developments in this area of applied artificial intelligence. The purpose of this article is not to give a complete review, but rather to explain certain subareas and directions that, in my opinion, are critical for using machine learning in medical diagnostics. The naïve Bayesian classifier, neural networks, and decision trees are highlighted in the historical review. The author offer a comparison of several cutting-edge systems that represent each field of machine learning when applied to a variety of medical diagnostic tasks. Two case studies are used to highlight future tendencies. The first discusses a newly discovered technique for dealing with the consistency of classifier choices, which seems to have promise for intelligent data analysis in medicine. The second presents a method for utilizing machine learning to validate certain inexplicable occurrences in complementary medicine, which is not (now) recognized by the orthodox medical community but may play a significant role in total medical diagnosis and treatment in the future.

KEYWORDS: Machine Learning, Medical Diagnosis, Accuracy, Validity, Comprehensibility.

1. INTRODUCTION

Machine Learning (ML) is a set of methodologies, strategies, and tools that may be used to solve diagnostic and prognostic issues across a wide range of medical fields. For the examination of the significance of clinical factors and their combinations, machine learning is being utilized[1]. Prediction of illness development, extraction of medical information for prognosis study on outcomes, treatment planning and support, and overall patient care Data analysis, such as the identification of regularities, is also aided by machine learning. Interpretation of continuous data in the data by properly dealing with faulty data in the Intensive Care Unit, data is utilized, and intelligent alerting is used, resulting in effective and efficient care. monitoring that is effective It is claimed that the effective application of ML techniques is dependent on the availability of data[2]. Can assist in the integration of computer-based technologies in a healthcare setting allowing possibilities for medical professionals' work to be facilitated and improved and ultimately, to enhance medical care efficiency and quality. We've included some examples below[3].

Enumerate some of the most important machine learning applications in medicine Medical diagnostic reasoning is a critical use of intelligent reasoning. Expert systems and model-based methods are used in this framework. offer tools for hypothesis development based on patient data as an example, Expert systems are built by extracting rules from expert knowledge[4]. Unfortunately, professionals may not know or be able to formulate, in many instances. What knowledge people really use to problem-solving learning via symbols to supplement learning, methods such as inductive learning by example are used. Expert systems with knowledge management skills: given a collection of clinical cases Learning in intelligent systems may be accomplished using ML as an example. Techniques that may provide a systematic description of the clinical characteristics that are important describe the clinical circumstances in a unique way. As a result, knowledge may be represented in a variety of ways. Simple rules or a decision tree are often used. This is a typical illustration of this kind of situation[5]. KARDIO is a system that was created to read ECGs. This method may be used in situations when there is no prior knowledge. Knowledge of how to analyze and comprehend medical data. To identify changes in a patient's state, cardiac physiology is used. Furthermore, according to a study these models may be used to generate early hypotheses that can lead to further research. experimenting[6]. Because datasets are large, learning from patient data presents a number of challenges. Incompleteness (missing parameter values), and incorrectness (systematic errors) sparseness (few and/or non-representable patient records), or random noise in the data available), and imprecision (inappropriate selection of parameters for the given task). Medical datasets have similar features, and ML offers methods to deal with them. Sub symbolic learning techniques, particularly neural networks, are capable of dealing with these situations. datasets, which are mostly utilized for their pattern matching capabilities and human-like appearance [7]. Features (generalization, noise robustness) in order to enhance medical care. Making decisions. Biomedical signal processing is another area of use. Because our knowledge of biological systems isn't comprehensive, there are a few things we need to know. Characteristics and information not immediately apparent in physiological signals evident. Furthermore, the impacts of the many subsystems are indistinguishable. Biological signals are characterized by significant variability, which may be produced by a variety of factors. Internal processes that arise spontaneously or as a result of external stimulation.

The connections between the Different parameters may be too complicated to solve using traditional methods. ML techniques make advantage of these collections of data, which are simpler to generate and may aid in the discovery of new information. Model and extract parameters from the nonlinear connections that exist between this data as well as characteristics that may help enhance medical treatment. Computer-based medical image interpretation systems make up a significant portion of the market. Medical diagnosis is greatly aided by this application area[8]. Most of the time, the creation of these systems is seen as an effort to solve a problem. Imitate the doctor's skill in detecting cancerous areas in minimally invasive surgery techniques using invasive imaging (e.g., computed tomography, ultrasonography, etc.) endoscopy, confocal microscopy, computed radiography, or magnetic resonance imaging are some of the techniques used). The goal is to improve the expert's capacity to spot cancerous areas. Thereby reducing the need for assistance and preserving the capacity to make correct decisions diagnosis. It may also be feasible to investigate a wider region, examining living conditions. tissue in vivo, perhaps from afar , and therefore mitigate the drawbacks of biopsies, such as patient pain, diagnostic delays, and a restricted number of biopsies samples of tissue [9].The need for more effective early detection techniques, such as those described above It is self-evident what computer aided medical diagnostic systems seek to offer.

