

An Analytical Study on health prediction system using machine learning

Priyanka Kadu, Dr. Amar Buchade

Department of Computer Engineering,
Pune Institute of Computer Technology, Pune, India

Abstract: There has been unprecedented growth and proliferation of technology into every sector of human life. The technology has been envisioned as a source of increasing convenience in just about all the walks of life. Health care is also one of those areas where technology has influenced the inner workings. Health care has been augmented with the help of Machine Learning tools that enable us to predict and revolutionize the way we deal with bedridden and elderly patients. Most of these applications collect a lot of data which is manually processed before being analyzed, which increases the time taken and decreases the accuracy of the system. Therefore, an innovative technique for health prediction has been proposed that utilizes Hidden Markov Models for providing accurate and timely predictions to the Healthcare Professionals.

Index Terms - Hidden markov model, Health prediction, K- nearest neighbor clustering

I. INTRODUCTION

Health Care has been a subject of large debates recently as there has been increased interest in individuals for increasing the quality of the healthcare provided. Even when active research has been going on for centuries, there has been increasing benefits that are obtained from a better healthcare system. This is evident by the fact that life expectancy has gradually risen throughout the years to about a theoretical maximum of about 70 years. This is largely due to the extensive amount of research being done.

Humans have been actively searching for cures of various diseases and also creating cost-effective and efficient cures for the diseases that have already been cured to provide access to the less privileged among society. The Healthcare system has only one motto, the psychological and physical well-being of humans and the systems that deal with the maintenance of health.

Health care is usually centered around the identification of symptoms that a diseased human would go through, the identification then can be utilized to ascertain what kind of an ailment the person has contracted and necessary counter-measures are taken. Health care systems are usually very complicated as there is a plethora of diseases that a human being encounters and each disease has a various set of symptoms that are unique to that disease.

The doctors or medical practitioners are trained to identify the various ailments with the help of the symptoms that are being demonstrated. Most of the time some symptoms overlap with other related diseases, the doctor has to carefully isolate the actual condition that is being manifested in the human. This is done by eliminating possible and improbable diseases and diagnosing the patient accurately.

The process of diagnosis is a very complicated one where the communication between the doctor and the patient must be clear and unadulterated. As communication is the main ingredient for the doctor to understand the symptoms of the disease. This is not always the case though, as most of the time the doctor would be distracted or the patient is forgetful and does not mention a critical symptom to the doctor. There could also be a language barrier between them and this would lead to a lot of misunderstanding between them, which can be very dangerous.

To ascertain the cause of the disease more accurately in this regard, the doctor has to perform some tests, such as the blood test to analyze the chemistry of the blood and if there are any imbalances in the level of the hormones or the different cells that are in the blood. Any discrepancies in the tests can be an indication of an ailment. The doctor could also perform other tests, such as X-Ray in the even of pain or injury to ascertain if there are any fractures in the bone.

The doctor can also perform an ultrasound, which is basically a technique for imaging the various parts of the body with the help of ultrasonic waves and determining if there are any irregularities in the structure. There can be ECG or Electrocardiogram performed to determine if the symptoms are heart-related. Any changes in the rhythm of the beating of the heart can be a sign of cardiac disease. Most of these tests are performed by professionals and they utilize special equipment for these procedures.

As some patients can be old or have restrictive movement and therefore, are unable to perform the tests. Many times there are discrepancies in the data represented and also most complex tests take a lot of time to be evaluated. In an emergency when the patient has to be treated right away to not face any complications, the various tests would prove to be highly dangerous and can ultimately lead to fatal consequences.

This can be eliminated to some extent with the introduction of machine learning into the systems and utilizing it for achieving a higher level of understanding with the help of faster processes and accurate results can be achieved. This is very useful in the area of Healthcare as it is one of the most crucial aspects of the application. Machine Learning is one of the most versatile technique to be used for the application of Healthcare.

K Nearest Clustering is one of the most common techniques for the clustering of the dataset according to the mentioned parameters. It is one of the supervised learning methods that have to be trained prior to use on the dataset for the clustering purpose. K- Nearest Neighbors is very popularly used across many applications such as pattern recognition, mining of data, intrusion detection etc. The K- Nearest Neighbor can be used for the purpose of classification or calculating the regression as it is a very versatile algorithm.

