Interpretable Machine Learning and Artificial Intelligence for Sustainable Healthcare Monitoring

DR. A. JAYANTHILA DEVI[1], DR. P. S. AITHAL[2]

[1] *Professor & Post-Doctoral Research Fellow, Institute of Computer Science and Information, Srinivas University, Mangalore, India.*

[2] *Vice Chancellor, Srinivas University, Mangalore, India.*

Abstract- A machine learning model is said to be interpretable if it is able to provide an explanation as to why it made a particular prediction. The availability of traditional machine learning measurements such as area under the curve (AUC), precision (PR), and recall (R) may not be sufficient in many domains where user trust in the predictions of machine learning systems is necessary. The process of recognizing, comprehending, and responding to the human emotional responses that are included in written language is referred to as sentiment analysis. Using an innovative combination of natural language processing and sentiment analysis, data scientists have developed algorithms that are capable of inferring human emotion from text. These algorithms were designed using natural language processing. With their help, we were able to do this. Artificial intelligence (AI) systems have the potential to commit errors, which could have severe effects for patients or lead to other difficulties within the healthcare system. It is impossible to make assumptions about the advantages of developing technology, and there is always the risk that these technologies could have unforeseen impacts. However, it is also feasible to establish assumptions about the benefits of developing technology. It is vital for future study to address both the positive advancements as well as the problems that could have potentially detrimental effects.

Indexed Terms- AI, Machine Learning, Sentiment Analysis, Healthcare Monitoring

I. INTRODUCTION

The development of artificial intelligence (AI) technologies, particularly cutting-edge machine learning (ML), is becoming an increasingly crucial factor in the progression of medical advances [1]. Examples of this include the use of AI-led health chatbots for telemedicine and the development of intelligent assistive technologies for the care of the elderly and those with dementia; the application of deep learning and computer vision in radiology and dermatology; natural language processing approaches to mental health screening[2]; and so on. Other examples include the use of deep learning and computer vision in radiology and dermatology; the application of deep learning and computer vision in radiology and dermatology; and so on. Deep learning and computer vision have also found applications in other fields, such as radiology and dermatology, amongst other fields[3].

The applications of artificial intelligence (AI) in medicine encounter challenges in the form of societal barriers, legal barriers, and economic constraints [4]. However, these applications contain a significant amount of potential. A few of the difficulties that have been brought up include the reliability and accountability of predictive AI systems, the revolutionary influence that these systems will have on clinical decision-making and the dynamics between doctors and patients, and the dependability of AI-powered equipment. It has been suggested by researchers and developers of AI that if the user (the physician or the patient) trusted the AI, then the AI would be able to manage certain situations more efficiently [5- 8].

II. HEALTHCARE SYSTEM – BACKGROUND

When assessing whether or not it is appropriate, we will also investigate whether or not it is acceptable to inquire about explanations from machine learning systems and whether or not it is preferable to maintain an explanation-agnostic stance with regard to healthcare systems. As a result of this, a doctor may have a great deal of curiosity regarding the reason why a machine-learning system, as opposed to the diagnosticians at a hospital, is indicating a cancer diagnosis in a patient [9] [10]. It is normal for the planner working in the emergency department to find that the reasoning behind the hourly emergency department arrival forecasts generated by an automated learning system is of little importance to them. In the context of healthcare, we also study how the aforementioned criteria relate to the many different machine learning systems and algorithms that are now in use [11]. In order to investigate the limitations and drivers of using explainable machine learning algorithms in a variety of healthcare contexts, we make use of the findings from our research on the performance comparison of interpretable models on real-world problems such as risk of readmission prediction, ED utilization prediction, and hospital length of stay prediction. One of these challenges is estimating how long a patient will have to remain in the hospital [12]-[15].

Although methods of machine learning have been used for several decades, their recent proliferation into new fields such as healthcare has prompted a renewed interest in explaining how these systems operate. This interest has been prompted by the fact that machine learning methods have been used for several decades [16]. This interest has been sparked by the fact that methods of machine learning have been in use for several decades already. When practitioners and administrators in the healthcare business are trying to decide whether or not to make use of models, one of the most important factors they take into account is the interpretability of the model predictions. There is an urgent need for explanations of expected outcomes as machine learning algorithms are being gradually adopted in numerous locations along the continuum of care for patients [17] [18].

