Career Point International Journal of Research(CPIJR)
©2022 CPIJR | Volume 4 | Issue 1 | ISSN : 2583-1895

AI Services Implementation through Machine Learning Tools, Analyzing and visualizing the data for early Disease prediction

Dr. Abid Hussain

Associate Professor School of Computer Application, Career Point University, Kota, India

Abstract- The body's sugar levels become abnormally high over time in diabetes mellitus. Consequently, it causes damage to numerous body systems, including neurons and blood vessels. This sickness has a high guess for early location, which can assist with saving living souls. Examination of information includes acquiring and assessing information to get experiences that might be utilized for independent direction. Huge data sets and metrics are visualized using charts, graphs, and other visualizations. Data visualization and analysis will be presented in this paper, resulting in recommendations for algorithms and methods for early diabetic prediction. We used Azure plat form and Rapid Miner to comprehend the data.

Keywords: - Rapid Miner, Azure, Diabetics, Machine learning, Analysis, Analytics.

I. INTRODUCTION

Numerous chronic illnesses are widespread worldwide, both in developing and industrialized nations. Diabetes is a metabolic condition in which insulin levels are altered by either increasing or decreasing. [2]Human real parts, for example, the eyes, kidneys, heart and nerves are completely impacted by diabetes. As a result, we have gathered the information from the end user and a well-known pathology lab in Pune through a Google form. [3]To fully comprehend the data, we have decided to conduct data analysis and visualization prior to its actual implementation. The process of cleansing, converting, and modeling data in order to uncover information that can be used for corporate decision-making is referred to as data analysis. A decision can be made as a result of the data analysis. Data visualization is the graphic representation of data and information. By employing visual components like maps, graphs, charts, and graphs, these tools make it simpler to examine and comprehend data.

We pre-processed and visualized the data with Rapid Miner Studio and the free Azure AI service.

II. LITERATURE REVIEW

Numerous businesses and academic fields have paid close attention to AI's recent significant advancements. Deep Learning (DL), a collection of processes and algorithms that automatically enable computers to automatically detect complex patterns in large datasets, is the most successful technique driven by advancements in ANNs. These advancements are fueled by more readily available computing power, user-friendly software frameworks, and increased access to data—also known as "big data," which enables the widespread application of deep neural networks. When neural networks outperformed other methods in a number of high-resolution image analysis criteria, DL gained prominence in image processing.

In 2012, a CNN model reduced the second-highest error rate in image classification work by 50% in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [12]. Before that, it was thought that computers had a hard time recognizing objects in natural images. In ILSVRC, CNN has even outperformed human performance to the point where ILSVRC classification is practically solved. For a variety of computer vision issues, DL techniques have emerged as the objective standard. The application of DL methods to the diagnosis of acute human diseases has been suggested by numerous studies.

Multiple scenarios based on ML and DL models have been used by researchers to predict conditions like liver disease, heart disease, Alzheimer's disease, and a variety of cancers that can only be treated early . Using pediatric chest radiographs, some researchers have used DL techniques to diagnose and differentiate bacterial pneumonia. Critical endeavors have additionally been made to recognize the various elements of chest CT imaging qualities of

different illnesses. In various studies, new hybrid models based on Case-Based Reasoning were proposed for the diagnosis of various skin diseases. The application's output from the model could diagnose and recommend treatment for a variety of skin conditions. In healthcare, personalized real-time monitoring systems based on ANN techniques are widely used to obtain vital body information. This device can assist patients with health management, particularly in critical situations. ANN models were used by researchers in to accurately predict diabetes disease.

Data is automatically analyzed using ML classification techniques to diagnose or predict various diseases. Scientists fostered a customized ongoing electronic medical services checking framework to get essential data from the body, for example, pulse or circulatory strain. This device can assist patients with health management, particularly in critical situations.

In the healthcare system, AI and the Internet of Things (IoT) can improve healthcare technology and treatment procedures. A solid IoT-based framework involving ML calculations for medical services was proposed to screen human exercises and the general climate through the body sensor organization, BSN-Care. To predict type 2 diabetes (T2D), another study suggested a hybrid IoT model that combined a healthcare monitoring system with the Random Forest method. The random forest classifier, which performed better than other algorithms, was also used to investigate the risk of T2D among individuals based on personal lifestyle information.

Using the random forest classifier, a mobile platform for real-time tuberculosis disease (TD) antigen-specific antibody detection was developed with 98.4% accuracy. Using RNN and LSTM networks, a research study with 97.057% accuracy proposed an AI-based framework for classifying multiple gastrointestinal (GI) diseases.

