Lungs Cancer Prediction Model Using Machine Learning Techniques

Muhammad Wasim[1,] Uzair Ahmad[2], Nasir Abbas[3], Shahzad Amir[4], Komal Arshad[5], Muhammad Rizwan Ullah Siddiqui[6]

[1]Bahauddin Zakariya University (BZU), Multan, Pakistan

[2]National College of Business Administration & Economics

[3]WHO PEI Lahore

[4,5,6]Department of Health, Okara

Corresponding Author: Muhammad Wasim, Email: arainwasim1@gmail.com

***Abstract*:-**The obvious reason for malignant growth-related mortality all through the globe is "Cellular breakdown in the lungs". As a result, early detection of cellular breakdown in the lungs including location, anticipation, and finding has become crucial as the clinical board that follows accelerates and gets better. This study presents a model for the cellular breakdown in the lungs expectation by using Machine Learning (ML) methods. To raise the advancement and drug of carcinogenic circumstances, ML methods have been used because of their precise results. Different kinds of AI calculations (ML) like Naive Bayes, Support Vector Machine (SVM), Logistic relapse, and Artificial Neural Networks (ANN), have been applied in the medical services area for examination and anticipation of cellular breakdown in the lungs. In this research, the proposed model achieves more than 95% accuracy by using ML techniques and in the future, this model will also apply to predict other diseases with better performance.

Keywords: - *Lung Cancer, Prediction Model, Machine Learning Techniques*

1 Introduction

The National Lung Screening Trial (NLST) in the USA, showed a 20% decrease in the cellular breakdown in lung mortality, and the resulting choice through the U.S. Habitats for Medicare as well as Medicaid Facilities to cover cellular breakdown in the lungs screening under Medicare have made ready for the boundless cellular breakdown in the lungs separating the USA. The significance of low portion processed tomography (LDCT) in the early recognizable proof of cellular breakdown in the lungs was additionally featured by this choice. However, the moderately high misleading positive pace of LDCT-based screening is one of the known downsides. For example, within the initial double rounds of the LDCT limb, the pace of positive assessment tests was practically 27%, and in the 3rd year of assessment, it was 17%. A screening CT was thought positive if it showed any troubling irregularities or a non-calcified knob with a long hub of something like 4 mm. All through 96% of these positive screens throughout the three rounds were bogus up-sides, and 72 percent required demonstrative development (Kadir and Gleeson, 2018)

The American College of Radiology (ACR) Lung Imaging Reportage, as well as Information Scheme (Lung RadsTM) device (2) for normalized detailing of CT-based cellular breakdown in the lungs evaluating, embraced a limit for strong knobs of six mm for its classification two where no extra demonstrative effort-up is suggested as well as the subject is imaged once more at yearly separating request to resolve this issue. A six-month follow-up LDCT is suggested for new nodules that are 4 mm or larger, though, due to the fact that they have a higher likelihood of being malignant(Alam et al., 2018).

The effect of Lung-RADS was broken down in a review analysis of the NLST (3). Lung-RADS remained displayed to decrease the general screening bogus plus rate to 12.8% as well as 5.3% at standard as well as stretch imaging separately on the expense of a decrease of responsiveness after 93.5% in the NLST to 84.9% utilizing Lung-RADS on the benchmark as well as 93.8% in the NLST as well as 84.9% utilizing Lung-RADS later pattern. Nonetheless, even though Lung-RADS diminishes the in general misleading plus a percentage, the bogus positive pace of positive screens, i.e., Lung-RADS three or more, continues extremely elevated by 93% at standard as well as 89% when benchmark; of 3,591 Lung-RADS three or more panels, 3,343 remained bogus up-sides on the pattern as well as of 2,858 Lung-RADS three or more panels when gauge 2,543 have been misleading up-sides. Subsequently, though the reception of Lung-RADS container decreases the all-out quantity of harmless knobs existence stirred up inside a broadcast program, at an expense of fair shy of 10% misfortune in responsiveness, here stay an extremely huge number of harmless knobs being examined, and the knob characterization task stays a difficult one. Taking on PC helped finding (CADx) innovation as a device for radiologists and pneumonic medication specialists is one method for managing this issue. Such techniques attempt to convey a quantifiable result connected with the gamble of cellular breakdown in the lungs given an information CT and possibly extra relevant patient subtleties. One could say that such systems have two objectives in mind(Awai et al., 2004).