The potential of machine learning in this field is enormous, since it supplies us with a wealth of information. techniques for collecting, altering, and updating knowledge in a computer Medical image interpretation systems that are intelligent, and in particular, learning systems tools to aid in the induction of knowledge from instances or data Especially novel imaging concepts in minimally invasive imaging techniques, such as ML techniques may be helpful in a variety of situations, including fluorescence imaging and laser scanning microscopy. There aren't any algorithmic answers, there aren't any formal models, or Due to a lack of prior experience, understanding of the application area is limited. In the interpretation of the collected pictures, you should have experience and/or medical competence[10]. Many real-world datasets have an unbalanced class distribution, with a majority class containing normal data and a minority class containing aberrant or important data. These datasets include fraud detection, network intrusion, and medical diagnosis; however, unlike other machine learning applications, the medical diagnostic issue does not stop once we have a model to categorize new cases. That example, if the instance is categorized as ill (the most serious category), the produced information should be able to offer medical personnel with a fresh perspective on the issue. This may aid in the timely use of a medical therapy in order to prevent, postpone, or reduce the disease's occurrence.
Then, in addition to classification accuracy, the comprehensibility of diagnostic information should be considered. We must also address a separate issue: the identification of relevant characteristics (or risk factors). We should concentrate on changeable characteristics (those that can be altered with medical treatment), such as blood pressure or cholesterol levels, rather than non-changeable variables like age and sex (which are typically excellent categorization attributes). This makes the categorization job much more difficult.

Another significant consideration is that medical datasets utilized in machine learning should be indicative of the disease's overall occurrence. This is necessary in order to share the information produced with other populations. As a result of the artificial manipulation of the datasets, over-sampling and under-sampling techniques commonly used to balance the classes and improve the minority class prediction of some classifiers may generate biased knowledge that is not applicable to the general population. As a result, we suggest a new approach that aims to improve the classification accuracy of the minority class while preserving the original dataset. As a result, each step of our algorithm is aimed towards achieving this goal. Because we're working with binary classification issues, we can assume that the majority class accuracy will be assured. The technique of our algorithm .The performance of our method is then compared to that of various conventional classifiers, emergent classifiers, and classifiers specially designed for unbalanced datasets.

This comparison refers to the acquired findings' accuracy, comprehensibility, and validity (only for symbolic classifiers). We analyze the findings, we offer our conclusions and plans for future study. An invited presentation on endoscopic procedures and the use of machine learning approaches in this setting. The speaker discussed the present limits of endoscopic methods, which are linked to endoscopy's access restrictions to the human body. Technical constraints in this regard include: limitations in manual capacities to handle human organs via a tiny access, difficulties in seeing tissues, and limitations in obtaining diagnostic information about tissues. International technological advancements are focusing on the solutions to these issues. Moustakis and Charissis surveyed the role of machine learning in medical decision making and provided an extensive literature review on various ML applications in medicine that practitioners interested in using ML methods to improve the efficiency and quality of medical decision making systems might find useful.

The need of moving away from accuracy measurements as the primary criterion for evaluating learning algorithms was emphasized in this paper. The problem of comprehensibility, or how effectively a medical expert can comprehend and use the findings of a system that uses ML techniques, is critical and should be carefully addressed throughout the assessment.

