Machine learning in Healthcare

Riya Joshi
LY IT B, 1814028
KJSCE
Mumbai, India
riya.joshi@somaiya.edu

Darshan Kalwani
LY IT B, 1814029
KJSCE
Mumbai, India
darshan.kalwani@somaiya.edu

Sagar Kamat
LY IT B, 1814030
KJSCE
Mumbai, India
sagar.kamat@somaiya.edu

***Abstract*—The significance of AI approaches in healthcare has increased significantly in the last decade, owing to the rising availability of huge multimodality data and developed computational models and algorithms. This new trend has sparked increased interest in the use of advanced data analytics and machine learning techniques in a range of healthcare applications. This paper intends to critically examine the use of machine learning in the healthcare business by determining how the technology is altering the industry and what its ultimate potential is.**

Keywords : Machine learning , Deep learning, Health care, Medicine, Model, Algorithm, Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, Big data

I. Introduction

Many parts of healthcare have been transformed by machine learning approaches, from new models that aid doctors in making more informed decisions to new devices that assist individual patients better control their own health. Clinicians have recognised the potential of algorithms to save lives since the 1950s, when Kaiser established the first computerised records for chest X-ray reports and blood test results, as well as the introduction of the pacemaker. As new advancements in image processing, deep learning, and natural language processing alter the healthcare business, this rich history of machine learning for healthcare influences innovative research today. In this paper, we have explored the various applications where machine learning has potentially increased the accuracy of treatment regimens and health outcomes through various algorithmic procedures.

II. Benefits

Machine learning is being increasingly frequently utilised in healthcare, and it is assisting patients and professionals in a variety of ways. Machine learning is most commonly used in healthcare to automate medical billing, clinical decision support, and the establishment of clinical care standards. There are several significant applications of machine learning and healthcare ideas in medicine. The first medical machine learning system to predict acute toxicities in patients getting radiation treatment for head and neck malignancies has been created by researchers. In radiology, deep learning in healthcare automatically detects complicated patterns and assists radiologists in making informed judgments when analysing pictures such as traditional radiography, CT, MRI, PET scans, and radiology reports. Machine learning-based automated detection and diagnostic systems have been demonstrated to perform as well as an expert radiologist. Google's healthcare machine learning programmes were trained to identify breast cancer and reached an accuracy of 89 percent, which is on par with or better than radiologists. These are only a handful of the numerous applications of machine learning in healthcare.

III. Applications

A. Machine Learning Based Medicine Distribution System

In this research paper, a medicine based vending machine using the concept of text recognition in images with the help of machine learning has been proposed. Support Vector Machine (SVM) algorithm has been used along with Connected Components Labelling algorithm to determine the text region to find the largest connected region to locate the medicine name.Then the method of "fragment link" for text segmentation is used , which divides the text into two elements: fragment and link, and combines the whole word according to the set rules. Finally, we directly use Optical Character Recognition (OCR) for character recognition. The experimental results have shown that this method has a high accuracy. This medicine based vending machine not only brings convenience to the patients, but also improves the working efficiency and management level of the hospital and pharmacies to a large extent by eliminating the possibility of incorrectly reading the doctors' handwritten orders and saves a lot of time. Also, systems like these can work 24/7 bringing a relief to the patients who can purchase medicines even at night in urgent cases.

B. A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection

The purpose of this research paper is to use an ensemble of supervised machine learning algorithms so as to detect "depression" (which is a state of mind where a person's mood is always low and shows loss of interest and aversion to activities.This is a very complex disease which can severely affect a person's day to day activities , thoughts, behavior, feelings and sense of well being.) by the use of smartwatches. Different supervised learning algorithms like Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor, and Naïve-Bayes algorithms are used for creating the ensemble model. The dataset used for training the model is taken from kaggle which contains 334 records and

31 different attributes. Smartwatches collect the data from the wearer, analyze it, and produce respective results which can be helpful to detect depression at an early stage.

