Literature Review:

Implementing Machine Learning tools and/or techniques in Medical Diagnosis

The medical industry is an innovative, constantly evolving sector with advancements in technology providing continuous improvements to diagnosis and the treatment of patients. Many of the key technological advancements of the past decade are in the field of Artificial Intelligence and more specifically, machine learning. Naturally, these advancements are of great significance to the innovation of the medical industry. This literature review will investigate the innovations of Machine learning in medical diagnosis, the increase in utilisation and adoption of the technology and identify potential limitations.

Early detection and accurate diagnosis of diseases such as cancer can significantly increase the chances of full recovery. It not only increases the chances of survival but also reduces morbidity and the overall costs of treatment (Etzioni et al., 2003). At a time of increasing pressure on the NHS and with waiting times for hospital treatments in England at the highest level since 2007 (Baker, 2023), the quick, accurate diagnosis of disease is crucial to reduce the strain and budgetary pressures on the NHS.

The use of artificial intelligence can be seen in many common applications in our day to day lives, from chatbots and virtual assistants, to customer profiling and predictive modelling. Machine learning is a branch of Artificial Intelligence in which a computer is programmed to use vast amounts of sample data to create its own algorithms to solve future problems (Alpaydin, 2020). According to Naqa et al. (2015) machine learning is;

> *'an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment.'*

Many of the largest technology companies in the world such as Google and Netflix take advantage of machine learning to increase profits and minimize risks (Portugal, 2018) which fuels the innovation and advancement of these tools. However, this technology can be utilised in a vast range of use cases and industries. Machine learning can be used to identify

trends and patterns in any application requiring the use of big data. Given the vast amount of patient data collected and stored in patient records and databases, there is an abundance of sample data to train machine learning tools to identify trends which would be impractical to find otherwise.

The use of machine learning tools to assist in diagnosis is not a new trend. According to Kononenko (2001), machine learning was used as early as the 90's in oncology, liver pathology, cardiology, and many other areas of medicine. While this paper is dated in terms of the algorithms it discusses, particularly given the significant advancements of technology over the last 20 years, it is useful to understand the historical context of machine learning in the medical field. Interestingly, this paper recognises the slow acceptance of machine learning technology and the underutilisation of this new technology at the time. One of the reasons it provides for this is that machine learning would add to the complexity of the already complex job of a physician.

The increasing popularity of machine learning techniques is illustrated by Foster et al. (2014) who observed a significant volume of academic papers relating to machine learning in biomedicine at the time. Focusing only on support vector machine (SVM), a type of supervised learning model, they found a wide range of research relating to, as an example, ECG analysis. This can be attributed to the wide availability of machine learning software packages and easy access to vast amounts of ECG data allowing many to experiment with limited experience or training in either medicine or computer science. This paper in particular questions the competence and reliability of some of these tools and studies.

Bhavsar et al. (2021) discuss the wide range of machine learning methods in use today such as Artificial Neural network (ANN), Bayesian Classifier (BC), Random forest (RF) and Support Vector machine (SVM) among others. They also provide a thorough overview of the wide range of medical diagnostic use cases across a wide range of disciplines. The main advantages they identify are the increase in accuracy of disease diagnosis, reducing costs while supporting doctors by assisting them with decision making and providing a second

opinion. This paper evidences the significant increase in utilisation of machine learning models in real life medical applications.

Guncar et al. (2018) further illustrates the value that machine learning can have to medical diagnosis. They created 2 Random Forest (RF) predictive models to detect changes to blood parameter values. Not only did their models diagnose diseases with a degree of accuracy at a similar level to haematology specialists and significantly higher than general specialists, but they also detected changes which would have been overlooked with traditional quantitative techniques. Furthermore, their models made accurate diagnosis with less blood parameter data points than would normally be required with a similar accuracy. The findings of this study could be of significant value to the development of blood analysis techniques.