As a tool to improve patient outcomes, efficiency in the operating room, and financial outcomes, machine learning technologies are becoming an increasingly significant part of the healthcare industry. The processes of diagnosis, clinical care pathways, and patient risk stratification are just a few examples of the many areas that could potentially benefit from judgments that are based on predictions produced by machine learning [19]. Medical doctors and other experts are interested in the reason behind the forecast in order to make decisions that are of such a crucial nature. This course will examine a number of explanations that are utilized in machine learning and provide an in-depth grasp of the subtle distinctions that set each type apart from the others [20].

III. PROPOSED METHOD

We will discuss the myriad of facets that explainability encompasses in machine learning. Machine learning models, input data, model parameters, and techniques are some examples of these aspects; nevertheless, this list is not exhaustive. In addition to this, the individual who is using the system has a sizeable amount of say over the kind of explanation that is given to him.

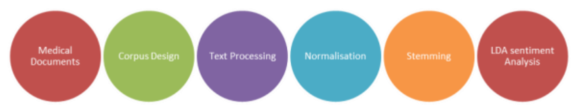

Figure 1: Proposed Modelling

As a consequence of this, a straightforward linear model that has highly developed and intricate features may be less interpretable than a deep learning model that contains features that are uncomplicated and easy to understand. We provide a methodology for evaluating interpretable machine learning systems that is predicated on a detailed literature assessment of themes that are pertinent to the area. This assessment was carried out by looking at previous research that has been done on these topics.

Then, we take this structure and combine it with a variety of different techniques to machine learning in order to solve a wide variety of concerns relating to medical care. In addition, we give a complete study of the limits and hazards associated with interpretable machine learning in the medical business from the point of view of an expert in the healthcare domain. This analysis is based on the perspective of a professional in the healthcare domain.

- LDA Based Topic Modelling

The results of this research indicate that the LDA model was utilized in order to carry out subject discoveries included within the tweets that were already in existence. The statistical generative model known as Latent Dirichlet Allocation (LDA) makes use of Dirichlet distributions as one of its key building blocks.

The authors start with a document pool of size D and determine in advance that there will be at least X subjects to be located in the pool. The outputs will not only include the topic model, but also documents of type D that are a combination of the X various subjects. When utilizing latent semantic analysis, one is able to determine the significance of the connections between documents and subjects, in addition to the connections between subjects and words. This is possible because latent semantic analysis focuses on the latent meanings of words. The application of linear discriminant analysis (LDA) has the benefit of making use of feature data in order to construct a new axis, which, in turn, reduces variability and raises the class distance.

The basic objective of LDA is to reduce the total number of dimensions that are required to project individual features. The three following steps are those that will be of assistance to you in achieving your objective:

Before we can move forward with anything else, the first thing we need to do is determine the average distance that separates two or more classes (also known as the separability between classes). The term inter-class variance is one that may be used to characterize this occurrence. It is possible to express this phenomenon using the equation that is presented below: (1).

$$S_b = \sum_{i=1}^{g} N_i * \left(\overline{x}_i - \overline{x}\right)\left(\overline{x}_i - \overline{x}\right)^T$$

The second thing you need to do is figure out how large of a gap there is between the mean of each group and the total number of people in the sample. The equation that best defines it can be found here, and it is also known as the variation that takes place within the group.

$$S_w = \sum_{i=1}^{g} \left(N_i - 1\right) * S_i$$

Where:
S_b = class seperability.
S_w = class distance.
N = instances.
$\bar{x}$ = Mean.

In the end, but certainly not least, we will need to build a lower-dimensional space that maximizes the diversity that exists across classes while simultaneously decreasing the variance that exists inside each class. When we talk about Fisher criteria, also known as P, what we mean is a projection of higher-dimensional space onto lower-dimensional space. This is what we mean when we talk about Fisher criteria.