C. ResNet-Based Model for Cancer Detection

The paper proposes a ResNet-based deep learning network to identify metastatic cancer from cancer scan images by using the PatchCamelyon (PCam) benchmark dataset. The model also uses Test Time Augmentation to improve accuracy of detection. It is a computer aided diagnostic system to help detect Cancer. The ResNet model uses identity shortcut connection to train many layers which improves computational efficiency.

D. Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease

This research paper predicts the presence of heart disease by deploying 6 different Supervised Machine learning algorithms, namely; Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine, K-Nearest Neighbor and logistic Regression algorithms on the heart disease data-set. Then the performance of each algorithm was analyzed and compared with each other on 4 parameters : accuracy, F1-score, and AUC. The purpose of this paper is to ensure that heart risks are predicted early to ensure timely medication and better chances of survival.

E. Real-time machine learning for early detection of heart disease using big data approach

This work proposes a real-time heart disease prediction system based on Apache Spark, a powerful large-scale distributed computing platform that can be successfully employed for streaming data events versus machine learning via in-memory calculations. Streaming processing and data storage and display are the two key components of the system. The first applies a classification model to data events using Spark MLlib and Spark streaming to forecast cardiac disease. The vast volume of created data is stored in Apache Cassandra by the seconds.

F. Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease.

In the paper, the performance of the predictions of the algorithms are analyzed using a pre-processed dataset. The performance analysis is done based on the accuracy of the results, prediction time, ROC and AUC Curve and error rate. The comparison of the algorithms will suggest which algorithm is best fit for predicting the chronic kidney disease.

IV. Results and Discussion

A. Machine Learning Based Medicine Distribution System

1) Database : The database used here was self-made and the images were from the medicine vending machine itself consisting of 10 pictures of each medicine and in total 100 images were there. The dataset was divided in the ratio of 70:30 having 70 samples for training and 30 for testing.

Fig. 1. Medicine images.

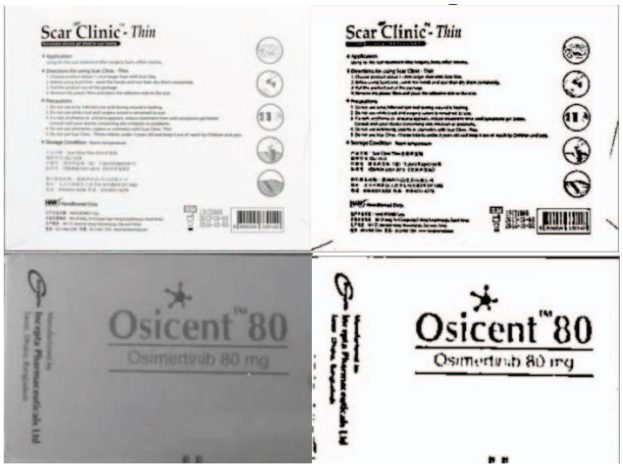

Fig. 2. The medicine image after gray level binarization.

2) Image processing: The medicine's image was pre-processed into 256*256 format and matlab image cutting function was used.

3) Grayscale and binarization of the image:

4) Finding the largest text connected region: The MSER algorithm was applied to find the region with the same gray value inside the text and the connected region in the image was extracted as the candidate region of the text. Then, with the help of the SVM classifier the connected region of the text and non-text was distinguished and the largest connected region of the text was obtained.

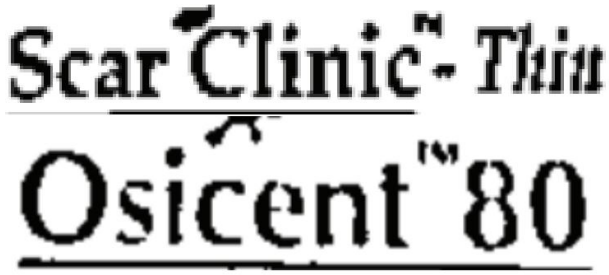

Fig. 3. The largest connected text region obtained.