Published a study on the use of inductive ML techniques in the medical diagnosis of stroke. The suggested method used the See5 algorithm, which is an improved version of the C4.5 algorithm. This method demonstrated the capacity to learn from examples and manage missing information by building a decision tree that could be converted into IF/THEN rules in the tests described. By contacting medical specialists, special emphasis was paid to determining the complexity and comprehensibility of the obtained decision rules. The Magnus Assistant decision tree learner and the Bayesian classifier were employed in the article by for the diagnosis and prognosis of first cerebral paroxysm. Despite the fact that the naïve Bayesian classifier produced the best predictions, the Magnus Assistant decision tree learner produced the most intriguing findings from a medical standpoint. Expert neurologists regarded data and characteristics to be apparent and useless, but they turned out to be crucial for automated diagnosis and prognosis. In this instance, machine learning techniques gave a different estimate of certain clinical characteristics, which prompted doctors to develop new hypotheses and, as a result, enhance their conventional diagnostic and prognostic procedures. Developed an interactive method for determining visual perception problems. Using a modified version of the ML algorithm Charade, the system conducted data analysis and identified intriguing correlations between visual perception problem and brain injury. The suggested method seems to be successful in rehabilitating people with specific brain abnormalities, according to preliminary findings. Built on earlier work on medical expert systems for ECG diagnosis by integrating machine learning techniques to continually enhance a medical expert system's knowledge base. According to the findings, the new system shows continuous learning capabilities by using an expanded version of the ID3 algorithm to extract a set of diagnostic criteria based on a training set of ECGs from time to time. Duplicates are eliminated and the extracted rules are integrated into the older ones. A knowledge management subsystem monitors the performance of the final rules in terms of diagnostic accuracy and changes the knowledge base in order to improve the system's performance showed the necessity for Health Care Unit's medical imaging models, as well as the idea of a generic and the development of novel manipulation methods combining robots and intelligent sensor devices for more precise endoscopic treatments. This new generation of sensor devices is recognized for contributing to the development and dissemination of intelligent systems in medicine by supplying data for ML techniques to analyze.

Suturing in cardiac surgery and other therapeutic areas are examples of current uses. Several research organizations have said that they are concentrating their efforts on the development of new endoscopic visualization and diagnostic technologies. New imaging concepts, such as fluorescence imaging or laser scanning microscopy, and machine learning techniques have enormous promise in this area. Early identification of malignant tumors in phases when local endoscopic treatment is feasible is the clinical rationale underlying these advances. Although technical advancements in this area are promising, clinical findings are still waiting, and continuing study will be required to explain the deductive/inductive model of operation, which enables scheduling, forecasting, and accounting. Using this method, multiple agents collaborated to generate hypotheses, each with a distinct role to play in assessing various aspects of the data, and neural networks were used to find connections in the dataset. These technologies have a lot of therapeutic promise. First, we must consider why using machine learning to medical diagnostics is so challenging.

The poor performance of the utilized classifiers in terms of AURC is an example of this, since the best performance in certain instances did not exceed 70% (Cardiovascular Disease and Breast Cancer domains). Onea explanation for this is because the majority of the datasets are derived from longitudinal medical studies, which include tracking the progression of an illness in a group of people over time. At the conclusion of the research, each participant is assigned to one of two categories: healthy or ill. However, an individual who presented clear risk factors during the study period but died from a cause other than the studied disease (i.e. an accident), or who did not present the disease at the end of the study (it is highly likely that he developed it shortly after the end of the study), is classified as healthy, and this situation tends to perplex the classifiers. Despite this inconvenience, REMED demonstrated consistent performance across all domains, ranking top in terms of AURC and sensitivity. Nave Bayes and 1-NN were two more classifiers that consistently delivered consistent findings, but they were not symbolic classifiers, and the outputs were not very comprehensible. When it comes to comprehensibility, REMED always creates rule systems that are just two rules long.

Ripper generated basic rule systems in all of the situations examined, although not as precisely as REMED. In the hepatitis domain, for example, both classifiers seem to generate comparable rule systems, but REMED much outperforms Ripper in terms of AURC and sensitivity, whereas C4.5 consistently produces bigger rule systems than REMED and Ripper. REMED also has the benefit of avoiding producing rules with enclosed intervals, which is useful in medical areas. This is significant since the probability of getting an illness is directly proportional to the rise or decrease of a risk factor, making it uncomfortable in medical diagnostics. Furthermore, an increase or decrease in a risk factor may be linked to two distinct illnesses (i.e. hypothyroid and hyperthyroid).

1.1 Application Of Machine Learning In Medical Diagnosis:

- *Identifying Diseases and Diagnosis:* One of the most important uses of machine learning in healthcare is the detection and diagnosis of illnesses and disorders that are otherwise difficult to identify. This may range from malignancies that are difficult to detect in their early stages to other hereditary disorders. IBM Watson Genomics is an excellent illustration of how combining cognitive computing with genome-based tumor sequencing may aid in the rapid detection of cancer. Berg, the biopharmaceutical behemoth, is using artificial intelligence to create medicinal solutions in fields like cancer. Predict (Predicting Response to Depression Treatment) is a project by P1vital that seeks to provide a commercially viable method to diagnose and treat depression in regular clinical situations.