The two most important aspects of hypertension healthcare control and awareness are the reduction of stroke and cardiovascular disease. In this regard, researchers evaluated AI and digital healthcare technologies and suggested a privacy protection system for the collection and storage of individuals' data . In addition, numerous studies on disease prediction have been conducted by researchers to identify and anticipate diseases in their early stages. Based on the Internet of Things, a novel hybrid ML model with a precision of 99.50% and an accuracy of 100% was proposed for the initial phase of disease detection . A

method for predicting cardiovascular disease based on various characteristics has been proposed by researchers in another work. They achieved 88.7% accuracy using a hybrid random forest classifier .

An ML algorithm known as XGBoost was proposed in a study on the detection of positive urine culture results. The accuracy of this model, which was superior to that of other developed models, ranged from 0.826 to 0.904. Another Vol.: (0123456789) Discover Artificial Intelligence (2023) 3:5 | https://doi.org/10.1007/s44163-023-00049-5 Review 1: The CNN model was used to extract features from images of malaria-infected blood cells in this study . An ML model was also used to predict malaria infection in another study. In order to identify acute exacerbations in patients with chronic obstructive pulmonary disease, researchers employed a variety of machine learning (ML) techniques, including Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbor, and Naive Bayes. They discovered that the SVM model performed the best. Based on data from the National Survey on Drug Use and Health from 2015 to 2017, researchers used three ML algorithms—ANN, distributed random forest, and gradient boosting—to predict opioid abuse among adolescents in other studies. The area under the receiver operating characteristic curve (AUROC) prediction performance ranges from 0.809 to 0.815. Similarly, other researchers proposed a model for detecting COVID-19 from an X-ray image dataset with 98.91 percent accuracy using multiple ML algorithms like CNN, RF, SVM, DT, and AdaBoost. Individual stress levels can be detected using ML and DL methods. One methodology is to utilize physiological signs, for example, pulse or breath,

to recognize pressure. For instance, a large study looked at various ML models based on heart rate variability for stress levels. ML Random Forest outperformed other approaches in this study. To predict diabetes, various ML models were utilized by other researchers. In their work, Calculated Relapse and Backing Vector Machines performed well. KNN, SVM, ANN, Decision Tree, Logistic Regression, Naive Bayes, Random Forest, and XGBoost were used in a comprehensive study to predict the risk of chronic type 2 diabetes. With 0.91 AUC, the Random Forest model outperformed the other models in this study. An extended DL model known as 3DCellSeg has recently outperformed basic models when it comes to analyzing and distinguishing image-based diseases. There is only one hyper parameter required for this DL approach, which uses a light deep CNN.

III. METHODOLOGY

Artificial intelligence techniques in disease diagnosis and prediction

In order to succeed and acquire more precise knowledge regarding dangerous disorders and diseases, AI technologies are increasingly being utilized in medicine. In the medical field, AI is increasingly used for disease diagnosis and prediction due to its positive interaction with image data. The two most important tools for effectively implementing AI methods in the health care system are learning algorithms and big data derived from medical records or wearable devices. These tools can be used to improve disease diagnosis, disease classification, decision-making

processes, activities, walking aids performance, providing the best treatment options, and ultimately helping people live safer and longer lives. In a short amount of time, AI is used to improve medical analysis and diagnosis.

For instance, this technology is able to identify potentially harmful tumors in medical images, enabling pathologists to treat the disease rather than sending samples of tissues or lesions to a laboratory for long-term research. Unencoded, rare, and undiagnosed patients can all be successfully identified using AI-based algorithms. As a result, AI models for disease diagnosis afford patients ample opportunities for early diagnosis.

The use of ML and DL strategies to analyze heart illnesses is expanding fundamentally. In cardiology, there is a wide variety of medical imaging techniques, such as CT, ECG, and echocardiography, so DL can be used to analyze and review cardiovascular data accurately and effectively. Common cardiovascular disease with severe disability and morbidity is coronary atherosclerotic heart disease. This disease can be detected early, which has a significant impact on treatment. In this period, the diagnosis of coronary atherosclerotic heart disease has made significant progress using ML and DL methods. For instance, a powerful tool for predicting major adverse cardiac events would be CT-Fractional Flow Reserve (CT-FFR) based on ML, which can speed up diagnosis and reduce time [81, 82]. Additionally, CT-FFR based on the DL can simplify computation, speed up prediction, and reduce time. In 170 patients, researchers used SVM and ANN techniques to identify various heart conditions early [85]. SVM and ANN models were used to investigate CHD, CAD, arrhythmia, and cardiomyopathy. For arrhythmia, cardiomyopathy, coronary heart disease, and coronary artery disease (CAD), the SVM algorithm achieved an accuracy of 89.1%, 80.2%, 83.1%, and 71.2 percent, respectively. In a similar vein, the accuracy of the ANN algorithm for CAD, cardiomyopathy, arrhythmia, and CHD was 69.6%, 72.7%, and 85.8%, respectively. The South African Heart Disease dataset of 462 samples was used in another study to predict coronary heart disease (CHD). They used Naive Bayes, Support Vector Machine, and Decision Tree supervised learning methods to diagnose CHD and boost its prediction rate. In their study, the accuracy of the library data for cardiovascular diseases was 83.9 percent, while the accuracy of the library data for diabetes was 95.7 percent. The SVM classifier outperformed other ML methods in predicting CHD with 95 percent accuracy, according to another study. In addition, another group of researchers looked into the ability of SVM, ANN, and Decision Tree algorithms to predict CHD disease in 502 samples. With an accuracy of 92.1%, SVM outperformed the other two algorithms in these three accuracy tests. The widespread applications of AI in the medical field have also enabled precise diagnosis and prediction of brain diseases.

