International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 13 • Issue 1 pp. 46-51 Sept 2021 - March 2022 www.csjournals.com

Overview of Tools and Applications of Machine Learning in Healthcare

Jaswinder Singh
Dept. of Computer Science and Engineering
Guru Jambheshwar University of Science and Technology, Hisar, India
jaswinder_singh_2k@rediffmail.com

Abstract: Computer science includes the sciences of machine learning. This is the technology that have been studied the most and employed in many industries, such as healthcare, education, retail, etc., to lessen the need for human labour. It has been demonstrated that ML improve disease detection, identification, diagnosis, and treatment. Due to the climate and people's lifestyle choices, many diseases affect people nowadays. Hospitals keep a variety of records that are referred to as electronic health records. Manual data processing and disease diagnosis provide challenges. It is necessary to develop a procedure or approach to solve these issues. ML came in existence in order to address these issues. This study discuss about applications, tools and parameters of machine learning used in healthcare sector.

Keywords: Machine Learning, Applications, Tools, Health Sector

Introduction

Since its development, machine learning (ML) has experienced significant growth. It is employed in a variety of industries, including healthcare. The goal of machine learning (ML) is to discover or extract information from a dataset's unseen pattern. With usable knowledge, intriguing and exploitable ML algorithms can be created. This kind of knowledge can be used to improve the precise orientation of decision-making processes' efficiency and efficacy. Healthcare organizations generate and keep enormous volumes of information. A method of automating the production of this much data is necessary. ML approaches are frequently employed in the healthcare industry to make judgements, identify diseases, and provide patients with better care at a lower cost. ML is a cognitive method for extracting hidden patterns from a vast amount of data. The enormous data that is available can be used to retrieve crucial information and relate all aspects to make decisions. ML is needed to process that enormous amount of data efficiently. This paper's main goal is to give an outline of how ML might benefit the healthcare industry. Numerous methods, models, and algorithms in ML can be used to forecast disease, which is advantageous for both patients and medical professionals. These methods allow clinicians to diagnose their patients with less time and effort. By combining these methods, a hybrid model or method can be created to improve the effectiveness of the healthcare system. Models for diagnosis are selected based on specifications. Additionally, it benefits its organizations.

There are sections in this work. The comprehensive introduction to ML is given in section 1. In Section 2, some significant ML-based research in the field of healthcare has been discussed. Tools, Parameters and, Applications of ML in the healthcare industry have been covered in later sections.

Literature Review

In [1] R. Krishnamoorthi et al. (2022) anticipated a framework using four ML algorithms: "Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF)". The author also performed hyper-parameter tuning to enhance the performance of corresponding algorithms. Dataset used was Pima Indian Diabetic (PIDD) downloaded from "The University of California, Irvine (UCI) repository". Python was used for the implementation. The anticipated technique attained an accuracy of 83% with a minimal error rate.

In [2]D. Deepika and N. Balaji (2022) proposed an efficient heart prediction approach using "Multi-Layer Perceptron for Enhanced Brownian Motion based on the Dragonfly Algorithm (MLP-EBMDA)" and an optimized unsupervised technique used for feature selection. Dataset was taken from UCI. The proposed system had a 94.28% accuracy rate.

In [3]Suneetha Chittineni (2022) proposed a system to identify breast cancer using RF and Arithmetic optimization algorithms (AOA). The proposed system worked in different phases: Preprocessing phase to remove noise or unwanted data from the image, feature extraction using CNN, and Rat swarm optimization algorithm after that classification phase. The image dataset was obtained from mammographic image analysis society. Matlab was used for implementation. The suggested AOA-RF with the CNN-RSO system yields 100% accuracy.

In [4] Jan Carlo T. Arroyo and AllemarJhone P. Delima (2022) proposed an optimization system using ANN and Genetic Algorithm (GA) to predict heart disease. Dataset was picked up from Kaggle and Python was used for implementation. The result of the system showed an accuracy of 73.43%.

In [5] M. Nishat et al. (2021) discussed eight supervised ML algorithms: "Decision Tree (DT), Naive Bayes (NB), LR, KNN, SVM, RF, Multilayer Perceptron (MLP), and Quadratic Discriminant Analysis (QDA)" for the prediction of CKD. Dataset was occupied from UCI and Python was used for implementation. RF achieved an accuracy of 99.75%.

In [6]A. Sharma et al. (2021) discussed supervised ML algorithms: DT, NB, Artificial Neural Network (ANN), and LR for the prediction of Diabetes. The dataset was PIDD, and it was obtained from the "UCI machine learning repository." WEKA 3.8.4 was used for the experiment. LR performs better than others.

In [7]M. Durgadevi (2021) suggested a hybrid algorithm using Ant Miner and GA for the prediction of Diabetes. Ant Miner selected the best features and GA optimize the attributes. Dataset was taken from UCI. The proposed model's results were compared to those of existing techniques such as linear regression, LR, and DT. The experiment was carried out using WEKA and showed an accuracy of 98.92%.