For the aim of carrying out this research, we relied on the pyLDAvis package, which is an open-source Python framework for the visualization of dynamic topic models. This allowed us to get a clearer picture of the relationships between the variables.

IV. RESULTS AND DISCUSSIONS

The authors started out by doing a sentiment analysis using a novel approach. Because it takes into account the relative relevance of the words that are being used, this approach requires substantially less data cleansing than traditional methods do. This is just one of the numerous benefits that this method offers in addition to its many other advantages. After the research had been completed and analyzed, it was found that positive sentiment was the highest across all industries. This indicates that it will have a large effect, both favorably and dramatically, on international tourism over the course of the next few years. In addition to this, it has the ability to reduce our impact on the environment and make the planet we live on more habitable and sustainable.

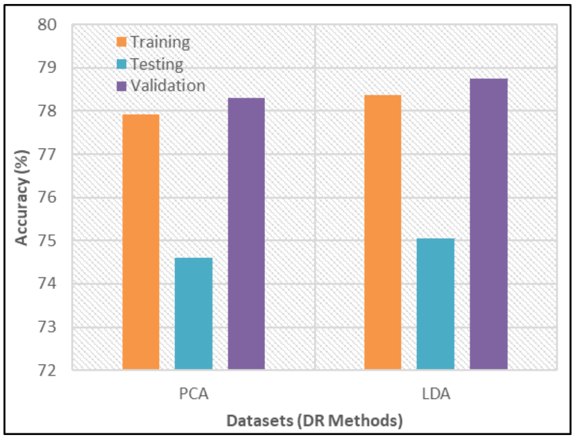

Figure 2: Accuracy of LDA

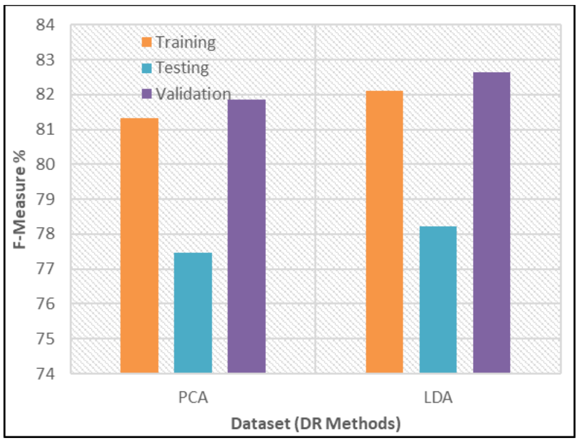

Figure 3: F-Measure of LDA

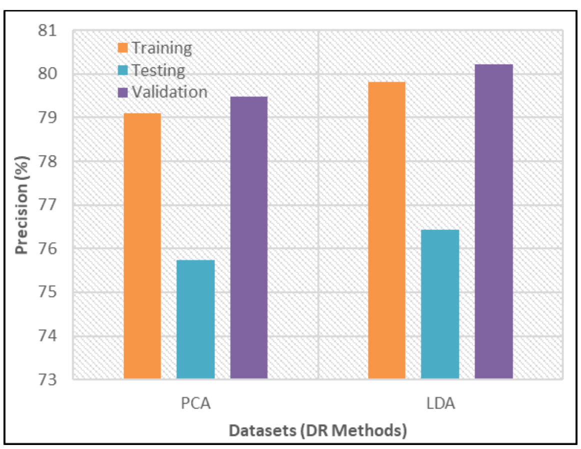

Figure 4: Precision of LDA

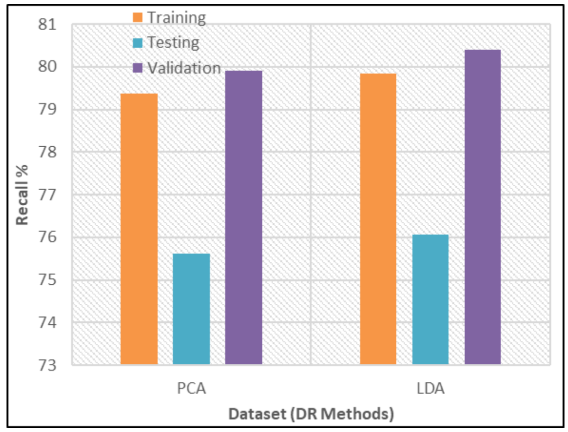

Figure 5: Recall of LDA

The utilization of topic modeling was yet another method that was put into action in order to locate patterns. These patterns included the recurrence of particular phrases as well as the extent to which they differed from one another in degree. A subject model categorizes comments and remarks that are made frequently into groups that are analogous to one another and puts the results into a hierarchy. When you have access to all of this material, it is quite simple to identify the topics that are covered in the various collections of works since you can easily make connections between them. The LDA paradigm analyzes each text as though it were a collection of topics taken from the corpus. This approach is called topic modeling. It is possible to link the meaning of each word in the text back to a specific category contained within the model. A cutting-edge deep learning algorithm was used to build a sophisticated model for predicting sentiment, and that model was then evaluated using a variety of various epoch sizes. This brought the total number of steps in this process up to five.

CONCLUSION

This research establishes the framework for a long-term paradigm that is practically possible and is based on an analysis of deep learning as it applies to healthcare. The research is based on an examination of how deep learning can be implemented. Utilizing an LDA provides a number of benefits, one of the most significant of which is the fact that it can be simply understood and maintained by people. In this way, the semantic orientation may be gathered, and then it can be placed into one of three categories: neutral, positive, or negative. In contrast to sentiment analysis, which extracts subjectivity and polarity from a body of text, semantic orientation evaluates the force and polarity