Cancer Detection Using Machine Learning

Vismay Patel[1], Aniket Tripathi[2], Prof.Ajayakumar T. Shah[3]
[1, 2] Dept of Computer Engineering
[3]HOD, Dept of Computer Engineering
[1, 2, 3] Alpha College of Engineering and Technology

Abstract- *In the proposed product users will be able to upload histology images. Histology images are also known as microanatomy of biological tissues. Histology is the microscopic counterpart to gross anatomy, which looks at larger structures visible without a microscope our machine learning algorithm will predict and tell you the probability of a having cancer in an uploaded image. Users will also be able to upload multiple images divided into two or more classes and train their own machine learning algorithm that they can use it their own product or project.*

Keywords- Cancer, Breast cancer, Machine Learning, prediction, risk, Artificial Intelligence, ANN, Artificial Neural Network, Early Stopping algorithm

I. INTRODUCTION

Machine learning is a branch of artificial intelligence which contains modules of data science, autonomous driving, fraud detection, image detection, image recognition and many more applications of machine learning are available.

The increasingly number of applications of machine learning in healthcare allows us to glimpse at a future where data, analysis, and innovation work hand-in-hand to help countless patients without ever realizing it. Soon, it will be quite common to find ML-based applications embedded with real-time patient data available from different healthcare systems in multiple countries, thereby, increasing the efficacy of new treatment options which were unavailable before.

Here are some examples of ML-based application in Healthcare: Identifying Diseases and Diagnosis, Drug Discovery and Manufacturing, Medical Imaging Diagnosis, Personalized Medicine, Machine Learning-based Behavioral Modification, Smart Health records, Clinical Trial and Research, Better Radiotherapy, Crowdsourced Data collection, Outbreak Prediction.

Our product focuses on Medical Imaging Diagnosis in which you take an image and make some prediction using machine learning model. This model is trained on millions of images from past patients and it continues to improve as time goes by.

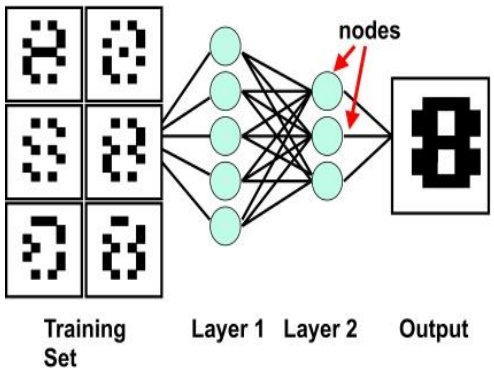

Fig 1: How a machine learner is trained to recognize images using a training set

II. LITERATURE REVIEW

A. Machine learning is a branch of artificial intelligence research that employs a variety of statistical, probabilistic and optimization tools to "learn" from past examples and to then use that prior training to classify new data, identify new patterns or predict novel trends. Machine learning, like statistics, is used to analyze and interpret data. Unlike statistics, though, machine learning methods can employ Boolean logic (AND, OR, NOT), absolute conditionality (IF, THEN, ELSE), conditional probabilities (the probability of X given Y) and unconventional optimization strategies to model data or classify patterns. These latter methods actually resemble the approaches humans typically use to learn and classify.

B. At present Mammograms are commonly used to screen for breast cancer. If an abnormality is detected on a screening mammogram, your doctor may recommend a diagnostic mammogram to further evaluate that abnormality. Breast ultrasound uses sound to waves to produce images of structures deep within the body.

C. Artificial Neural networks (ANN) or neural networks are computational algorithms. It intended to simulate the behaviour of biological systems composed of "neurons". ANNs are computational models inspired by an animal's central nervous systems. It is capable of machine learning as well as pattern recognition. These presented as systems

of interconnected "neurons" which can compute values from inputs.

D. A neural network is an oriented graph. It consists of nodes which in the biological analogy represent neurons, connected by arcs. It corresponds to dendrites and synapses. Each arc associated with a weight while at each node. Apply the values received as input by the node and define Activation function along the incoming arcs, adjusted by the weights of the arcs.

E. From our analysis of the literature several trends were noted. As has been remarked previously, the use of machine learning in cancer prediction and prognosis is growing rapidly, with the number of papers increasing by 25% per year. While it is clear that machine learning applications in cancer prediction and prognosis are growing, so too is the use of standard statistically-based predictive methods. In particular, we looked at the frequency with which "cancer prediction prognosis methods" and "cancer risk assessment prediction methods" occurred in PubMed.