Hidden Markov Models are based on the Markov models that have been presented by scientists in the early 1960's. It has been in popularly used in various applications since in its inception as it has been the easiest form of the Dynamic Bayesian Network. The Markov model and the Hidden Markov Model differ only in one aspect in their workings. The states that govern the outcome of the model are visible in the Markov Model whereas they are hidden in the Hidden Markov Model.

Hidden Markov Model has been used in a variety of applications, they have been mainly used for determining various handwritings and matching it to the particular person. It has also found applications in the field of OCR or Optical Character Recognition. As the Hidden Markov Models work with hidden states, they are particularly useful for speech recognition and bioinformatics as they benefit from this particular characteristic of the Hidden Markov Models.

In this paper, section 2 is dedicated for literature review of past work and Finally Section 3 concludes this paper.

II. LITERATURE REVIEW

M. Healy [1] explores the different ways through which there can be a better understanding of the conditions a particular patient is going through. The authors try to link the demeanor of the patients to a better understanding of their situation which will help the doctors to provide personalized healthcare for the patient. For this purpose, the authors gauged the patients in an emergency situation with the help of a camera. The patients were authenticated by the facial recognition features of the Intel Real Sense camera. The authors achieved higher accuracies by implementing the Support Vector Machine.

Research Gap – The authors have implemented this mechanism utilizing specialty cameras from Intel called the Real Sense and have only implemented this technique on Wi-Fi, there should be an implementation in the form of 5G networks for better reach.

M. Krajnak [2] explores the mortality rates of the people being admitted in the ICU. As the ICU is a very critical place and patients that reach the ICU are usually on the verge of death. It is essential for doctors to know what is the possibility that a particular patient might survive. Therefore, the authors presented an innovative technique of performing various operations through a rule-based model and structure to the Fuzzy-rule based system that can ascertain with a high degree of accuracy the rate of mortality of the patient. The technique was tested on the Physio Net challenge dataset and returned very promising results.

Research Gap – The proposed technique has a very high computational complexity and very high time complexity that is an exceptional hurdle as longer wait times will negatively impact the proceedings in the ICU.

J. Yoo [3] explores the incidence of seizures in patients suffering from epilepsy. The authors have stated that there is a correlation between the incidence of a patient suffering from an episode. But this isn't utilized by the doctors as every individual has a particular pattern which differs from patient to patient. To ameliorate this effect, the authors have implemented a technique utilizing machine learning in the form of an LSVM, Non-linear SVM, and DDA-SVM. These machine learning tools extract the patterns from the ictal patterns of the patient.

J. Alcaraz [4] expresses concern over the rehabilitation for the patients undergone a deadly disease or a treatment, as it is one of the most difficult time for the patient. The research aims to achieve real-time acoustic feedback with the help of wearable devices. The authors implemented this technique in conjunction with the machine learning principles and were successfully able to provide care and rehabilitation to people who cannot afford to be rehabilitated in the hospital with exorbitant fees. The proposed system has been proven to be better than traditional rehabilitation methods and is comparatively quicker and efficient.

Research Gap – The authors have implemented a system of acoustic feedback but it is very simple and does not employ any signal processing methods for increasing the accuracy.

J. Wojtusiak [5] explores the semantics of healthcare data, as one of the most important aspects of the data. The authors have tried to extract and utilize a lot of semantics that has originated from the Healthcare data. The researchers used the Unified Medical Language System, UMLS for the extraction of the semantic data. A rule-based learning method was implemented and the data was extracted from a lot of sources with the help of supervised as well as an unsupervised learning process.

Research Gap – The proposed technique is highly computation extensive as the various algorithms also take a lot of storage space for their calculations.

M. Mozaffari-Kermani [6] explains a technique for the security of data utilized for the purpose of machine learning. There have been a lot of studies that indicate that machine learning is one of the most vulnerable when it is in the training phase, where it can be corrupted with the help of spurious data. These attacks are commonly known as Poisoning Attacks. To ameliorate this effect, the authors have implemented a scheme to introduce errors into the training data to get arbitrary errors on the output.

Research Gap – The technique does solve the problem of the data poisoning attacks but is highly inefficient and needs to be further developed before using.

