IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 7, March 2024

Enhancing Medical Diagnosis through Machine Learning: A Review

Sudhir Kumar Thakur[1] and Dr. Savya Sachi[2]

Research Scholar, Department of Computer Science and Engineering[1]

Assistant Professor, Department of Computer Science and Engineering[2]

Rajiv Gandhi Proudyogiki Mahavidyalaya, Bhopal, M.P, India

Abstract: *The medical industry is seeing remarkable increase in computer-aided detection using machine learning (ML) and deep learning (DL). The true source of the relevant data needed for a disease diagnosis is thought to be medical pictures. One of the most critical aspects of reducing the death rate from cancer and tumours is early detection of the disease through the use of several modalities. Radiologists and medical professionals can better understand the internal anatomy of a discovered disease by using modalities to extract the necessary features. Due to big data sets, ML is limited by current modalities, while DL functions well with any volume of data. DL is therefore seen as an improved method of machine learning (ML), in which ML employs learning strategies and DL gathers information on how machines should react around people*

Keywords: Machine Learning, Deep Learning, Disease Diagnosis, Computer-Aided Detection, Cancer, Tumours

I. INTRODUCTION

From prior years, it is evident how important disease classification and prediction are. To pinpoint the precise source of an illness as well as its symptoms, it is imperative to be aware of the crucial characteristics and attributes provided in a dataset. Classification and decision support are two areas where artificial intelligence (AI) has demonstrated encouraging outcomes. A subset of artificial intelligence called machine learning (ML) has sped up a lot of medical research. On the other hand, DL, a subset of ML, works with neural network layers and analyses the precise characteristics needed for disease identification. The research that has been done so far, from 2014 to the present, covers a wide range of applications and algorithms designed to improve the medical profession by giving patients reliable findings. Through data, ML has fuelled breakthrough advances in computer vision, automatic speech recognition, natural language processing, and other fields to create reliable systems like automated translation and driverless vehicles. Medical applications of machine learning continued to face risks despite all the advancements. Many of these problems arose from medical treatment, where the aim was to use the data gathered and the medical systems management to make precise forecasts. In order to extract the necessary features or highlights from a vast amount of data, AI analyses a particular dataset using a variety of methodologies. This makes it challenging to locate the best possible arrangement of important features and to eliminate redundant ones. Accuracy measures become inaccurate when such features are taken into account, which is uncomfortable. Therefore, selecting a limited subset of attributes from a large list will improve the model's effectiveness. The removal of cumbersome and repetitive features would therefore reduce the dimensionality of the data and accelerate the learnt model in a manner akin to boosting. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two useful techniques that are used to extract the significant features from the existing features. Selecting a feature in particular has two crucial goals: first, it improves the arrangement's presentation; second, it reduces the number of features needed to solve the dimensionality problem. Thus, choosing characteristics is seen as a crucial step towards achieving the aforementioned goals. Subsequently, choice-based multi-target strategies were used to increase study on features improvement. Therefore, the tactics for selecting effective features will be the main emphasis of this review. Multiple methods were used to identify cancer illness. methods for segmenting images, choosing features, and performing regression using Root Mean Square Error (RMSE), including parameters for object detection, pattern recognition, and picture classification. Using six classifiers and Transfer Learning (TL) algorithms for picture segmentation with brain magnetic resonance imaging (MRI), a brain tumour was identified.

II. STATISTICAL ANALYSIS OF ML AND DL TECHNIQUES USED FOR MEDICAL IMAGE DIAGNOSIS

(i) Machine learning- A subset of artificial intelligence called machine learning (ML) uses data to find patterns and make decisions automatically with little to no human input. The most crucial qualities of a machine learning model are its ability to adapt on its own, learn from past computations, and generate accurate results when repeatedly applied to fresh datasets. The two main aspects are (i) machine learning techniques that enable doctors to quickly and accurately interpret medical images using computer-aided design (CAD); and (ii) algorithms that are used for complex tasks such as brain tumour segmentation with MRI, breast cancer and mammography, and CT scan segmentation. The methods used in traditional machine learning models, which operated on structured datasets, were specified for each step and would fail if any of Steps were not followed. It is crucial to assess the quality of the data that ML and DL algorithms employ. On the other hand, new algorithms modify the data that is omitted according to the algorithm's robustness requirements. The method by which ML algorithms anticipate and diagnose diseases is shown in Figure 1.