The first goal is to reduce the heterogeneity in how various interpretive doctors assess and report the risk of lung cancer. Computer-aided techniques have been demonstrated to improve physician dependability in a number of clinical settings, including mammography screening and nodule detection, thus it stands to reason that such decision support techniques might have a similar beneficial impact on nodule classification. Second, CADx may improve classification performance by supporting physicians who lack specialist knowledge or experience in assessing the risk that a particular lesion will progress to cancer. In this study, we examine the advancements made in the creation and validation of nodule categorization CADx software and lung cancer prediction models. While we do not intend for this to be an exhaustive analysis, we do want to give an overview of the key strategies used up to this point and list some of the obstacles still standing in the way of bringing this technology into widespread clinical use(Awai et al., 2004; Khanday et al., 2020)

2 LITERATURE REVIEW

An efficient approach employing SVM for lung cancer detection, diagnosis, and the ability to predict the probability of lung cancer is described in Jane et al.'s suggested "multi-stage lung cancer detection and prediction using multi-class SVM classifier." The Grey Equal Cooccurrence Technique (GLCM) procedure is utilized to apply an algorithm that utilizes picture managing methods such as image improvement, recognition as well as separation, detection, as well as quality extraction. SVM is utilized to make up classifications. Using the binarization method, forecasts are made. A database after UCI ML is bought, as well as it contains five hundred infected as well as five hundred uninfected CT pictures. Out of 130 photos, the suggested approach found 126 as being contaminated. as well as out of 100 previously designated photos, 87 were expected to be malignant. According to the experimental analysis, identification accuracy was 97 percent and prediction accuracy was 87 percent (Ausawalaithong et al., 2019).

A Computer-Aided Diagnosis Method for Detection of Lung Cancer Nodules Utilizing Intense Learning Machine was defined by M. Gomathi et al. Used for the study of disease in CT pictures, a CAD model is made in this paper. Establishing the area of importance in the feedback CT pictures is the primary step in CAD. Lung region segmentation comes before lung region withdrawal, as well as the Fuzzy Chance C mean cluster procedure is used to find cancer nodules. The analytical laws are produced utilizing a full Drawable Sphere strength rate. Next, using the ELM's assistance, these rules are put into practice to aid in learning (Kyzalas et al., 2019).

The effect of Lung-RADS remained broken down during a review examination of the NLST (3). Lung-RADS remained displayed decrease the general broadcast bogus optimistic amount to 12.8% as well as 5.3% at standard as well as stretch imagination separately on the expense a decrease of responsiveness after 93.5% into the NLST to 84.9% utilizing Lung-RADS on the benchmark as well as 93.8% in the NLST as well as 84.9% utilizing Lung-RADS later pattern. Nonetheless, even though Lung-RADS diminishes in the general misleading optimistic value, the bogus optimistic pace of optimistic screens, i.e., Lung-RADS three or more, continues extremely elevated at 93% at standard as well as 89% when benchmark; of 3,591 Lung-RADS three or more panels, 3,343 remained bogus up-sides a pattern as well as of 2,858 Lung-RADS three or more panels later gauge 2,543 remained misleading up-sides. Subsequently, though the response of Lung-RADS can decrease the all-out amount of harmless knobs actuality stirred awake inside a broadcast program, at an expense of fair shy of 10% misfortune in responsiveness, here stay an extremely huge number of harmless knobs being examined, and the knob characterization task stays a difficult one. Taking on PC helped to find (CADx) innovation as a device for radiologists and pneumonic medication specialists is one method for managing this issue. Such techniques attempt to convey a quantifiable result connected with the gamble of cellular breakdown in the lungs given an information CT and possibly extra relevant patient subtleties (Majeed et al., 2020).

