A Novel Platform For Disease Detection Using Machine Learning

Asfahan Shah, Aniket Purkayastha, Karthikeyan Rathore, Anrup Sharma

I. Introduction

INtoday's world of medical science, the detection rate of diseases such as heart diseases, brain tumor etc, has drastically increased; along with this the errors in detecting the diseases has become more frequent by ever by radiologists/doctors [1]. Keeping this in mind, the use of AI in this field becomes imperative. Let us see an example: In a study where the efficacy of AI algorithm was judged in detecting breast cancer was to determined, datasets from 5 institutions in USA, South Korea [2]. The AI algorithm and 14 radiologists participated in this study to examine the mammography examinations. The AI algorithm's performance on was more than 94%and that of the radiologists a was 82%. The AI algorithm developed with large-scale mammography data showed better diagnostic performance in breast cancer detection compared with radiologists.

Another reason to use AI in the medical field is The exhaustion and overwork felt by many medical professionals is impacting on their performance. Long hours, heavy workloads, and a lack of support are all contributing factors. Physicians make difficult and life-changing decisions on a daily basis, yet they frequently lack the space or time to efficiently manage their duties. This is where AI comes as a solution. The Ai algorithm will be able to process large quantities of images thus decreasing the workload on the health workers and the administration in preliminary detections. There have been many different ML models which are used to predict the disease, at a preliminary stage, from a given dataset, but those models majorly are utilized in individual disease prediction. In our project, we intend to consolidate various ML models into one platform so that the user can choose from the available diseases and detect its presence.

In our project, we aim to develop a platform where we can process the preliminary prediction of diseases such as heart disease, pneumonia, Diabetes.

The human heart is the most important organ in the body. In a nutshell, it controls and regulates the flow of blood throughout our bodies. Any irregularity in the functioning of the heart might lead to obstacles in the functioning of the other parts of the body. Heart disease is defined as any interruption in the normal functioning of the heart. In today's society, heart disease is one of the most common causes of death. Heart disease can be caused by an unhealthy lifestyle, such as smoking, alcohol and high intake of fat which may cause hypertension. The rate of cardiovascular disease in India is one of the highest in the world. The yearly number of casualties from heart diseases is projected to drastically increase from 2.26 million (1990) to 4.7 million (2020).

Pneumonia is a respiratory distress infection caused due to bacteria, viruses, or fungi. It leads to inflammation of the air sacs in the infected lungs as well as fills the lungs with fluid. 15% of all fatalities in children under the age of five years are caused due to pneumonia. It is particularly frequent in poor and impoverished nations, wherein overpopulation, congestion, pollution, and unsanitary environmental factors aggravate the problem, and health services are few. As a result, early detection and treatment can help prevent the disease from progressing to a deadly stage. The use of magnetic resonance imaging (MRI), computed tomography (CT), or radiography (X-rays) to examine the lungs for diagnosis of pneumonic conditions is common. The lungs can be examined by X-ray imaging, which is a non-invasive and reasonably an affordable procedure.

The WHO says – "Each year pneumonia has resulted in fatalities of around 1.4 million children who are under the age of 5 years". The main data and resources on the basis of which pneumonia detections can be done include chest X-rays, CT of the lungs, ultrasound of the chest and MRI of the chest. Infiltrates, or white patches on the pneumonic X-ray , identify a person being in pneumonic state from a healthy state. Chest X-ray scans for pneumonia diagnosis, on the other hand, are vulnerable to individual variation [2, 3]. As a result, an automated approach for detecting pneumonia is needed.

Diabetes is one of the most prevalent ailments on the earth. Diabetes is induced usually by factors like obesity, high blood glucose level etc. Diabetes alters the hormone insulin, causing aberrant carbohydrate metabolism and improving blood sugar levels. When the body does not produce enough insulin, diabetes onsets. In 2014, the World Health Organization (WHO) had projected that 8.5 percent of individuals over the age of 18 had diabetes worldwide. In India's adult population, diabetes is expected to affect 72.96 million people. The frequency in urban regions varies from 10.9 percent to 14.2 percent, whereas the prevalence in rural India was 3.0-7.8% among 20 years and above individuals, with a significantly greater frequency among those who were above 50 years of age. WHO also states that there are approximately 422 million people worldwide who suffer with diabetes, with the majority of them living in low- and middle-income nations.

