International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

Health1: A Disease Detection Tool

Akshay Anil Baviskar[1], Divya Pandey[2], Alok Singh[3], Pratyush Mahalle[4], Suyash Varma[5], Prof. Ashvini Jadhav[6]

[1,2,3,4,5]UG Scholar, Dept. Of Information Technology, MIT ADT University, Pune
[6]Professor, Dept. Of Information Technology, MIT ADT University, Pune

---

Abstract - *Artificial intelligence is radically changing medical practice. Advanced levels in machine literacy, digital data collection, and calculating structure have permitted artificial intelligence operations to penetrate into fields believed not long ago to be concerned with the elite dispatch of mortal specialists. This review essay outlines the fiscal, legal, and social consequences of artificial intelligence when it comes to the setting of health care operations. Many diseases are happening to human beings due to the environment condition and living habit. So, predicting the disease in the early phase has become an important task. But detection using symptom is too difficult for doctors to get accurate output. So, disease detection should be the most difficult task. But by considering data mining, the problem of disease prediction can be solved. On the other hand, supervised machine learning algorithms showed great potential to outperform the current standard disease diagnosis process and medical professionals in the early detection of high risk diseases. The objective is to identify the patterns between various supervised ML models and performance metrics to identify disease detection trends.*

A limited number of advancements in AI technologies and their biomedical operations are also summarized as well as the obstacles to the far restraint of additional developments in AI for Health Cure Application. Some of the major Features of the Health Cure Application consists of: It can able to detect 7 type of diseases which are –: Covid-19 detection, brain Tumour detection, breast cancer detection, Alzheimer's detection, diabetes detection, pneumonia detection and Heart disease detection.

Key Words: **CNN, Machine Learning, XGBoost**

1.INTRODUCTION

Disease Inference System is the most relevant and efficient for machine learning used to predict human diseases is to provide the relevant symptoms. Our system uses powerful machine learning algorithms for symptoms from the user's prediction of diseases. The health care system has used and produced quite a large amount of data that can be used to develop knowledge about a patient's unique disease. The healthcare information will be previous treated for the treatment of the patient's health with effective and best possible treatment. This area also needs some improvement using informative data in health science. However, the main problems are the extraction of information from that data because the data is presented in huge amount and that some data mining and machine learning techniques are used. The integration of the trio: machine learning, Python, and Flask has, gratefully, established a new way in which many medical diseases are detected and managed In the paradigm shifting, ever-changing healthcare civilization. Our project's vector, "Health Cure Solution", is grounded on specific non-negotiables centre on tirelessly nourishing accessible and accurateness available in healthcare, and providing it in suitable amount in our evolving health matter alignment, attributable to many factors that instantly renew the arena of the formation of modern creativeness for living and wellness of today. The platform's health-focus is structured is seeded on the instant, convenient observation and detection of several crucial diseases, accumulated through a diversity grade of machine learning Platform, and the program language Python, and the robust useable interface in the Flask framework. Python is the platform's dynamic language that is versatile with the extensive and coherent number of official and member released packages that plan on making autonomous Python functions downstream and can be effectively employed source files. Python packs containing NumPy, SciPy, and Pandas, this packs deliver a lucid foundation onto which extra engaged scientific software production uses reason without fear of exerting low volumes of analogies. Similarly, many discipline-applied packages are available for forecasting needs from the meteorological field. Also, State-of-the-art Machine Learning algorithms have also been implemented.

An expected outcome of these project is predating the disease in advances, and it can prevent the threat of life in time and save people's life and reduces the cost of treatment to some extent. Detective diseases are 1. Breast cancer detection 2. Diabetes detection. 3. Heart disease detection. Recent work on deep learning has focused on several largest areas in machine learning. Machine learning models will learn and understand after the raw data's hierarchical representation with some preprocessing data and it has been declared as a concept called as big data technology and more attention towards disease prediction has been devoted.

1.1 Breast Cancer Detection

Breast cancer is recognized as a multifactorial disease in the world and the most common cancer in women, with 30% of all cancer cases in women; 15 million women are

diagnosed with breast cancer worldwide and 5,00,000 of demise each year. The disease has increased over the past 30 years, whereas the death rate has decreased. However, the decline in death rate from mammography screening is 20%.