(ii) Deep learning- Because medical image analysis techniques have advanced, DL models allow machines to reach the accuracy. Using the labelled chest X-rays, the heart illness was identified in. The cardiologist went over and relabelled all the data, removing the images that were not heart failure or normal. Data augmentation and TL were utilised with 82% accuracy, 74% specificity, and 95% sensitivity for heart failure to extract the precise information from the images. With the least amount of human labour, an autonomous feature selection system was created in utilising histopathological images labelled as positive and negative cancer images. While a single-layer network of K-means centroids was utilised for unsupervised feature learning, two networks, dubbed Deep Neural Network (DNN) 2-F and DNN1-F, were utilised with PCA to reduce features in DNN was applied. Afterwards, a comparison was made between the outcomes of supervised (94.52%) and uncontrolled (93.56%) learning. To manage data effectively, the DL model automates the feature extraction process. The method by which DL algorithms anticipate and diagnose different diseases is shown in Figure 2.

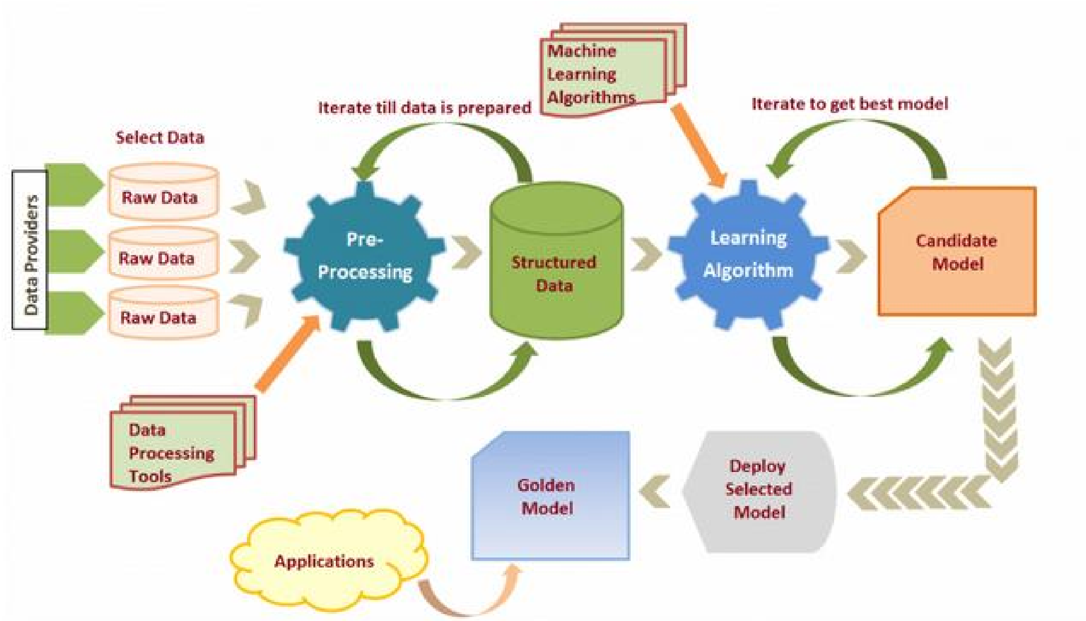

Figure 1- ML process

Figure 2- DL process

III. ML AND DL TECHNIQUES FOR MEDICAL IMAGING

It provides an overview of the current techniques of ML and DL for medical imaging. Further, followed by various parameters considered for selecting the classifiers and the evaluation metrics used to evaluate classification models. The existing literature review is divided according to the diseases such as breast cancer, brain tumor, lung disease, diabetes, multiple disease detection, etc.

(i) Breast disease- this articles related to breast disease symptoms, detection, classification, prediction and diagnosis using ML and DL methods are discussed. In, significant features were identified using BI-RADS (Breast Imaging Reporting and Data System) to develop a CAD system for obtaining breast ultrasound. Also, 10-fold cross validation technique was used upon the benign and malignant lesions. As a result, 77% accuracy was achieved using the SVM classifier. However, some methods with a few algorithms handling the vast variety of data need to be understood and analyzed precisely. CNN was used to train the system with the available clinical data and to comprehend the complex structure. Moreover, it was suggested to study radiomics and expansion of CADx to get the tumor signs using a CAD system. Breast cancer disease was classified using the parameters like Area Under Curve (AUC), sensitivity, and specificity. A CAD system was developed using CNN where a large number of features were required, using multiview features. These features provide the maximum details of the image data to be extracted for the accuracy of detection and classification.