of the information being presented. The power and polarity of the content are evaluated using the concept of semantic orientation. In this method, adjectives and adverbs are employed to emphasize certain facets of the text overall meaning in order to direct the reader attention to them (in our case, a tweet). The value of the sentiment orientation is then determined by using the combination of the adverbs and adjectives. In addition to this, everything of value can be traced back to a single origin. When undertaking a study of the general emotion expressed in the tweets, we are able to arrive at an average by making use of a technique that is known as pooling.

REFERENCES

[1] Panesar, A. (2019). *Machine learning and AI for healthcare* (pp. 1-73). Coventry, UK: Apress.

[2] Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, *27*(3), 491-497.

[3] Nordling, L. (2019). A fairer way forward for AI in health care. *Nature*, *573*(7775), S103-S103.

[4] Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: a mapping review. *Social Science & Medicine*, *260*, 113172.

[5] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31-38.

[6] Wang, F., & Preininger, A. (2019). AI in health: state of the art, challenges, and future directions. *Yearbook of medical informatics*, *28*(01), 016-026.

[7] Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *Journal of global health*, *8*(2).

[8] Kaushik, S., Choudhury, A., Sheron, P. K., Dasgupta, N., Natarajan, S., Pickett, L. A., & Dutt, V. (2020). AI in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in big data*, *3*, 4.

[9] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6*(2), 94.

[10] Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, *5*(6), 493-497.

[11] Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., ... & Akour, I. A. (2021). IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet*, *13*(8), 218.

[12] Ahmad, M. A., Eckert, C., & Teredesai, A. (2018, August). Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics* (pp. 559-560).

[13] Manne, R., & Kantheti, S. C. (2021). Application of artificial intelligence in healthcare: chances and challenges. *Current Journal of Applied Science and Technology*, *40*(6), 78-89.

[14] Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2021). Impact of machine learning on management, healthcare and agriculture. *Materials Today: Proceedings*.

[15] van Der Schaar, M., Alaa, A. M., Floto, A., Gimson, A., Scholtes, S., Wood, A., ... & Ercole, A. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Machine Learning*, *110*(1), 1-14.

[16] Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: addressing ethical challenges. *PLoS medicine*, *15*(11), e1002689.

[17] Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, *20*(5), e262-e273.

[18] Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, *395*(10236), 1579-1586.

[19] Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, *112*(1), 22-28.

[20] Iliashenko, O., Bikkulova, Z., & Dubgorn, A. (2019). Opportunities and challenges of artificial intelligence in healthcare. In *E3S Web of Conferences* (Vol. 110, p. 02028). EDP Sciences.