5) Text fragment links: The fragment link predicts the fragment and link at each location on the image. Next, the detected fragment is combined by the detected link to get the detection result of the whole word. The detection result of the whole word will be represented by the rectangular box, which

is the center point of the rectangular box, and is the length and width of the rectangular box.

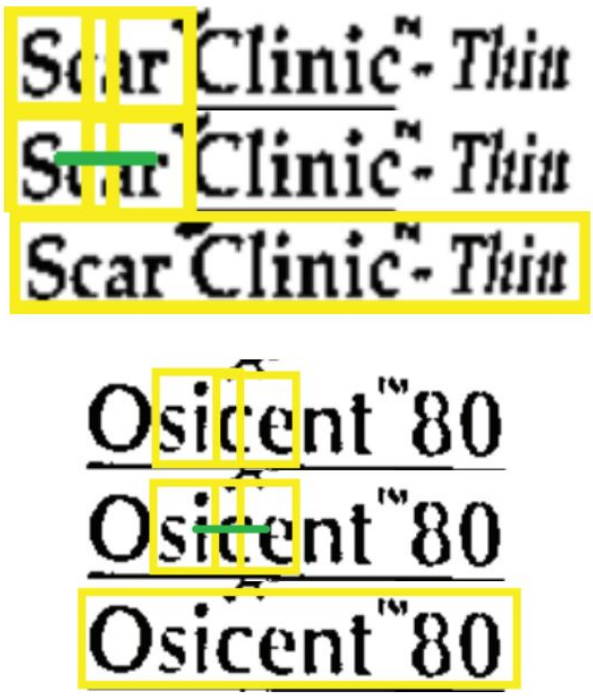

Fig. 4. Fragment link diagram.

6) Identify the text linked from the fragment.: After we have obtained the detection information linked from the connected region segment, the OCR software is used to identify the name of the medicine.

As compared to the other available methods , this has higher accuracy as it was able to correctly detect even the superscript information from the image.

B. A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection

The result and discussion are divided into two parts. A) Data acquisition and visualization, where we acquired the data from Kaggle dataset and visualized it. B) Ensemble Model implementation where we implemented these models to achieve necessary results.

1) Data Acquisition and Visualization: The count of 1 or the orange bar represents the people suffering from depression which is a total of 86 people out of 334. Here also the count of 1 or the orange bar represents the people having anxiety which are 100 in total. By looking at this pie-chart we can make an inference that people suffer more from anxiety as compared to other symptoms of mental illness like : lack of concentration, anxiety, mood swings, etc. Since the data is of unemployment, we also compared people who are unemployed and people who are at least engaged in some kind of part-time job and from the above figure we can see that around 27.5% out of 334 were unemployed and these suffered depression at a greater level.

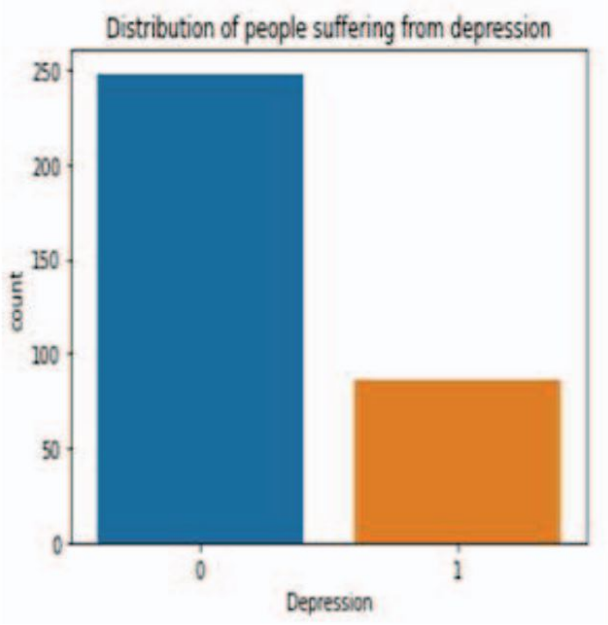

Fig. 5. Distribution of people suffering from depression.