In [8]Bonomalikhunti (2021) proposed a new classification technique using KNN and Standard deviation (SDKNN). In this study, distance calculation was based upon Standard Deviation of points. Dataset was taken PIDD acquired from UCI. Dataset was fragmented into 90% and 10% training and testing data respectively. The proposed technique showed an accuracy of 83.2%.

In [9]Parlar (2021) proposed an optimization technique based on Wolf Search Algorithm (WSA) for diagnosing Parkinson's disease. They used feature selection techniques viz. Information Gain (IG), ReliefF, and WSA along with the ANN model. Their result concluded that ReliefF performed better

than others for Parkinson's disease. The experiment was carried out using Cleveland dataset downloaded from UCI and performed on WEKA software.

In [10]Pramanik (2021) discussed Parkinson's detection method that consists of incremental decision trees and training instances. "Systematically Developed Forest, Decision Forest" by Penalizing Attributes, and Random Forest were used to construct these incremental decision trees. The experiment was conducted in two phases. Feature selection was performed in the first phase and in the second phase, training & analysis were conducted. Dataset was provided by "Department of Neurology in Cerrahpasa, Faculty of Medicine, Istanbul and Departamento de Matemáticas, Universidad de Extremadura, Cáceres, Spain."

Tools

In this section tools used by the researcher will be discussed. There are software, such as MATrix LABoratory, WEKA, Rapid Miner, KNIME, Apache Mahout, DataMelt, Sisense, and SSDT to implement Machine Learning techniques.

WEKA "Waikato Environment for Knowledge Analysis (WEKA)" provides the implementation of existing popular algorithms. The framework enables the development of numerous algorithms utilizing data extraction as well as the generation of algorithms from diverse applications using the Java programming language. WEKA includes tools for data pre-processing, regression, clustering, feature extraction, association rule generation, and visualization [11].

R-Studio It includes a set of designs that help in more productive with R and Python [12]. It comes with a console, a syntax-highlighting editor that enables direct code execution, and several powerful graphing, history-viewing, tracing, and workspace management capabilities.

MATLAB Mathematics Laboratory (MATLAB) helps in implementing Machine Learning algorithms. It offers toolboxes to implement algorithms by selecting values for the required parameters. If an end-user tries to update the algorithms, simply can write code in the editor window. It provides a toolbox for Classification, Regression, Clustering algorithms, and Data visualization. It executes programs faster than any open-source software for machine learning computation.

Python It helps in implementing ML algorithms with the help of built-in libraries and functions. Some components make Python an important language for machine learning. Libraries like Jupiter Notebook, NumPy, Pandas, and Scikit-learn help in implementing machine learning algorithms.

WEKA, R-Studio, and Python are open-source software whereas MATLAB is not open. Users must purchase this software for work.

Parameters

Several parameters are there, by which the performance of the ML algorithms can be evaluated. These parameters will be chosen according to the requirement of the experiment. Precision, Recall, F1, Accuracy, and ROC curve are used to analyse performance. To improve the performance, parameter tuning is performed. The process of identifying the suitable set of hyper-parameters for a learning algorithm is known as parameter tuning, also known as hyper-parameter optimization. The goal is to find an optimal combination of hyper-parameters to minimize a predefined loss function to deliver better outcomes.

Precision: It provides the answer regarding accurately classifying true positives. It is the ratio of True Positive and Actual Result as mentioned 1.

$$\text{Precision} = \frac{\text{True Positives}}{\text{Actual Results}} \quad (1)$$

Recall: How many important findings are extracted, and identified by using this metric [13]? It is the ratio of True Positive and Predicted Results as shown in 2.

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad (2)$$

F1-Measure :"F1 score, also known as the F score or F measure", is a metric for determining the accuracy of a test. It is the weighted harmonic mean of the precision and recall of the test [14]. The mathematical formula for F1-Measure is shown in 3.

$$\text{F} - \text{Measure} = \frac{2*(\text{Recall} + \text{Precision})}{\text{Recall} + \text{Precision}} \quad (3)$$

Accuracy: The number of accurate observations is divided by the total number of input samples. It only works well if the same number of observations resides in each class [15].

Area Under the ROC Curve (AUC): It is a graphical representation that recapitulates the intact location of the ROC curve rather than subject to a specific operating point. "The ROC curve is generated by plotting the true positive rate versus the false positive rate at various thresholds" [16].

Applications

Following are a few applications of machine learning. Predictions: It is one of the main concerns of research in ML. Using such information and behavior, one can accurately predict results[21].

1) Medical Sector: In the field of medicine, ML helps with predictions, reviews, and categorisation. The disease is categorized as either prevalent or dangerous[17].

2) Banking and Stock Market: ML could be very important for stock price forecasting. Prediction will help you in significantly prevent deficits on the stock market. The foundation of ML algorithms is supervised learning. The algorithm develops its model using the earlier data [18].

3) Speech Recognition: A popular ML application is speech recognition. This method makes advantage of linguistically sound models. The sound model analyses and eliminates words based on the sounds it hears. The linguistic model tries to match the sound of the word with similar terms[19].