- *Drug Discovery and Manufacturing:* One of the most common clinical uses of machine learning is in the early stages of drug development. This includes research and development technologies like as next-generation sequencing and precision medicine, which may aid in the discovery of new treatment options for complex illnesses. Unsupervised learning, which can detect patterns in data without making predictions, is now used in machine learning methods. Microsoft's Project Hanover is utilizing machine learning-based technologies for a variety of projects, including creating AI-based solutions for cancer therapy and customizing medication combinations for AM patients.

- *Medical Imaging Diagnosis:* The revolutionary technique known as Computer Vision is the result of combining machine learning and deep learning. This has been accepted by Microsoft's Inner Eye project, which focuses on image diagnostic tools for image analysis. Expect to see additional data sources from various medical images becoming a part of this AI-driven diagnosis process as machine learning becomes more accessible and their explanatory ability grows.

- *Personalized Medicine*: By combining individual health with predictive analytics, personalized therapies may not only be more successful, but they are also ripe for further study and improved illness evaluation. Physicians are now restricted to choose from a list of diagnoses or estimating the patient's risk based on his clinical history and accessible genetic information. However, machine learning in medicine is progressing rapidly, and IBM Watson Oncology is at the forefront of this trend, using a patient's medical history to produce numerous therapy choices. More gadgets and biosensors with advanced health monitoring capabilities will reach the market in the coming years, enabling more data to become easily accessible for cutting-edge ML-based healthcare solutions.

- *Machine Learning-based Behavioral Modification:* Behavioral modification is an essential component of preventive medicine, and with the rise of machine learning in healthcare, a slew of new companies have emerged in the areas of cancer prevention and detection, patient treatment, and so on. Somatix is a B2B2C data analytics business that has developed a machine learning-based software that recognizes movements we make in our everyday lives, enabling us to understand our unconscious behavior and make required adjustments.

- *Crowdsourced Data Collection:* Crowdsourcing is all the rage in the medical profession these days, enabling academics and practitioners to access a massive quantity of data that individuals have contributed with their permission. This real-time health data will have a significant impact on how medicine is viewed in the future. Users may utilize Apple's Research Kit to access interactive applications that employ machine learning to cure Asperger's and Parkinson's illness. Based on crowdsourcing data, IBM has collaborated with Medtronic to interpret, collect, and make diabetes and insulin data accessible in real time. With IoT advances, the healthcare sector is constantly finding new methods to utilize this data to handle difficult-to-diagnose situations and aid in overall diagnosis and treatment improvement.

- *Outbreak Prediction:* Machine learning and AI-based technologies are now being used to monitor and forecast epidemics all around the globe. Scientists now have access to a vast quantity of data gathered from satellites, as well as real-time social media updates, website information, and other sources. Artificial neural networks aid in the collection of this data and the forecasting of anything from malaria outbreaks to severe chronic infectious illnesses. Predicting epidemics is particularly beneficial in third-world nations, where medical infrastructure and educational systems are lacking. The Proem-mail, an Internet-based reporting tool that monitors developing and emerging illnesses and delivers real-time epidemic notifications, is a prime example of this.

In general, it appears that, as the healthcare environment becomes increasingly reliant on computer technology, the use of machine learning methods can provide useful aids to physicians in many cases, eliminate issues associated with human fatigue and habituation, provide rapid identification of abnormalities, and enable real-time diagnosis.

Next, we'll go through some of the successful applications of machine learning techniques presented during the Workshop on Machine Learning in Medical Applications, which was sponsored by the ECCAI Advanced Course on Artificial Intelligence for 1999 (ACAI '99) in Chania, Crete, Greece, on July 15th, 1999. The workshop's goals were to foster fundamental and applied research in the application of machine learning methods to medical problem solving and research, provide a forum for reporting advances in the field, determine whether machine learning methods can support research and development on intelligent systems for medical applications, and identify areas where more research is likely to be conducted.

Eleven refereed papers and one invited article were presented during the workshop, which focused on the ideas and techniques that underlie the application of machine learning technologies in medicine. A variety of research agenda suggestions were made, covering both technical and human-centered problems. A short summary of these contributions is given in the next section, and the chapter ends with a broad discussion of the development and use of machine learning techniques in the medical field.

2. DISCUSSION

Most of the ML research for medical applications has been focused on technical problems and is mostly application driven. However, it is critical to improve our knowledge of ML algorithms and give mathematical explanations for its characteristics in order to answer basic issues and get valuable insight into the performance and behavior of ML techniques in a medical setting. Furthermore, we must overcome a number of challenges related to the practice of acquiring knowledge, such as display of learnt information, extraction of comprehensible rules from neural networks, and detection of noise and outliers in data. Other difficulties that emerge in ML applications in medicine, such as overfitting management and the scaling characteristics of ML techniques so that they can be used to solve problems with big datasets and high-dimensional input (feature) and output (classes-categories) spaces, need to be investigated further.