Another field where machine learning diagnosis models have been shown to be successful is in the diagnosis of heart disease. Terrada et al. (2020) used Artificial Neural Network (ANN), Adaboost and DT algorithms to create a predictive model with an accuracy of up to 94% in predicting coronary artery disease, higher than comparable non-machine learning techniques. Ahmad et al (2022) achieved even more accurate results of up to 100% accuracy using a combination of Extreme Gradient Boosting classifier (GBC) with GridSearchCV across some of the same databases as the previous paper. This illustrates the continuous advancements in the field and the potential for even more accuracy as the models are improved and improved machine learning software is designed.

As with all technology, there are limitations to using machine learning technology in the healthcare industry. As discussed previously, Foster et al. (2014) discovered a significant number of low-quality models created by researchers with unknown training or experience. The low barrier of entry due to the widespread availability of machine learning tools and sample data means that the number of models, of varying quality, for each discipline is significant. Identifying the highest quality and most useful models is a massive task requiring a cross disciplinary team of medical professionals and computer scientists. With the current restraints and pressures on the healthcare industry, finding the resources to thoroughly

investigate and introduce new techniques is difficult. Furthermore, machine learning models are only as good as the sample data provided to them. Jarrett et al. (2019) note that models are trained based on previous expert opinions, assuming this data as the truth, however they may struggle when problems are not well defined and as such may be of limited use with conceptual clinical challenges where specialist doctors may continue to make better judgements. As Kononenko (2001) discussed, clinical staff are not always accepting of new technologies and techniques, which may be a barrier to widespread use of these models.

In conclusion, while the advancements of machine learning technology have been significant, and the use cases within the healthcare field have continuously grown since the 90's, many of the potential use cases are experimental and have not yet reached widespread use. From the findings of this literature review, I believe future research would be worthwhile into the current attitudes of clinical staff and patients into the use of Artificial Intelligence in the healthcare industry and the likelihood of widespread acceptance of these new technologies and techniques in the future. As advancements in technology have progressed over the past 30 years and more of the models are being utilised, research should be made into how well these models have been received over time to identify any barriers to further widespread use in the future. Machine learning is a valuable tool for the industry and overcoming barriers to get these models into clinical use would have value to patients, doctors and the industry as a whole.

References

Ahmad, G., Fatima, H., Ullah, S., Saidi, A. & Imdadullah (2022) Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV. *IEEE Access* 10:80151-80173

Alpaydin, E. (2020) *Introduction to machine learning.* 4th ed. The MIT press.

Baker C. (2023) NHS Key Statistics: England, July 2023. Available from: https://researchbriefings.files.parliament.uk/documents/CBP-7281/CBP-7281.pdf [Accessed 24/09/23]

Bhavsar, K., Abugabah, A., Singla, J., AlZubi, A., Bashir, A., Nikita (2021) A comprehensive review on medical diagnosis using machine learning. *Computers, Materials and Continua.* 67(2): 1997-2014.

El Naqa, I. & Murphy, M. (2015). *Machine Learning in Radiation Oncology.* Springer, Cham. Available from: https://link.springer.com/chapter/10.1007/978-3-319-18305-3_1#citeas [Accessed 25/09/23]

Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., Radich, J., Anderson, G. & Hartwell, L. (2003) The case for early detection. *Nature Reviews Cancer* 3: 243–252

Foster, K., Koprowski, R. & Skufca, J. (2014) Machine learning, medical diagnosis, and biomedical engineering research - commentary. *BioMed Eng OnLine* 13:94.

Gunčar, G., Kukar, M., Notar, M., Brvar, M., Cernelc, P., Notar, M. & Notar, M. *(2018)* An application of machine learning to haematological diagnosis. *Scientific Reports* 8:411

Jarrett, D., Stride, E., Vallis, K. & Gooding, M. (2019) Applications and limitations of machine learning in radiation oncology. *The British Journal of Oncology.* 92(1100):

Kononenko, I. (2001) Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial intelligence in Medicine.* 23(1):89-109.

Portugal, I., Alencar, P., & Cowan, D. (2018) The use of machine learning algorithms in Recommender Systems: A systemic review. *Expert systems with Applications* 97: 205-227.

Terrada, O., Hamida, S., Cherradi, B., Raihani, A. & Bouattane, O. (2020) Supervised machine learning based medical diagnosis support system for prediction of patients with Heart disease. *Advances in Science, Technology and Engineering Systems Journal 5(5): 269-277*