The diagnosis of various brain and neurodegenerative diseases like Alzheimer's disease (AD), Parkinson's disease (PD), and brain tumor, which has always been very difficult to detect in the early stages, is where the most recent ML and DL approaches are used. Man-made intelligence has made it conceivable to process and examine an enormous measure of mind cues and information to find

bits of knowledge and connections which are not totally clear to the natural eye. The most broadly involved calculation for infection

identification is DL-based CNN models. Pre-trained models for predicting and detecting Alzheimer's disease were the subject of a recent study. In this review, the EfficientNetB0 model beat different models and got a precision of 92.98%. A blend of various simulated intelligence calculations was utilized for the early finding of Parkinson's sickness as of late.

With an accuracy of 95.58 percent, the combination of the genetic algorithm and random forest produced the best result, making it the best result in recent research in this area. Advanced AI algorithms also greatly aid in early breast cancer detection and prediction. Breast cancer is a disease that kills a lot of women and kills millions of people every year. However, early diagnosis is essential for treating and controlling the condition. The Wisconsin Bosom Disease Dataset (WBCD) is a broadly utilized dataset for specialists researching ML techniques to analyze bosom disease. WBCD was successfully used to diagnose breast cancer with the least-squares support vector machine (LSSVM) algorithm, which achieved a classification accuracy of 98.53% . On the WBCD, a hybrid fuzzy-artificial immune system with a k-nearest neighbor algorithm was also proposed, and its classification accuracy was 99.14 percent. The feature selection was combined with the SVM algorithm.

IV. RESULT & IMPLEMENTATION

The diseases and their respective accuracy levels that the patient is currently suffering from will be the outcome of the developed system. The degree of accuracy for a particular disease will be determined by a variety of factors, including the patient's age, gender, and medical history, when the data are analyzed in accordance with the steps of the Knowledge Discovery process depicted in Figure 2.2. The clinical doctors will benefit from the outcome of the subsequent data mining in order to enable them to treat patients with greater accuracy for specific diseases.

By simply entering the user's health data and symptoms, a health prediction system will enable doctors and medical staff to reduce the amount of effort they need to put into their clinical decision-making process. The system will use data-mining algorithms to cleverly deduce the patient's disease by correlating the patient's information with the health information provided by doctors and medical professionals and stored in a database. This would effectively cut down on the amount of time doctors spend making clinical decisions and the difficult work they have to do. The system will also help doctors and patients talk to each other by pointing them to doctors who are right for their diagnosis and relevant to their particular medical fields.

The goal of this project is to create a system that is easy for patients and doctors to use for diagnosing illnesses and providing appropriate guidance for their current health issues. PCs will be the only platforms for which the software can be installed.

The following features must be included in the proposed system:

4.1 Patient Registration To use the system, patients would need to sign up for the first time using their username and password.

4.2 Patient Access:
Patients would need to use their username and password to access their system.

4.3 Viewing Patient Details In order to familiarize themselves, doctors and patients may view each other's details.

4.4 Disease Prediction Answer multiple questions and use data mining to identify the most accurate symptoms to identify the illness or diseases that the user is attempting to describe.

4.5 Search Doctors and Patients Patients and doctors can search for each other by specialty, diseases they've been diagnosed with, and other references.

4.6 Providing Inputs
Specialists and patients might give input that might act as an extra data to be seen.

4.7 Adding New Diseases and Symptoms Administrators have the ability to add new diseases and symptoms to the system so that patients and doctors can look at them.

4.8 Doctor Login To use the system, doctors must log in using their username and password.

4.9 Administrators could add a new doctor to the system, register them, and give them a username and password.

4.10 Administrator Login Doctors must use their username and password to access the system.

4.11 View Diseases The system's database provides administrators with access to a variety of disease-specific information.

4.12 Sharing Data
Specialists could share data of a sickness or patient to one more specialist for confirmation.

V. CONCLUSION

The diagnosis and prediction of diseases could be completely transformed by DL and ML methods. The most important aspect of the treatment process is the accuracy and correctness of the disease diagnosis. AI has demonstrated significant accuracy in both the prediction of treatment outcomes in terms of survival rate and treatment response and the detection of image-based diseases. The huge amount of picture information requires execution into

handling stages through prompt, solid, and precise processing power given by computer based intelligence strategies. Issues like detection accuracy, effective treatment, and ensuring patients' well-

being are crucial in disease diagnosis.