The need for comprehensibility of the learning outcome, relevance of rules, criteria for identifying applications and assessing their feasibility, integration with patient records, as well as the description of the appropriate level and role of intelligent systems in healthcare are all common concerns for ML applications in the medical context. Because technological, organizational, and societal problems are all interwoven, these challenges are very complicated. Previous research and experience suggest that successful implementation of information systems and in particular decision support systems, in the healthcare sector is dependent on the successful integration of technology with the organizatio191190 nal and social context in which it is used. Medical data is essential for patient diagnosis and treatment, therefore the ethical concerns that arise throughout its life cycle are critical. Understanding these problems is becoming more important as these technologies become more widely used. Some of these difficulties are system-centered, i.e., they are linked to the ML research's intrinsic problems. Humans, not algorithms, are the only ones who can behave as moral actors. This implies that only humans are capable of recognizing and dealing with ethical problems. As a result, it's critical to look at the developing difficulties and ethical concerns from a human-centered standpoint, taking into account the motives and ethical dilemmas of researchers, developers, and medical users of machine learning techniques in medical applications.

3. CONCLUSION

Machine learning methods to create a systematic judgment for medical picture diagnosis and prediction. The representation of machine learning has been examined, i.e. based on a variety of techniques that concentrate on prediction, based on known characteristics acquired from the training data. The accuracy rate of current techniques is low, according to the literature review, therefore improvements are needed to make them more consistent, since Naive Bayes surpasses KNN and SVM in terms of accuracy. The most essential objective is to make a benchmark database of Ultrasound scanned pictures available to the public so that various algorithms based on CAD systems may be actively compared and evaluated. REMED is a highly competitive algorithm that may be utilized in medical diagnostics, by the findings.

However, it is important to note that REMED does not claim to be the panacea of machine learning in medical diagnostics, but rather a good approach with the desired features to solve medical diagnostic tasks, such as good performance, comprehensibility of diagnostic knowledge, ability to explain decisions, and the algorithm's ability to reduce the number of tests required to obtain reliable diagnostic results. It's also worth noting that the REMED method can be scaled to operate with datasets far bigger than those utilized in our tests. This is because REMED's complexity is O(n2), and regardless of how many instances there are or how many starting attributes there are, REMED always generates basic rule systems with only two rules (including the default rule: else class = 0) and a maximum of m conditions per rule. However, we still need to work on improving REMED's performance; one possibility is to combine REMED with Boosting methods or Cost-Sensitive strategies. We also aim to improve REMED's flexibility by adding features that enable it to take into account discrete characteristics, deal with multi-class issues, and, in certain instances, create rule systems with enclosed intervals. This will be done so that REMED may be used in domains with unbalanced datasets.

REFERENCES

[1] M. Gagliano, "Applications of Machine Learning in Medical Diagnosis," *Int. Conf. Comput. Sci.*, 2017.

[2] H. Zhou, J. Tang, and H. Zheng, "Machine Learning for Medical Applications," *Scientific World Journal*. 2015, doi: 10.1155/2015/825267.

[3] A. El-Baz, G. Gimel'Farb, and K. Suzuki, "Machine Learning Applications in Medical Image Analysis," *Computational and Mathematical Methods in Medicine*. 2017, doi: 10.1155/2017/2361061.

[4] "Application of Machine Learning in Medical Image Analysis," *Med. Imaging Inf. Sci.*, 2017, doi: 10.11318/mii.34.66.

[5] A. Criminisi, "Machine learning for medical images analysis," *Medical Image Analysis*. 2016, doi: 10.1016/j.media.2016.06.002.

[6] V. Bolón-Canedo, B. Remeseiro, A. Alonso-Betanzos, and A. Campilho, "Machine learning for medical applications," 2016.

[7] H. Li, D. Li, C. Zhang, and S. Nie, "An application of machine learning in the criterion updating of diagnosis cancer," 2005, doi: 10.1109/icnnb.2005.1614594.

[8] J. M. Tomczak and M. Zięba, "Probabilistic combination of classification rules and its application to medical diagnosis," *Mach. Learn.*, 2015, doi: 10.1007/s10994-015-5508-x.

[9] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*. 2017, doi: 10.1007/s12194-017-0406-5.

[10] G. Wu, D. Shen, and M. R. Sabuncu, *Machine Learning and Medical Imaging*. 2016.