Authors suggested “Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system” The suggested technique is done in 5 stages; a few locales are portioned by utilizing the k-implies grouping strategy. After portioning the lung bends, by utilizing pre-handling calculations dubious locales are fragmented from that of lung districts that produce a twofold picture which comprises of openings because of the division interaction, tidal wave - the full calculation is utilized to fill this opening as knob up-and-comer is considered as strong articles. This framework might contain general districts and lung knobs including veins. To isolate this handle, Law-established classifiers together with 2D and 3D characters are utilized. In the long end, the deceptive plus things are recognized using Linear Discrimination Analysis (LDA). The planned procedure remained separated on a dataset that integrates 1454CT pictures collected after 34 decided patient role to have 63 lung handles(Causey et al., 2019)

The authors portrayed a structure used for assessing the CAD impact on radio therapists' pneumonic handles classification. The proposed approach involved image handling procedures for lung as well as intrapulmonary construction classification dim level limit, 3D marking methods as well as statistical morphological procedures that remain utilized used designed for lung division. Designed for a group of intrapulmonary shapes Top-cap shift strategy is used, on an information picture to distinguish the curved picture. Strainer station is rummage-sale the recognizable proof of essential likely knobs, formerly underlines of aspiratory knobs are extricated to distinct honest handles after that of misleading optimistic buttons, ANN is attuned to conclude the likelihood of district of attention considering an image contain. Revision of this structure upgraded aspiratory knob inhabitant's finding of CT checks (Alam et al., 2018).

Cheran et al proposed that "PC helped to find for lung CT using counterfeit life models" This CAD model is made by sending various computations. In any case, the ribcage still hanging out there by using a 3D district creating computation, and a short time later the unique shape method is executed to cultivate an exact locale for the imminent underground bugs which are reallocated to encourage an exact as well as exact changing of the vascular tree as well as pleura. To recuperate the bronchial as well as the vascular trees fake living models are utilized. By utilizing dynamic structure models, it is settled whether the late-built wings contain handles as well as moreover to recognize whether the handles are related to the pleura. By using snakes and bit improvement estimation better computation is made to restrict the handles (Cheran and Gargano, 2005).

Lakshmana Prabu et al fostered the Optimal Deep Neural Network as well as Linear Discriminate Analysis and established the order pattern used for CT pictures. Lung knob order is finished by utilizing LDR as well as advancement is finished by utilizing the Revised Gravitational Look for Procedure to anticipate cellular breakdown in the lungs. Standard CT information base is utilized for the trial examination which involves 50-low measurement cellular breakdown in the lungs CT pictures. This paradigm corresponds to the present model like KNN, NN, DNN SVM, etc., as well as the prosecution investigation illustrates good quality outcomes intended for the created standard with 94.56% precision, responsiveness, as well as explicitness 96.2% as well as 94.2% individually (Khan et al., 2021).

Worawate et al fostered a strategy "Programmed Lung Cancer Prediction from Chest X-beam Images utilizing Deep Learning Approach", Writers utilized DensNet-121 (121 layers Convolutional brain organization) related to deep learning for grouping utilizing chest pictures. The model is prepared on double datasets i.e., Chest X-beam 14 as well as JSRT to distinguish the knobs. The model acquired an exactness of around 74.43±6.01%, responsiveness, as well as particularity of around 74.68 ±15.33% as well as 74.96 ± 9.85% individually(Ghoshal and Tucker, 2020).

Christoph et al suggested an expectation model considering tomography cellular breakdown in the lung's pictures. A CNN is used for highlight removal by tweaking pre-prepared ResNet18 as well as multimodal highlights CNN is prepared with the Cox standard for danger expectation. Lung1 dataset is utilized for trial investigation which let go be gotten to after "The Cancer Imaging Archive" (TCIA) that involves 422 NSCLC pictures for 318 of 422 affected roles (Majeed et al., 2020).