A. WongKoblap [7] expresses concern over the rising incidences of depression among the populace of late. As it is one of the most common types of mental illness and is quite difficult for the person suffering to go through such an ailment. The patients lose interest in most of the activities and then start to become introverted, which ultimately leads to suicide. As there has been a lot of research done on this topic involving various machine learning tactics, therefore the authors have proposed a framework that utilizes a deep learning predictive model of the prediction and prevalence of depression among individuals.

Research Gap – The innovative framework is quite efficient but the performance fails to impress as the performance has to improve. The authors also aim to implement further participation from the candidates with comments and replies.

H. Sharma [8] presents an innovative technique for the purpose of conjugating the rule-based machine learning techniques and the statistical methods that are used to predict certain behaviors. The researchers have utilized the features from the unstructured texts with the help of UMLS. The extracted features have been combined with the OMOP data model which increases portability between various institutions and data systems. The authors have also utilized various NLP features such as extend, adapt and reuse by using the various components from the OMOP data model stored by the rules-based system.

Research Gap – The authors have successfully implemented this technique with promising results, but the increased amount of time required to process all the data is quite large and needs improvement.

J. Park [9] states that there's a rising need for a highly accurate means of determining a person's psychological wellness. The traditional methods have been unable to give accurate responses and are using the old techniques for their evaluation. These factors are detrimental to the process and introduce various inaccuracies in the system which reduces the overall reliability of the system. The authors, therefore, propose a technique based on machine learning for the purpose of developing a psychological wellness index.

Research Gap – The authors have utilized various machine learning algorithms such as K-nearest neighbor, regression neural network, etc. All these components were used to make the prediction but there have been issues with its accuracy that need to be fixed.

A. Charleonnann [10] explains that there is a need for supporting the doctors so that they can provide timely care to critical patients, as there is always a shortage of time, the authors developed a system for the identification of the accurate treatment for the patients in a short duration of time. This system was capable of detecting kidney diseases in patients by utilizing various algorithms such as Logistic regression, K-nearest neighbors. These techniques are highly powerful and have been trained with Chronic Kidney Disease Dataset.

N. Chetty [11] states that there has been a lack of a system that can accurately predict the disorders of the liver including diabetes. Therefore, the authors developed a technique called data mining that can be applied to the healthcare data, which in turn will detect the disorders. The researchers have classified their data with the help of Fuzzy classification and then it is classified again with the help of K-Nearest Neighbors. The classified data is then clustered through K-means. The proposed technique has demonstrated to be quite effective.

Research Gap – The technique discussed by the authors have shown a marked improvement but this has been on the same dataset. The authors plan to verify the technique for other datasets.

V. Krishnaiah [12] proposes that efficient prediction is possible in the field of healthcare with the effective use of Data Mining. As there have been countless implementations done in this field, there hasn't been much innovation in the field anymore. Heart diseases are one of the most dangerous conditions to have and require the utmost amount of care. Therefore, the authors have decided to implement a technique that classified the data utilizing the Euclidean distance fuzzy KNN. The fuzzy system is quite powerful and effective in this application.

I. Stancluesco [13] has observed the statistics about the premature babies, as they are kept in intensive, most of them tend to have a very late onset of neonatal sepsis. It is a very debilitating condition, especially for the baby, as it has a high fatality rate. The traditional technique for detection of the neonatal sepsis is very slow and not at all efficient. Therefore, the researchers present a technique for the detection of neonatal sepsis with the help of an Autoregressive Hidden Markov Model. The technique has been tested extensively and has provided exceptional results.

S. Huang [14] proposes an accelerator for the machine learning process in a Healthcare Application. As most of the time, machine learning is a very important and useful addition to the healthcare application, but most of the time it is very time consuming and inefficient. Therefore, the authors proposed a framework that can accelerate the machine learning process by implementing Support Vector Machine and Naïve Bayes Classifier to hasten the process of machine learning drastically. It has been tested against the traditional techniques which provided excellent results.

D. Chen [15] discusses that there is an increased amount of difficulty in working with data that has been censored, as the disease prognosis model creates a lot of problems when used with censored data. This difficulty increases with the increase in the heterogeneous prognosis and presentation. Therefore, the prognostic tools have to be resilient to the censored data to be able to function without failure. The authors have tested their implementation on the data for chronic lymphocytic leukemia for the resilience of their proposed framework on a heterogeneous Disease.