Diabetes is prevalent in a variety of countries, including Canada, China, and India etc. The exponential growth in India's population has resulted in an increase of a total 40 million diabetics. Diabetes is the world's leading cause of death. Diabetes may be managed and a human life saved if diagnosed early. To achieve this goal, diabetes prediction using

a variety of diabetes-related characteristics is essential.

Many regions around the globe are faced with a shortage of experienced health workers and radiologists who carry out the required tests in order to detect the disease. This is where the demand of AI solutions in preliminary detection is increasing. Data on numerous health-related issues is acquired by medical groups all around the world. This data may be used to acquire useful insights using a variety of machine learning approaches. However, the amount of data collected is enormous and noisy as well. Machine learning algorithms can readily examine these datasets, which are too large for us humans to process. As a result, these algorithms have recently proven to be quite effective in precisely predicting the existence or absence of illnesses.

II. Literature Review

A.singh et.al in their study used machine learning algorithms for prediction of heart diseases [3]. The machine learning algorithms that were used in the study were k-nearest neighbor, decision tree, linear regression, and support vector machine (SVM). The models were trained and tested using UCI repository dataset the study concluded that the KNN model performed better as compared to the rest of the model when the confusion matrix was used as a matric.

V. V. Vijayan et.al. proposed a system that uses AdaBoost algorithm with Decision Stump as base classifier for diabetes classification [4]. Additionally other algorithms like Support Vector Machine, Naive Bayes and Decision Tree were also used as base classifiers for AdaBoost algorithm in this study. It was observed that accuracy obtained by AdaBoost algorithm with decision stump as base classifier was 80.72% which was greater as compared to that of Support Vector Machine, Naive Bayes and Decision Tree.

K. K. Joshi et.al provided a review of the machine learning technique that are used in medical diagnosis [5]. In the study a survey of various diseases was conducted along with the different machine learning solutions and models was conducted. Various datasets were used in the study. The goal of the study was to give a brief review of various machine learning techniques and various advancements that can be used in medical diagnosis.

Mirbabaie et.al conducted a critical review of ai landscape in medical diagnosis and gave a short summarization of future advancements that are currently going in this filed [6]. In the study the authors classified the literature according to various metrics such as, algorithms, dataset characteristics, and performance measurements which is useful for medical practitioner and medical researchers. The goal of the study is to provide a analysis of application of AI in medical diagnosis and seeks to guide the researchers for the future research in this field of AI.

J. Garstka et.al conducted a study in which they presented a self-constructed convolutional neural network which was trained on a small dataset for classification of lung X-ray images [7]. The classification was done into three categories: healthy, bacterial pneumonia, and viral pneumonia. The accuracy of the CNN model was up-to 85% while the sensitivity was 0.95.

L. Bishnoi et.al conducted a review of Artificial intelligence techniques that are used in medicine [8]. The purpose of the review was to focus and study AI techniques that are commonly used in the field of medical science. It was found that artificial neural networks were commonly used followed by Support vector machine and fuzzy logic. This observation was attributed to the fact that these techniques had very high efficiency rate in medicine. Another conclusion that was found in the study was pattern recognition techniques and deep learning techniques had a gradual increase in use from the last decade as these techniques can help in prediction of diseases before occurring.

Kim et.al conducted a study in which they developed an AI algorithm for cancer detection which the help of mammography data [2]. This data was collected from five institutions In South Korea, USA, and UK. The study aims to compare the performance of AI and radiologists. The performance was evaluated in terms of area under ROC curve and recall and specificity. The study was mainly specific for breast cancer detection. The authors concluded that AI model gave better performance and AI outperformed the radiologist in detection of early stage of invasive cancers. It was also shown that AI detected more cancers with asymmetry and distortions in architecture than the radiologist. The authors also suggested that these results signify that artificial intelligence that can significantly improve that diagnosis. Overall, the study and the authors concluded that the diagnostic results of radiologist improved the help of AI thus it shows that AI can be sued as a good diagnostic support tool for breast cancer detection.