(ii) Brain disease- The concept of TL was used for image segmentation where the MRI scan of the brain was segmented using voxel wise classification. ML classifiers were applied for the classification of multiple diseases. Later, the results obtained were compared with the existing results to detect the disease. A brief introduction of DNN in medical image analysis to diagnose the brain tumor using brain tissues is provided in. It indicated the ways for applying DL to the entire process of MRI scanning, image retrieval, segmentation and disease prediction. It also focused on image acquisition to image retrieval, and from feature segmentation to prediction of disease. The entire process was divided into two parts: (i) the signal processing of MRI including the image restoration and image registration, and (ii) usage of DL for disease detection and prediction- based reports in the form of text and images. Also, the influence of DL in medical imaging was discussed in. Image segmentation approaches using DL included tumor segmentation, brain and lung's structure with bone tissues or cells. Patches were taken as input and 2- Dimensional Convolutional Neural Network (2D-CNN) was used to preprocess these at a later stage.

(iii) Lung disease- DL has the ability to automate the process of image interpretation which enhances the clinical decision making, identifying the disease and predicting the best treatment for the patient by reviewing the pros and cons of the DL techniques. These techniques were used for the cardiovascular medication, following are the steps for implementing DL model: (i) problem identification, (ii) data selection, (iii) hardware and software selection, (iv) data preparation, (v) feature selection, and (vi) splitting of data for training as well as validation process. In, a disease was analyzed automatically using labeled data and achieved the accuracy by processing medical images using DL models. The automatic prediction of the disease using ML techniques and the concept of big data was summarized to detect the patterns. The advantages and disadvantages for each algorithm were also discussed.

(iv) Diabetes- A comparative analysis of the classification algorithms based on iris images, using an iridology chart, was done for the diagnosis of diabetes. Type-2 diabetes was detected by identifying the center of the pupil of an eye at the early stage using the I-Scan-2. Also, a filter-based feature selection method was used with the combination of five classifiers namely binary tree, SVM, neural network model, Random Forest (RF) and adaptive boosting model. Later, in a study was compiled using the textural, statistical and various features (62 features of iris) to detect the same disease, however, an iridology chart was not used. ML and DL techniques were used to diagnose the errors in existing diagnostic systems. These techniques were used to analyze the medical images and extract the features which are required for the diagnosis of errors in existing diagnostic systems. Both supervised and unsupervised algorithms were used for the prediction of the disease in specific datasets.

(v) Multiple disease detection- Multiple diseases were identified with different radiology techniques like MRI imaging for breast cancer along with brain tumor, CAD for breast cancer along with skin lesions, and X- Rays for chest analysis. Also, ML techniques were used to attain better accuracy with denoising techniques including homographic wavelet, soft thresholding, non-homomorphic and wavelet thresholding. A CAD system using CNN was proposed to diagnose breast lesions as benign and malignant to assist the radiologists. It was implemented using Inception- v3 architecture to extract the multiview features from Automated Breast Ultrasound (ABUS) images. For the implementation of the model, 316 breast lesions data were trained and evaluated. ML feature extraction scheme was compared with the given method, resulting in 10% increase in AUC value.

IV. CONCLUSION

This paper presents an overview of different machine learning (ML) and deep learning (DL) approaches for illness detection, covering categorisation, imaging modalities, tools, algorithms, datasets, and medical domain issues. The most often utilised modalities for diagnosing diseases are MRI and X-ray scans. Furthermore, MATLAB and SVM dominated, respectively, among all the tools and techniques tested. It was noted that researchers make extensive use of the MRI dataset. Additionally, a comparison of ML classifiers and DL models using the MRI dataset has been made possible through a series of trials, in which CNN (97.6%) and RF (96.93%) have surpassed other approaches. According to this study, denoising techniques with DL models should be used in the healthcare industry. Furthermore, it concludes that different traditional ML and DL Strategies are often used to address data uncertainty. DL techniques have gained a lot of traction among researchers recently because of their improved performance. With the help of this review, members of the medical community, doctors, and clinicians will be able to diagnose diseases more quickly and accurately by selecting the best ML and DL technique.

REFERENCES

[1]. Abdulbaqi AS, Younis MT, Younus YT, Obaid AJ (2022) A hybrid technique for EEG signals evaluation and classification as a step towards neurological and cerebral disorders diagnosis. Int J Nonlinear Anal Appl 13(1):773–781

[2]. Aceto G, Persico V, Pescapé A (2018) The role of information and communication technologies in healthcare: taxonomies, perspectives, and challenges. J Netw Comput Appl 107:125–154. https://doi.org/10.1016/j.jnca.2018.02.008

[3]. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M (2017) Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inf Sci 415:190–198.https://doi.org/10.1016/j.ins.2017.06.027

[4]. Aggarwal 1LP (2019) Data augmentation in dermatology image recognition using machine learning. Skin Res Technol 25(6):815–820. https://doi.org/10.1111/srt.12726

[5]. Al-Najdawi N, Biltawi M, Tedmori S (2015) Mammogram image visual enhancement, mass segmentation and classification. Appl Soft Comput 35:175–185. https://doi.org/10.1016/j.asoc.2015.06.029