The vast and varied data, algorithms, deep computing techniques, a variety of neural networks, and emerging techniques that AI encompasses are constantly evolving to meet the requirements of humans. The purpose of this study is to examine how well AI methods work for diagnosing and predicting various diseases. The results of this study indicate that SVM performs best in predicting heart diseases. Due to their high accuracy and quick image recognition, supervised DL networks, such as CNN-based models, are widely used, particularly for diagnosing respiratory, lung, skin, and brain diseases, with significant results. When diagnosing breast cancer, KNN and other networks, like SVM, are typically combined to produce high accuracy. Because of their impressive experimental results in detecting and classifying medical images, DL and ML have a significant impact on the success of numerous diseases that are the subject of this study. By optimizing the utilization of various resources, AI-based methods aid medical systems in diagnosing and predicting conditions. Also, doctors won't have to struggle as much in the near future to make an accurate diagnosis of various diseases thanks to the rapid advancement of AI technologies.

VI. REFERENCES

[1] R. Scheffler and D. Arnold, "Projecting shortages and surpluses of doctors and nurses in the OECD: what looms ahead," Health Economics, Policy and Law, vol. 14, no. 2, pp.274-290, 2019. [Online]. Available at: https://sci-hub.tw/https://doi.org/10.1017/S174413311700055X#. [Accessed 18th November 2019].

[2] S. Chowdhury, "A Review Paper on Health Data Integration with Secured Record Linkage in National Health Data Warehouse and Privacy & Compatible Solution for its Privacy and Security Problems in Bangladesh and other developing countries," International Islamic University Chittagong (IIUC), 2017. [Online]. Available at: https://works.bepress.com/sayem-binsarwarchowdhury/2/download/. [Accessed 18th November 2019],

[3] [3] J. Tulsky, M. Beach, P. Butow, S. Hickman, J. Mack, R. Morrison, R. Street, D. White and K. Pollak, "A research agenda for communication between health care professionals and patients living with serious illness," JAMA internal medicine, vol. 177, no. 9, pp.1361-1366, 2017. [Online]. Available at: https://sci-hub.tw/https://jamanetwork.com/journals/ jamainternalmedicine / article-abstract/2635330#. [Accessed 19th November 2019],

[4] [4] C. Kuziemsky, "Decision-making in healthcare as a complex adaptive system. Healthcare management forum," The Canadian College of Health Leaders, vol. 29, no. 1, pp. 4-7, 2016. [Online]. Available at: https://sci-hub.tw/https://doi.org/ 10.1177/0840470415614842 #. [Accessed 20th November 2019],

[5] [5] K. Lakshmi, D. Ahmed and G. Kumar, "A smart clinical decision support system to predict diabetes disease using classification techniques," IJSRSET, vol. 4, no. 1, pp.1520-

1522, 2018. [Online]. Available at: https://s3.amazonaws.com/academia.edu.documents /57021280/3909.pdf?response-content-disposition=inline%3B%20filename%3 DA_Smart_ Clinical _ Decision_Support_System.pdf &X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20191203%2Fus-east-1%2Fs3%2 Faws4_request&X-Amz-Date= 20191203T074834Z&X-Amz-Expires=3600&X-Amz-Signed Headers = host&X -Amz-Signature=6ab09f94ad28907675bca9865b785e8c649 eecccad66 47 d5696fd2233fde7859. [Accessed 20th November 2019],

[6] Dr. Amit Sharma. Development of android application services at Arokia and its architecture. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1072-1075

[7] Vijay Malav, Dr. Amit Sharma. Effect and benefits of deploying Hadoop in private cloud. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1057-1062

[8] Dr. Amit Sharma. Implementing the design of service oriented architecture. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1027-1030

[9] L.Dhanabal, Dr. S.P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms", 2016, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6

[10] Wathiq Laftah Al-Yaseen, Zulaiha Ali Othmana, Mohd Zakree Ahmad Nazri, "Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System", 2015, Expert Systems With Applications

[11] Krishnanjali Magade and Dr. Amit Sharma "Prediction Accuracy On Automating Of Overnight Patient Care" Advanced Engineering Science ISSN: 2096-3246 Volume 54, Issue 02, August, 2022

[12] Mr. Vinayak, Dr. Rajendra Patil, Mr. Kajima Tambe, "Future of 5G wireless system ", Research Gate Special Issues: April 2021.

[13] Mr. Karan Patil, Mrs. Chetna Achar, "THE FUTURE OF 5G WIRELESS SYSTEM", IRJMETS|Volume:04,Issue:5, May-2022