Liu et.al conducted a systematic review and a meta-analysis on comparing the accuracy of diagnosis done by health care professionals against popular deep learning models [9]. The author stated various challenges and advantages that these models have. Though the study concluded that in terms of sensitivity and specificity both the, health care professionals and models had equivalent performance, the author state that deep learning models can be useful in various other aspects in field of medical diagnosis. Besides that, authors also stated various challenges like availability of data, transparency etc. that the models need to overcome so that their performance and overall usage percentage can increase.

A. Šećkanović et.al conducted a review for the use of AI techniques in cardiovascular diseases [10]. The authors used various papers which were published in Medline database for the study. The study found that AI based models had a performance of above 83% for these cardiovascular diseases. Authors conclude that these results are hopeful as these would encourage further development in AI and ML models for this disease.

M. Sorić et.al conducted a study in which they demonstrated use of CNN for pneumonia classification [11]. In this study, they build convolutional neural network (CNN) model using a supervised dataset. This dataset was divided into two classes, pneumonia, and non-pneumonia images. The study concluded that CNN model is satisfactorily good for pneumonia detection as the model achieved good classification metrics. The model achieved accuracy of 90.38%, 98.21% recall and 87.84% precision

N. D. Parbate et.al conducted a study that mainly focuses on Machine learning Algorithms to check whether a particular patient infected with COVID-19 will require ventilation or not [12].The methods used here are XG-Boost, Random Forest, CART etc. It was found in the study that ML models gave good results and therefore it can be used to analyze the risk in patients. The study shows promising results towards the use of AI and ML in the field of health care.

Sonu Subudhi et.al used the patient data collected from Mass General Brigham (MGB) Healthcare database regarding COVID-19 [13]. The authors developed the models and validated them using the Emergency Department data. The study showed that ensemble based models performed better than the other ML models. The models predicted the mortality as well as ICU admission. The study demonstrated the potential of machine learning in the health care and demonstrated that such models can be useful for future decision making in case of outbreaks like COVID-19.

III. Diagrams

We have demonstrated the flow of the system using block diagram 1. The block diagram shows overall working of the system.As demonstrated in the diagram we will have a front-end that would send request calls to our API. Upon receiving the request calls API will forwards these calls to the appropriate Machine Learning Model.The Model will predict the result and forward that to the API after which API will show the result in the front-end.

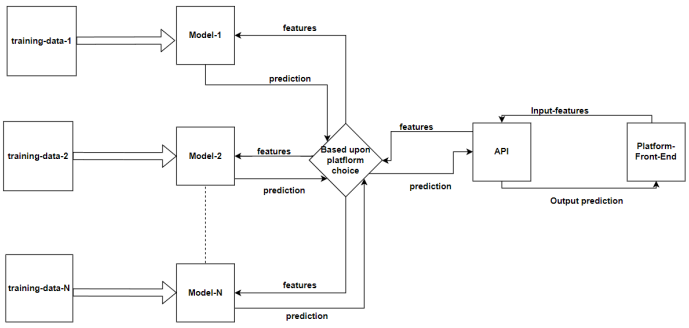

Fig. 1. Block Diagram.

As our system will consists of multiple models so in figure 2 we have shown the flow chart of each model of the system. Each Model in the system will first get checked whether the problem is supervised or not. In our case most of the problems are supervised in nature. After this the model will get trained upon the appropriate data then it will get evaluated for further optimizations. After being satisfied with the results the model will be deployed.