1.2 Diabetes Detection

Diabetes is a detrimental disease in this world. Diabetes is caused by obesity or high blood glucose and so on. It harms the hormone insulin, which causes improper crab metabolism and enhances blood sugar issues. Diabetes happens if your body doesn't produce enough insulin. Diabetes is one of the most frequent causes of death in this world today. Through the early prediction of diseases such as diabetes, these diseases can be controlled and human life can be saved. Therefore, this current work focuses on a diabetes prediction study with different diabetes disease related attributes.

1.3 Heart Disease Detection

The term "heart disease" is commonly applied to "cardiovascular disease". These are both terms for disorders that involve the bloodstream: Heart disease or stroke, – Chest pain or 'angina pectoris', – Heart attack. Other sorts of heart disease, such as those that are associated with the rhythm, valve, or muscle, are other sorts of heart disease.

2. Problem Identification

The system overviews with the predictions the likelihood of a chronic disease that is for a specific region and a specific community. This system predicts only some disease. In this system, disease risk prediction is done with Big Data & CNN algorithm. Whereas for a type S data, the system disease risk prediction is with machine learning algorithm i.e. K-nearest Neighbours, Decision Tree, Naive Bayesian. The accuracy of the system is up to 94.8%. The research paper "they streamline machine learning algorithms for effective prediction of chronic disease outbreaks in high-disease communities. They experiment with novel prediction models on actual hospital data collected from "Kaggle". They propose a convolutional neural network-based multimodal disease risk prediction algorithm using structured and unstructured hospital data. The system predicts chronic diseases that are region-specific and community-specific. This system predicts only some disease. In this system, disease risk prediction is with Big Data & CNN algorithm. Whereas for a type S data, the system disease risk prediction is with machine learning algorithm i.e. K-nearest Neighbours, Decision Tree, Naive Bayesian. The accuracy of the system is up to 94.8%.

The Health Cure Application, developed using artificial intelligence, deep learning, and machine learning, is a groundbreaking technology in healthcare that seeks to improve patient outcomes in medical care. Not only is this a challenge in itself, but the development process must also comply with legal provisions such as the Health Insurance Portability and Accountability Act. The main challenge in developing the Health Cure Application shall be the integration of ML, DL, and AI into enhanced health care.

Algorithm Bias and Interpretability: In the disease detection field, mitigating algorithmic bias and ensuring the interpretability of machine learning models are major challenges. Correcting biases that have been incorporated into training data and increasing the transparency of outputs of AI-driven diagnostics are crucial to building trust and confidence in the diagnostic process. This requires further research and development.

User Adoption and Accessibility The other main challenges are ensuring the disease detection platform can be adopted by the end-users and ensuring it can be accessed by a wide range of demographic users and from various geographical locations. The parity of adapting the platform interface to user demographic dynamics, language specificity and sensitivity, and cultural subtlety are crucial in enhancing user engagement, inclusiveness, and equitable healthcare access.

3. Proposed Work

For the proposed work, the aim is to create a robust disease detection platform that can serve immediate healthcare requirements and promote proactive approaches for disease prevention through the use of modern technologies. Disease detection platform – The primary goal will be to develop an inclusive platform named "Health Cure Solution" that can identify seven distinct diseases namely Covid-19, brain tumours, breast cancer, diabetes, Alzheimer's disease, pneumonia, and heart disease. Python and Machine Learning usage – The projects will include the use of Python language and machine learning models, namely the CNNs, SVMs, and decision trees. These models will help validate that the disease identification is accurate by incorporating data from medical sources of all forms. Flask framework integration – Further, the project will integrate the Flask framework to develop a patient and healthcare provider-friendly interaction pattern. There will be a model interaction between the patients and disease detection models and delink the healthcare provider and models through this simple interface. Scalability and continuous improvement – The solution will be developed as a constantly evolving system that would continue to be updated with newer technologies, improved performances, and changes. This way, the developed solution would be responsive, efficient, and adjustable to the needs of the healthcare system. Validation of accuracy and future planning – Finally, the project would discuss the claimed accuracy of the machine learning disease detection models and make future plans on how the models can be more accurate, inclusive of other disease detection, provide preventive recommendations, keep track of the detections, and provide a user-friendly interaction. Data encryption and privatization – To protect

the medical data from unnecessary access and unworthy leakage, the system would largely use data encryption with strong access controls.