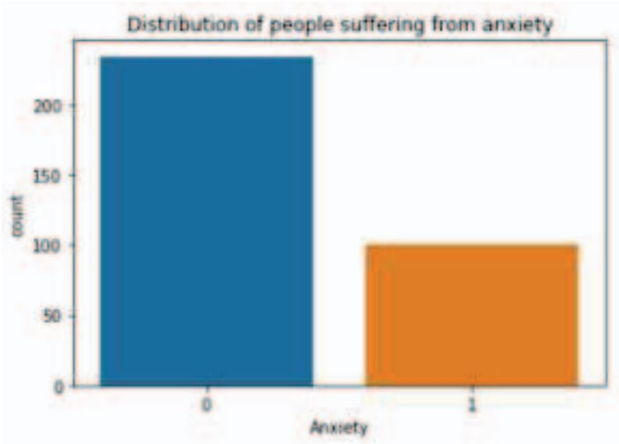

Fig. 6. Distribution of people suffering from anxiety.

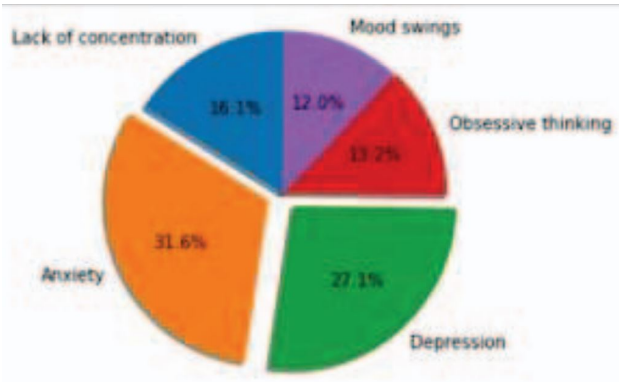

Fig. 7. Pie chart showing people suffering from different mental illness.

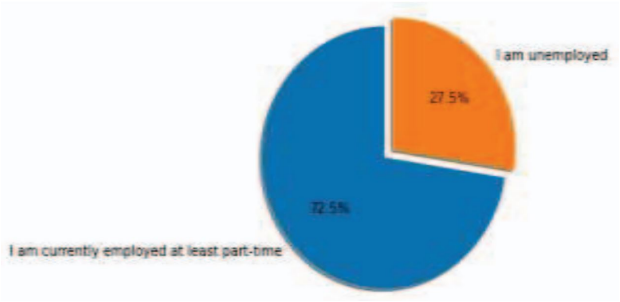

Fig. 8. Pie chart showing comparison between unemployed people and people working as a part-timer.

Model	Algorithms Applied	Accuracy
Ensemble Model 1	K-NN, Logistic Regression, SVM	89.603%
Ensemble Model 2	Decision Tree, Naïve Bayes, SVM	87.539%

Fig. 9. Representing ensemble models and their accuracies.

2) Ensemble model implementation:

C. Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease

Classifiers	AUC	Accuracy	F1 score	Precision
Random Forest	0.882439	0.835165	0.842105	0.888889
Support Vector Machine	0.862439	0.824176	0.836735	0.854167
Naïve Bayes	0.860000	0.824176	0.833333	0.869565
Logistic Regression	0.855610	0.802198	0.820000	0.820000
Decision Tree	0.789024	0.791209	0.804124	0.829787
K Nearest Neighbors	0.768537	0.725275	0.747475	0.755102

Fig. 10. Performance evaluation of supervised learning algorithms.

D. Real-time machine learning for early detection of heart disease using big data approach .

The accuracy of the diagnostic is kept at 87.50 percent.

Predicted	Actual	
	Heart disease	No heart disease
Heart disease	39	5
No Heart disease	6	38

Fig. 11. Confusion matrix of classifier used for classification of heart disease.