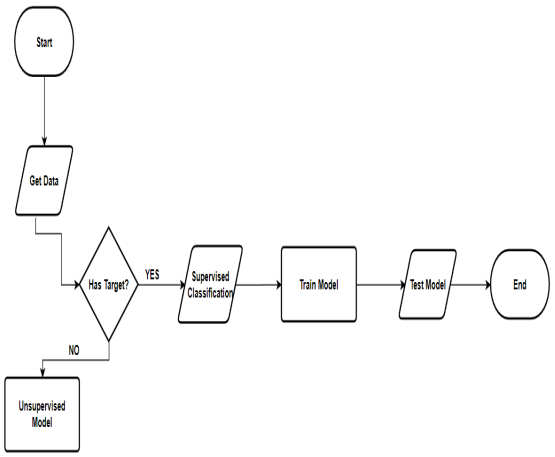

Fig. 2. Flow Diagram.

References

[1] M. D. Huffman, D. Prabhakaran, C. Osmond, C. H. Fall, N. Tandon, R. Lakshmy, S. Ramji, A. Khalil, T. Gera, P. Prabhakaran, S. D. Biswas, K. S. Reddy, S. K. Bhargava, and H. S. Sachdev, "Incidence of cardiovascular risk factors in an indian urban cohort," *Journal of the American College of Cardiology*, vol. 57, no. 17, pp. 1765–1774, Apr. 2011. [Online]. Available: https://doi.org/10.1016/j.jacc.2010.09.083

[2] H.-E. Kim, H. H. Kim, B.-K. Han, K. H. Kim, K. Han, H. Nam, E. H. Lee, and E.-K. Kim, "Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study," *The Lancet Digital Health*, vol. 2, no. 3, pp. e138–e148, Mar. 2020. [Online]. Available: https://doi.org/10.1016/s2589-7500(20)30003-0

[3] A. Singh and R. Kumar, "Heart disease prediction using machine learning algorithms," in *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 2020, pp. 452–457.

[4] V. V. Vijayan and C. Anjali, "Prediction and diagnosis of diabetes mellitus — a machine learning approach," in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2015, pp. 122–127.

[5] K. K. Joshi, K. K. Gupta, and J. Agrawal, "A review on application of machine learning in medical diagnosis," in *2nd International Conference on Data, Engineering and Applications (IDEA)*, 2020, pp. 1–6.

[6] M. Mirbabaie, S. Stieglitz, and N. R. J. Frick, "Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction," *Health and Technology*, vol. 11, no. 4, pp. 693–731, May 2021. [Online]. Available: https://doi.org/10.1007/s12553-021-00555-5

[7] J. Garstka and M. Strzelecki, "Pneumonia detection in x-ray chest images based on convolutional neural networks and data augmentation methods," in *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2020, pp. 18–23.

[8] L. Bishnoi and S. Narayan Singh, "Artificial intelligence techniques used in medical sciences: A review," in *2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 2018, pp. 1–8.

[9] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J. R. Ledsam, M. K. Schmid, K. Balaskas, E. J. Topol, L. M. Bachmann, P. A. Keane, and A. K. Denniston, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The Lancet Digital Health*, vol. 1, no. 6, pp. e271–e297, Oct. 2019. [Online]. Available: https://doi.org/10.1016/s2589-7500(19)30123-2

[10] A. Šećkanović, M. Šehovac, L. Spahić, I. Ramić, N. Mamatnazarova, L. G. Poković, A. Badnjević, and M. Kacila, "Review of artificial intelligence application in cardiology," in *2020 9th Mediterranean Conference on Embedded Computing (MECO)*, 2020, pp. 1–5.

[11] M. Sorić, D. Pongrac, and I. Inza, "Using convolutional neural network for chest x-ray image classification," in *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2020, pp. 1771–1776.

[12] N. D. Parbate, A. P. Palande, S. K. Wagh, A. Kamble, and S. Isave, "Icu admission prediction using machine learning for covid-19 patients," in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, 2021, pp. 1–6.

[13] S. Subudhi, A. Verma, A. B. Patel, C. C. Hardin, M. J. Khandekar, H. Lee, D. McEvoy, T. Stylianopoulos, L. L. Munn, S. Dutta, and R. K. Jain, "Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19," *npj Digital Medicine*, vol. 4, no. 1, May 2021. [Online]. Available: https://doi.org/10.1038/s41746-021-00456-x