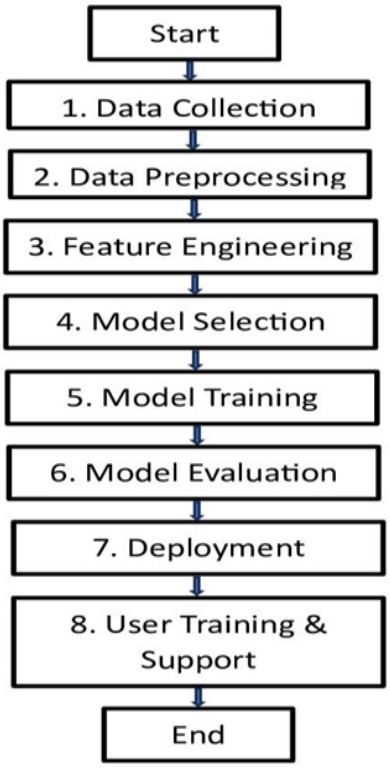

Fig 1 : Flow Chart of Project

Disease Predictor System – This system is used to predict most chronic diseases. This system will take structured and text type data to the machine learning model, and this system is used by end users. This system is used to predict disease based on symptoms and is based on machine learning technology. The final output of this is to predict the power of the random forest and the XG boost algorithm we are giving for prediction the final output will be in 0 or 1, for this the Logistic tree is used. Thus, the diseases are predicted by our system. It takes a structured data type to the machine learning model, and this system is used by end users that are patients or any user. That means the user will enter the symptoms he/she suffers from. These symptoms are fed to machine learning and predict the disease. We are applying algorithms & in this case, we are applying the maximum accuracy of algorithms. Thus, the system will predict diseases based on symptoms. The final output of this is the disease predicted by the model.

4. Hardware Requirement

- System should have 8 GB of RAM (Random Access Memory)
- The CPU must be x86 64-bit CPU i.e., INTEL/AMD architecture
- A reliable network must be in place.
- The network must be well designed with good bandwidth for data transfer and communication between the application, database server and users.
- Sufficient storage to store code, database and other files.

5. Software Framework And Tools

- Windows 10 OS
- Programming language such as language used to build application i.e. Java, Python etc.
- Anaconda
- Front-end Technologies HTML, CSS, JavaScript and frontend framework React, or Angular and Flask.

6. Proposed Model

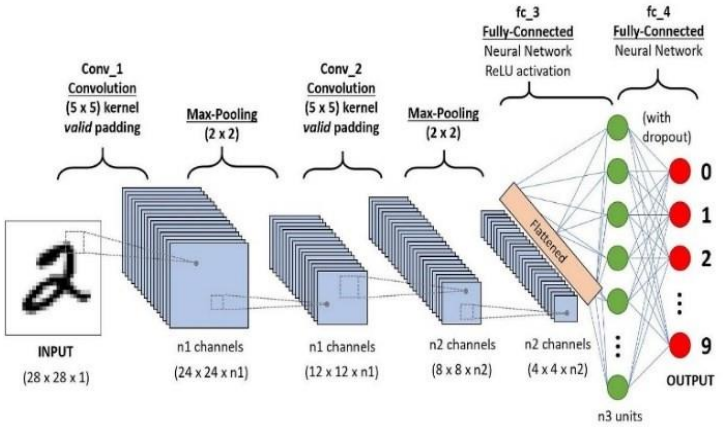

Fig 2: CNN Working

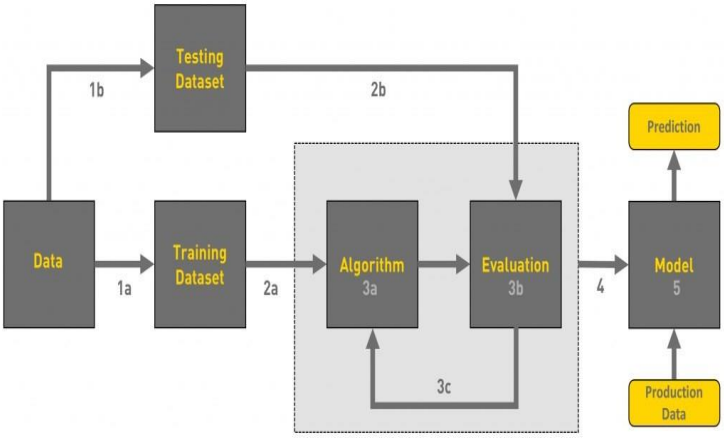

Fig 3: ML Model Working

7. Experimental Results And Discussion

Dataset Collection:

Here, first, we have to collect a dataset for our project. Once done with dataset selection, we divided the dataset into training and testing data. The data you use to learn the prediction model is called Training Data and the data you use to predict is called testing data. We used 70% of the data for training purposes and 30% of the data for testing purposes. For this project, the dataset which used was Heart Disease UCI. The dataset comprised of 76 attributes, only used 14 for this system.