4) Image Recognition: Image recognition is one of the most widely used applications of ML. The algorithms employ a variety of grouping and classification techniques. This is how they teach the model to distinguish among two images. It has numerous uses both in regular life and in unique circumstances [20].

Conclusion

ML is a rapidly expanding study with applications in many fields, including healthcare. Studies demonstrate that the medical industry is continually changing. Making informed decisions that enhance the patient's health may be possible with the use of information obtained utilizing ML

techniques. With the impending advancement of information and communication technology, ML will realize their full potential in uncovering previously undiscovered information in medical data. An overview of the applications, tools, and parameters used in the healthcare sector is provided in this study.

References:

[1] R. Krishnamoorthi et al., "A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques," Journal of Healthare Engineering, vol. 2022, 2022.

[2] D. Deepika and N. Balaji, "Effective heart disease prediction using novel MLP-EBMDA approach," Biomed. Signal Processing and Control, vol. 72, no. PB, p. 103318, 2022, doi: 10.1016/j.bspc.2021.103318.

[3] S. Chittineni and S. S. Edara, "Automated breast cancer detection system from breast mammogram using deep neural network," Indones. J. Electr. Eng. Comput. Sci., vol. 25, no. 1, pp. 580–588, 2022, doi: 10.11591/ijeecs.v25.i1.pp580-588.

[4] J. C. T. Arroyo and A. J. P. Delima, "An Optimized Neural Network Using Genetic Algorithm for Cardiovascular Disease Prediction," J. Adv. Inf. Technol., vol. 13, no. 1, pp. 95–99, 2022, doi: 10.12720/jait.13.1.95-99.

[5] M. Nishat et al., "A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms," EAI Endorsed Trans. Pervasive Heal. Technol., no. August, p. 170671, 2018, doi: 10.4108/eai.13-8-2021.170671.

[6] A. Sharma, K. Guleria, and N. Goyal, "Prediction of Diabetes Disease using Machine Learning Model," in Lecture Notes in Electrical Engineering (2021) 733 LNEE 683-692, 2021, no. March, doi: 10.1007/978-981-33-4909-4.

[7] M. Durgadevi, "Disease Prediction Model for Diabetes Mellitus Using Ant Miner – Genetic Algorithm," Psychol. Educ. 58(2) 10149-10155, vol. 58, pp. 10149–10155, 2021.

[8] Bonomali khunti, "Analysis and Prediction Of Pima Indian Diabetes Dataset Using SDKNN Classifier Technique," IOP Conf. Ser. Mater. Sci. Eng. 1070(1) 012059, 2021, doi: 10.1088/1757-899X/1070/1/012059.

[9] T. PARLAR, "A heuristic approach with artificial neural network for Parkinson's disease," Int. J. Appl. Math. Electron. Comput., vol. 9, no. 1, pp. 1–6, 2021, doi: 10.18100/ijamec.802599.

[10] M. Pramanik, R. Pradhan, P. Nandy, A. K. Bhoi, and P. Barsocchi, "Machine learning methods with decision forests for parkinson's detection," Appl. Sci., vol. 11, no. 2, pp. 1–25, 2021, doi: 10.3390/app11020581.

[11] "Weka 3 - Data Mining with Open Source Machine Learning Software in Java." https://www.cs.waikato.ac.nz/ml/weka/ (accessed Jul. 11, 2021).

[12] "Download the RStudio IDE - RStudio." https://www.rstudio.com/products/rstudio/download/ (accessed Jul. 11, 2021).

[13] P. Ariza-colpas and R. Morales-ortega, "Parkinson Disease Analysis Using Supervised and Unsupervised Techniques," vol. 1, pp. 191–199, 2019, doi: 10.1007/978-3-030-26354-6.

[14] "Accuracy, Precision, Recall or F1? | by Koo Ping Shung | Towards Data Science." https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 (accessed Jul. 05, 2021).

[15] "Metrics to Evaluate your Machine Learning Algorithm | by Aditya Mishra | Towards Data Science." https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 (accessed Jul. 05, 2021).

[16] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recognit., vol. 30, no. 7, pp. 1145–1159, 1997.

[17] I. Kononenko, "Machine Learning for Medical Diagnosis : History , State of the Art and Perspective,"Artificial Intelligence in Medicine, vol.23, Issue.1, pp. 89-109, 2001.

[18] K. Kangra and J. Singh, "Exploring the Potential of Machine Learning and Artificial Intelligence in HealthCare," J. Maharaja Sayajirao Univ. Baroda, no. July, 2021.

[19] T. M. Mitchell, "The Discipline of Machine Learning," Mach. Learn., 2006, doi: 10.1080/026404199365326.

[20] S. Das and N. Roy, "Applications of Artificial Intelligence in Machine Learning : Review and Applications of Artificial Intelligence in Machine Learning : Review and Prospect," no. May 2018, 2015, doi: 10.5120/20182-2402.

[21] S. Ellahham, N. Ellahham, M. Can, and E. Simsekler, "Application of Artificial Intelligence in the Health Care Safety Context : Opportunities and Challenges," 2019, doi: 10.1177/1062860619878515.