Most of the approaches have been used while employing and constructing several smart as well as intelligent frameworks like machine learning approaches(Ali et al., 2022, 2021; Ali Raza et al., 2022; Asif et al., 2021; Aslam et al., 2021; Chayal and Patel, 2021; Dekhil et al., 2019; Fatima et al., 2020; Ghazal et al., 2022; Khan et al., 2021; Muneer and Rasool, 2022; Saleem et al., 2022), Fuzzy Inference systems (Areej et al., 2019; Asadullah et al., 2020; Ihnaini et al., 2021; Saleem et al., 2019), Particle Swarm Optimization (PSO) (Iqbal et al., 2019), Fusion based approaches(Gai et al., 2020; Ma et al., 2020; Muneer and Raza, 2022; Sharma et al., 2021; Tabassum et al., 2021; Taher M. Ghazal, n.d.) ,cloud computing (Bukhari et al., 2022; Khan, 2022; Naseer, 2022; Siddiqui et al., 2021; Ubaid et al., 2022), transfer learning(Abbas et al., 2020) and MapReduce(Asif et al., 2021) that may require help in designing emerging solutions for the rising challenges in designing smart cloud-based monitoring management systems.

3 PROPOSED METHODOLOGIES

Cellular breakdown in the lungs is the significant reason for disease-related demise at this age, and it is normal to remain so for years to come. It is possible to treat cellular breakdown in the lungs assuming the side effects of the sickness are recognized early. It is feasible to build a supportable model for the therapy of cellular breakdown in the lungs involving ongoing improvements in computational knowledge without adversely affecting the climate. Since it will lessen the number of assets squandered as well as how much work important to finish manual jobs, it will set aside both time and cash. In this research, a machine learning approach-based framework is being designed for predicting lung cancer in the patient. This proposed framework may provide a better quality of life to the patients to maintain their health.

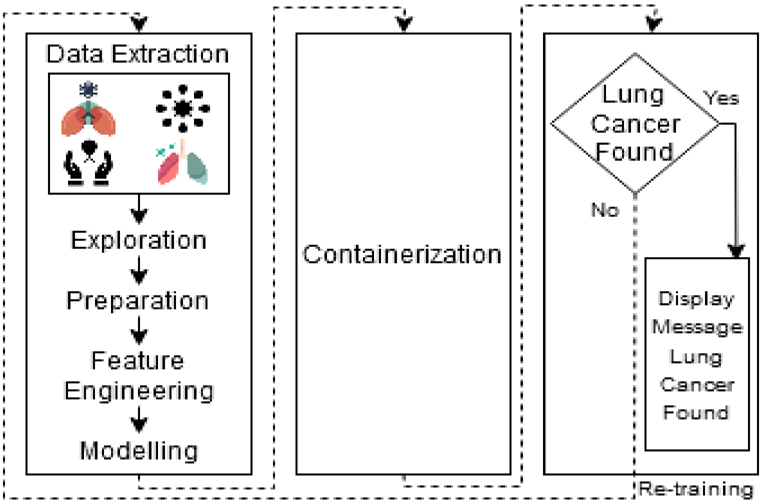

Figure 1: Proposed Model

It is shown in figure 1 that the suggested model is classified into three-step, where the first step is lung cancer data extraction is the process of collecting or retrieving disparate types of data from a variety of digital medical devices, many of which may be poorly organized or completely unstructured. After the extraction of data, the data is forwarded for preparation, which is the system of collecting, integrating, structuring, and managing information so it may be utilized in analytics and information visualization apps. The prepared data is then sent for feature engineering. Highlight designing is the extremely common approach of choosing, handling, and altering crude data into highlights that may be utilized in administered learning. To make AI function admirably on new assignments, it very well may be important to plan and prepare better elements. After the feature engineering, the data is sent for the modeling where machine learning algorithms may be applied for predicting the lung cancer in the patient based on the given set of parameters. After modeling the data, the data proceeded for containerization. A containerization is a typical part of software which bundles up code and all its needs, so the app runs firstly and consistently from one processing atmosphere to another. After containerization, it is checked whether the lung cancer disease is found in the patient. In the issue of yes, the letter will be displayed that lung cancer remains found. Whereas in the case of No, the process will be reoriented, as well as so on.