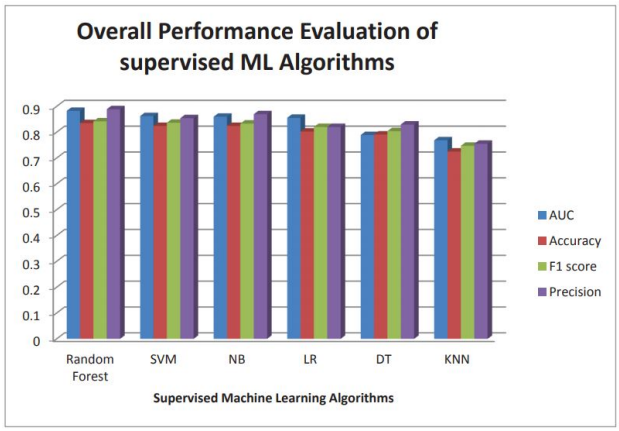

Fig. 12. Overall performance Evaluation of supervised machine learning algorithms.

V. Drawbacks

- Due to automation of processes in healthcare industry there will be mass displacement of workforce so many people will lose their jobs
- The accuracy of the machine learning models is not 100% so there is no guarantee that the diagnostics will be correct.
- There is also a huge possibility of data leaks of the personal and health records of patients which are stored in Electronic Health Records systems by the hospitals.

VI. Conclusion

We discussed some applications of machine learning in healthcare and saw the huge potential Machine learning has in the healthcare industry and how this sector can advance into a new realm and completely transform the healthcare operations. It can be applied to various domains such as in early stage detection and prediction of diseases, clinical distribution systems, easy wearables devices, suggesting preventive measures, administrative applications, electronic health records, telemedicine solutions, drug discovery etc. The accuracy of the models used can be increased by taking large datasets for training and by taking ensembles of machine learning models or by using deep learning approaches. Use of machine learning can enhance the productivity and lower the expenses and will help in serving more patients and hence will improve the healthcare outcomes to a large extent.

VII. Future scope of research

- For applications dealing with object recognition , different techniques can be implemented like MSER with Ada boost, State Estimation, Texture Based SVM, Chain Level Classification, CMSER Algorithm, Graph Model and Fuzzy Method.
- Future research can be done in the domain of using smart watches for depression detection so that they will be able to accurately predict the severity of depression that the wearer is going through and must be connected through

a chain of smartphones so that family members can keep a watch on the person and accordingly help them or seek guidance from doctors.Along with the smartwatch sensor: accelerometer , gyroscope , heart rate monitor can also be included in the device.

- Neural networks or deep learning models can also be used to get higher accuracies.
- More research can be done in cancer prediction by training the machine learning models for different sizes of cancer cells.
- It will be more efficient to include other big data technologies into our strategy. Traditional analytical tools make developing a distributed and real-time healthcare analytics system exceedingly difficult, but open source big data technologies make it more easier and more effective.
- Physical robots can also be made which can be used in hospitals for doing pre-defined tasks such as giving medicines , or food to the patients, to relocate or bring some things or delivering supplies.
- Apart from this, curing diseases with the help of machine learning should also be looked into in the near future.

REFERENCES

[1] Huiling Xia, Chunzhi Wang, Lingyu Yan, Xinhua Dong, Yichao Wang, "Machine Learning Based Medicine Distribution System", 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application, 2019.

[2] Piyush Kumar, Rishi Chauhan, Thompson Stephan, Achyut Shankar, Sanjeev Thakur, "A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection", 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence 2021), 2021

[3] Zheng, Zhuo; Zhang, Hao; Li, Xinjian; Liu, Shuai; Teng, Yukun. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE) Consumer Electronics and Computer Engineering (ICCECE), 2021 IEEE International Conference on. :325-328 Jan, 2021;

[4] Sujatha, P.; Mahalakshmi, K.. In: 2020 IEEE International Conference for Innovation in Technology (INOCON) Innovation in Technology (INOCON), 2020 IEEE International Conference for. :1-7 Nov, 2020;

[5] F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-7.

[6] A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2019, pp. 1-5.