Selecting Attributes: Attribute selection is an important element in building the prediction system. Generally, it is not good to add all the column values. Selecting the abstract attributes in the training and testing data is beneficial for model efficiency. Used different attributes concerning patients such as gender, chest pain, high blood pressures, cholesterol etc, to predict the result. A correlation matrix is used to select the attributes in the model.

Data pre-processing:

Data preprocessing is the act of converting the initial data into usable form upon launching your data model. Data pre-processing is required because the model accuracy requires clean data to process. Noisy, missing values. Data pre-processing involves a data import, dataset partition, attribute scaling, etc., the sub-model accuracy increases after data processing.

Data Balance:

There are two ways to balance an unbalanced dataset. They are under-sampling and over-sampling.

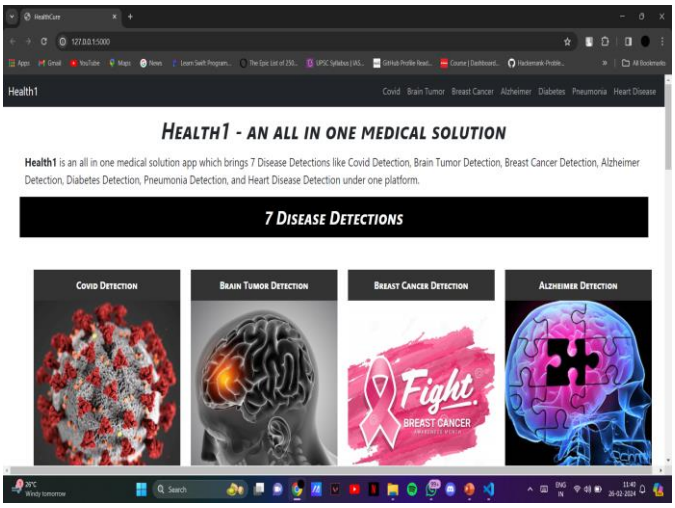

Fig 4: User Interface Of The Application

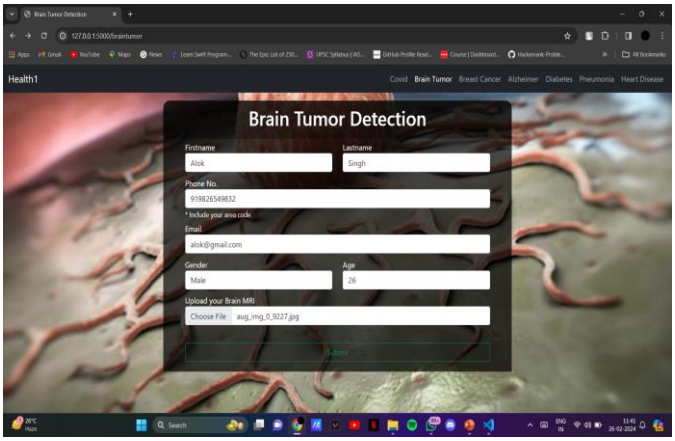

Fig 5: User updating details UI

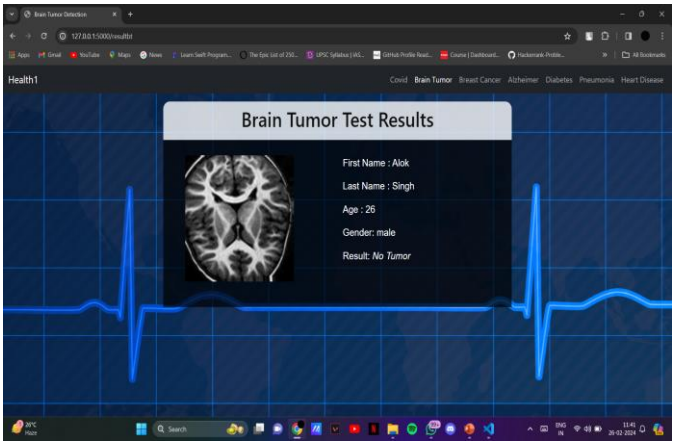

Fig 6: Result Page

Correspondingly, The Health Cure Application detects seven severe illnesses, such as brain tumours, diabetes, Alzheimer's disease, pneumonia, heart disease, and Covid-19 efficiently and accurately. Our application guarantees a high case detection rate, maintain the interaction with the application user-friendly, and susceptibility by utilizing the Flask framework and the machine learning approach designed with the Python language. The development process conducted in our project has addressed data security issues, processed more data to ensure robust performance, and maintained a high case detection rate. Therefore, the project illustrates the implementation of the ability and speed of these technologies in the diagnostic field.

8. Conclusion

Denotable, good and higher accuracy is achieved with our proposed system. This is to be utilized by researchers, doctors or physicians in ensuring the best treatment and medications possible based on the accuracy. Machine learning in healthcare enables effective treatment and well taking care of the patient. The case study is in relation to the implementation of some healthcare machine learning features on our system. In other words, unlike direct

diagnosis, if a patient is predicted to have a disease. ML is essentially a kind of machine learning algorithm and therefore healthcare can be made even smarter. The algorithms chosen for disease prediction from our dataset and the expected output when compared is that logistic regression algorithm and KNN algorithm are the best algorithms that had the highest performance compared to LDA algorithm. Machine learning provides the methods, practices, and easily comprehensible techniques that will help to solve the problem of diagnostic problems with the medical range with the simplest and modernized ML. Even ML has been used in the prediction and analysis of clinical trials among the people and is used to analyse the data process which includes the detection of an error in the dataset and to handle the incorrect data error present in our system. It is clear to give an argument about the integration of the perfect use and implementation of ML algorithm that is a great source of help with the computer system in the range of health care to improve and simplify the doctor's work which in terms lead to an increase in the efficiency and the quality of our medications