III. CONCLUSION

This research article has analyzed various works that have been published in the area of healthcare prediction and monitoring systems that are being used to help the elderly and the patients suffering from various critical ailments. There has been an increase in the use of machine learning for easing the process of healthcare and providing Medical Practitioners a helping hand. To increase the efficiency of the whole system and the workload on the doctors, the use of Hidden Markov Model has been proposed for the purpose of prediction of Healthcare.

REFERENCES

- [1] M. Healy and P. Walsh, "Detecting Demeanor for Healthcare with Machine Learning", IEEE International Conference on Bioinformatics and Biomedicine, 2017, pp. 2015-2019.
- [2] M. Krajnak, J. Xue, W. Kaiser, and W. Balloni, "Combining Machine Learning and Clinical Rules to Build an Algorithm for Predicting ICU Mortality Risk", Computing in Cardiology, 2012, pp. 401-404.
- [3] J. Yoo, "On-Chip Epilepsy Detection: Where Machine Learning Meets Patient-Specific Healthcare", International SoC Design Conference (ISOCC), 2017, pp. 146-147.
- [4] J. Alcaraz, S. Moghaddamnia, N. Poschadel, and J. Peissig, "Machine Learning as Digital Therapy Assessment for Mobile Gait Rehabilitation", IEEE International Workshop on Machine Learning for Signal Processing, Sept. 17-20, 2018, pp. 1-6.
- [5] J. Wojtusiak, "Semantic Data Types in Machine Learning from Healthcare Data", 11th International Conference on Machine Learning and Applications", 2012, pp. 197-202.
- [6] M. Mozaffari-Kermani, S. Sur-Kolay, A. Raghunathan and N. Jha, "Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare", IEEE Journal of Biomedical and Health Informatics, 2014, pp. 1893-1905.
- [7] A. Wongkoblap, M. Vadillo and V. Curcin, "Classifying Depressed Users with Multiple Instance Learning from Social Network Data", IEEE International Conference on Healthcare Informatics, 2018, pp. 436-436.
- [8] H. Sharma, C. Mao, Y. Zhang, H. Vatani, L. Yao, Y. Zhong, L. Rasmussen, G. Jiang, J. Pathak, and Y. Lao, "Portable Phenotyping System: A Portable Machine-Learning Approach to i2b2 Obesity Challenge", IEEE International Conference on Healthcare Informatics Workshop, 2018, pp. 86-87.
- [9] J. Park, K. Kim and O. Kwon, "Comparison of Machine Learning Algorithms to Predict Psychological Wellness Indices for Ubiquitous Healthcare System Design", International Conference on Innovative Design and Manufacturing, 2014, pp. 263-269.
- [10] A. Charleonnann, T. Fufang, T. Niyomwong, W. Chokchuepattanaki, S. Suwannawach, and N. Ninchawee, "Predictive Analytics for Chronic Kidney Disease Using Machine Learning Techniques", The 2016 Management and Innovation Technology International Conference, 2016, pp. 80-83.
- [11] N. Chetty, K. Vaisla and N. Patil, "An Improved Method for Disease Prediction using Fuzzy Approach", Second International Conference on Advances in Computing and Communication Engineering, 2015, pp. 568-572.
- [12] V. Krishnaiah, M. Srinivas, G. Narasimha, and S. Chandra, "Diagnosis of Heart Disease Patients Using Fuzzy Classification Technique", International Conference on Computing and Communication Technologies, 2015, pp. 1-7.
- [13] IoanStanculescu, Christopher K. I. Williams, and Yvonne Freer, "Autoregressive Hidden Markov Models for the Early Detection of Neonatal Sepsis", IEEE Journal of Biomedical and Health Informatics, VOL. 18, September 2014, pp.1560-1570.
- [14] S. Huang, Z. He, J. Han and X. Zeng, "A Flexible Low-Power Machine Learning Accelerator for Healthcare Applications", IEEE, 2016, pp. 613-615.
- [15] D. Chen, et al, "Predicting Time to First Treatment in Chronic Lymphocytic Leukemia using Machine Learning Survival and Classification Methods", IEEE International Conference on Healthcare Informatics, 2018, pp. 407-408.