[6]. Altan G, Kutlu Y, Allahverdi N (2019) Deep learning on computerized analysis of chronic obstructive pulmonary disease. IEEE J Biomed Health Inf 24(5):1344–1350. https://doi.org/10.1109/JBHI.2019.2931395

[7]. Anbeek P, Vincken KL, Van Bochove GS, Van Osch MJ, van der Grond J (2005) Probabilistic segmentation of brain tissue in MR imaging. NeuroImage 27(4):795–804. https://doi.org/10.1109/TMI.2014.2366792

[8]. Arya R, Kumar A, Bhushan M (2021) Affect recognition using brain signals: a survey. In: Computational methods and data engineering. Springer, Singapore, pp 529–552. https://doi.org/10.1007/978-981-15-7907-3_40

[9]. Arya R, Kumar A, Bhushan M, Samant P (2022) Big five personality traits Prediction using brain signals. Int J Fuzzy Syst Appl (IJFSA) 11(2):1–10. https://doi.org/10.4018/IJFSA.296596

[10]. Available at: https://techblog.cdiscount.com/a-brief-overview-of-automatic-machine-learning-solutions-automl/

[11]. Available at: Noisy Data in Data Mining | Soft Computing and Intelligent Information Systems (ugr.es)

[12]. Beam AL, Kohane IS (2018) Big data and machine learning in health care. JAMA 319(13):1317–1318.

[13]. Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A (2021) The state of the art of deep learning models in medical science and their challenges. Multimedia Syst 27(4):599–613. https://doi.org/10.1007/s00530-020-00694-1

[14]. Bhattacharya P, Tanwar S, Bodkhe U, Tyagi S, Kumar N (2019) Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications. IEEE Trans Netw Sci Eng 8(2):1242–1255

[15]. Bhattacharyya A, Pachori RB, Upadhyay A, Acharya UR (2017) Tunable-Q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals. Appl Sci 7(4):385

[16]. Bhushan M, Goel S (2016) Improving software product line using an ontological approach. Sādhanā 41(12):1381–1391

[17]. Bhushan M, Goel S, Kumar A, Negi A (2017) Managing software product line using an ontological rule-based framework. In: 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS). IEEE, pp 376–382

[18]. Bhushan M, Goel S, Kaur K (2018) Analyzing inconsistencies in software product lines using an ontological rule-based approach. J Syst Softw 137:605–617

[19]. Bhushan M, Goel S, Kumar A (2018) Improving quality of software product line by analysing inconsistencies in feature models using an ontological rule-based approach. Expert Syst 35(3):e12256

[20]. Bhushan M, Negi A, Samant P, Goel S, Kumar A (2020) A classification and systematic review of product line feature model defects. Software Qual J 28(4):1507–1550

[21]. Bhushan M, Kumar A, Samant P, Bansal S, Tiwari S, Negi A (2021) Identifying quality attributes of FODA and DSSA methods in domain analysis using a case study. In: 2021 10th international conference on System Modeling & Advancement in Research Trends (SMART). IEEE, pp 562–567

[22]. Bhushan M, Duarte JÁG, Samant P, Kumar A, Negi A (2021) Classifying and resolving software product line redundancies using an ontological first-order logic rule based method. Expert Syst Appl 168:114167

[23]. Caballé-Cervigón N, Castillo-Sequera JL, Gómez-Pulido JA, Gómez-Pulido JM, Polo-Luque ML (2020) Machine learning applied to diagnosis of human diseases: a systematic review. Appl Sci 10(15):5135.https://doi.org/10.3390/app10155135

[24]. Cabitza F, Rasoini R, Gensini GF (2017) Unintended consequences of machine learning in medicine. JAMA 318(6):517–518. https://doi.org/10.1001/jama.2017.7797

[25]. Caliskan A, Badem H, Basturk A, YUKSEL M (2017) Diagnosis of the parkinson disease by using deep neural network classifiers. IU-J Electr Electron Eng 17(2):3311–3318

[26]. Chaganti SY, Nanda I, Pandi KR, Prudhvith TG, Kumar N (2020) Image classification using SVM and CNN. In: 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE, pp 1–5

[27]. Currie G, Hawk KE, Rohren E, Vial A, Klein R (2019) Machine learning and deep learning in medical imaging: intelligent imaging. J Med imaging radiation Sci 50(4):477–487. https://doi.org/10.1016/j.jmir.2019.09.005

[28]. De Bruijne M (2016) Machine learning approaches in medical image analysis: from detection to diagnosis. Med Image Anal 33:94–97. https://doi.org/10.1016/j.media.2016.06.032