to the patients. This provided a comparative review of the recent fashion of work on prognosis and prediction of Alzheimer's disease by machine learning methodologies. This Comparison of the kinds of data used, and the role of machine learning methodologies in powerful early prediction side by side of Alzheimer's disease. The information hoisted lessen the accuracy of predicting the clinical identification as 100% perfect because in other words, the pathological corroboration was not supplied for the predicted outcomes hence it might have introduced some uncertainty.

9. Future Enhancement

Therefore, a computer does most of the heavy computation and delivers conclusions that correlate almost to reality. Machine learning, abbreviated as ML, is the application of several learning methods and algorithms that enable computers or machines to perform work that are a complex lie simplified manner. ML is also applicable in academia, that is for students or scholars, and also in industry to predict events accurately and use these different sources of data sets and information. All in all, we can say that we have increased in big data, machine learning, and data science etc. and we have been part of those industries which have been able to acquire such information and employ employees to produce their goods and services in an appealing way. The learning methods established for these industries and investigations give promising potential for further improvement in medical research and clinical care of a patient in the best way possible. Machine learning utilizes mathematical algorithms and techniques that are used to define the relationship between the variables used in the model and others. In other words, our article will describe the process of training a model and learning appropriate algorithm on screening to predict disease based on the properties of a tumor tissue sample. Although these algorithms work in unique and different ways depending on how they have been developed and utilized by researchers. One way to assess them is to think of their utmost goal. Our paper and statistical methods assess and come to a conclusion about data which have been collected from a variety of samples of our public. Despite the fact that so many techniques such as linear and logistics regression are able to predict diseases. E.g. with a model that describes and predicts the relationship between clinical variables and transience, we can track the yearly number of organ transplant surgeries, i.e. we need factors and variables that differentiate patients who have low mortality from those who have high mortality , We can make such predictions on the outcome and in the next near future arrive at a conclusion and reduce mortality rate in case of a high mortality rate, although it cannot be set that it is better than desired.

REFERENCES

[1] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

[2] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[3] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[4] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (Vol. 1, No. 2).

[5] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[6] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[8] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[10] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211-252.

[11] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Desmaison, A. (2019). PyTorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems (pp. 8024-8035).

[12] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16) (pp. 265-283).