F. When looking at the types of predictions or prognoses being made, the vast majority (86%) are associated with predicting cancer mortality (44%) and cancer recurrence (42%). However, a growing number of more recent studies are now aimed at predicting the occurrence of cancer or the risk factors associated with developing cancer. As a general rule, regardless of the machine learning method used, the type of prediction being made or the type of cancer being evaluated, machine learning methods appear to improve the accuracy of predictions by an average of 15–25% over alternative or conventional approaches

G. There is strong bias among scientists to use machine learning towards predicting outcomes or risks associated with breast (24%) and prostate (20%) cancer. This, no doubt, reflects the higher frequency of these cancers among patients in Europe and North America. Nevertheless, machine learning methods appear to have been successfully used in predicting outcomes or risks in nearly a dozen different kinds of cancer. This suggests that machine learning methods can be quite generally applied to cancer prediction and prognosis.

H. Almost 70% of all reported studies use neural networks as their primary (and sometimes only) predictor. Support vector machines are a distant second with 9%, while clustering and decision trees each account for about 6%. Genetic algorithms and other methods (naïve Bayes, fuzzy logic) are rarely used (Table 2). This is both surprising and a bit disappointing. ANNs are relatively old machine learning technologies which yield so-called "black-box" results. That is, their performance and classification processes are not easily explained or rationalized.

III. STUDY FINDINGS

A. In recent years, the availability of large datasets combined with the improvement in algorithms and the exponential growth in computing power led to an unparalleled surge of interest in the topic of machine learning. Nowadays, machine learning algorithms are successfully employed for classification, regression, clustering, or dimensionality reduction tasks of large sets of especially high-dimensional input data. In fact, machine learning has proved to have superhuman abilities in numerous fields (such as playing go, self-driving cars, image classification etc.). As a result, huge parts of our daily life, for example, image and speech recognition, web-searches, fraud detection, email/spam filtering, credit scores, and many more are powered by machine learning algorithms.

B. Breast cancer is the second leading cause of cancer deaths among U.S. women and screening mammography has been found to reduce mortality. Despite the benefits, screening mammography is associated with a high risk of false positives as well as false negatives. The average sensitivity of digital screening mammography in the U.S. is 86.9% and the average specificity is 88.9%. To help radiologists improve the predictive accuracy of screening mammography, computer-assisted detection and diagnosis (CAD) software4 have been developed and in clinical use since the 1990s. Unfortunately, data suggested that early commercial CAD systems had not led to significant improvement in performance and progress stagnated for more than a decade since they were introduced. With the remarkable success of deep learning in visual object recognition and detection, and many other domains, there is much interest in developing deep learning tools to assist radiologists and improve the accuracy of screening mammography. Recent studies have shown that a deep learning-based CAD system performed as well as radiologists in standalone mode and improved the radiologists' performance in support mode.

C. To perform classification or segmentation on large complex images, a common strategy involves the use of a classifier in sliding window fashion to recognize local patches on an image to generate a grid of probabilistic outputs. This is followed by another process to summarize

the patch classifier's outputs to give the final classification or segmentation result. Such methods have been used to detect metastatic breast cancer using whole slide images of sentinel lymph node biopsies34 and to segment neuronal membranes in microscopic images35. However, this strategy requires two steps that each needs to be optimized separately

IV. CONCLUSTION

The proposed system is built on flask framework which is built in python language. It provides user ability to upload histology images. Our website will be able to take image and evaluate the image through trained deep learning model. Our model is built using

Artificial neural networks. Neural networks are composed of layers of computational units (neurons), with connections among the neurons in different layers. These networks transform data – like the pixels in an image or the words in a document – until they can classify it as an output, such as naming an object in an image or tagging unstructured text data. This model is optimized to get maximum accuracy and to minimize false-positives and false-negatives.

V. ACKNOWLEDGMENT

We express our sincere thanks to Prof. Ajaykumar T. Shah Head of Department of Computer Engineering, Alpha College of Engineering and Technology for their Support and guidance for this project and care taken by them in helping us to complete the project work successfully.

REFERENCES

[1] Scientific Reports: https://www.nature.com/articles/s41598-019-48995-4
[2] TechTarget: https://www.nature.com/articles/s41598-019-48995-4
[3] Machine Learning model:https://tfhub.dev/google/tf2-preview/mobilenet_v2/feature_vector/4
[4] Applications of Machine Learning: https://journals.sagepub.com/doi/full/10.1177/117693510600200030