4 LIMITATIONS AND FUTURE DIRECTIONS

ML is a sub-field of AI as well as ML techniques are extensively utilized in the medical field, particularly in any type of disease prediction like Lung cancer, breast cancer, blood cancer, etc. Therefore, it is a need for a healthcare model that predicts lung cancer. Many previous traditional healthcare models used for this purpose, but it faces many limitations. For example, the deep learning-based model predicts lung cancer accuracy of 94.5%. In this research, the proposed model may be achieved better accuracy by employing ML techniques. In future, the proposed approach will be achieved better performance to forecast other types of cancer.

References:

1. Abbas, A., Abdelsamea, M.M., Gaber, M.M., 2020. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl. Intell.
2. Alam, J., Alam, S., Hossan, A., 2018. Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifie. Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018 1–4.
3. Ali, N., Ahmed, A., Anum, L., Ghazal, T.M., Abbas, S., Khan, M.A., Alzoubi, H.M., Ahmad, M., 2021. Modelling supply chain information collaboration empowered with machine learning technique. Intell. Autom. Soft Comput. 30, 243–257.
4. Ali, N., Ghazal, T.M., Ahmed, A., Abbas, S., Khan, M. A., Alzoubi, H.M., Farooq, U., Ahmad, M., Khan, Muhammad Adnan, 2022. Fusion-based supply chain collaboration using machine learning techniques. Intell. Autom. Soft Comput. 31, 1671–1687.
5. Ali Raza, S., Abbas, S., M. Ghazal, T., Adnan Khan, M., Ahmad, M., Al Hamadi, H., 2022. Content Based Automated File Organization Using Machine Learning pproaches. Comput. Mater. Contin. 73, 1927–1942.
6. Areej Fatima 1, M., Adnan Khan 1, Sagheer Abbas 1, M.W. 1, 2019. Evaluation of Planet Factors of Smart City through Multi-layer Fuzzy Logic (MFL) 11, 51–58.
7. Asadullah, M., Khan, M.A., Abbas, S., Alyas, T., Saleem, M.A., Fatima, A., 2020. Blind channel and data estimation using fuzzy logic empowered cognitive and social information-based particle swarm optimization (PSO). Int. J. Comput. Intell. Syst. 13, 400–408.
8. Asif, M., Abbas, S., Khan, M. A., Fatima, A., Khan, Muhammad Adnan, Lee, S.W., 2021. MapReduce based intelligent model for intrusion detection using machine learning technique. J. King Saud Univ. - Comput. Inf. Sci.
9. Aslam, M.S., Ghazal, T.M., Fatima, A., Said, R.A., Abbas, S., Khan, M.A., Siddiqui, S.Y., Ahmad, M., 2021. Energy-efficiency model for residential buildings using supervised machine learning algorithm. Intell. Autom. Soft Comput. 30, 881–888.
10. Ausawalaithong, W., Thirach, A., Marukatat, S., Wilaiprasitporn, T., 2019. Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach. BMEiCON 2018 - 11th Biomed. Eng. Int. Conf. 1–5.
11. Awai, K., Murao, K., Ozawa, A., Komi, M., Hayakawa, H., Hori, S., Nishimura, Y., 2004. Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on Radiologists' Detection Performance. Radiology 230, 347–352.
12. Bukhari, M.M., Ghazal, T.M., Abbas, S., Khan, M.A., Farooq, U., Wahbah, H., Ahmad, M., Adnan, K.M., 2022. An Intelligent Proposed Model for Task Offloading in Fog-Cloud Collaboration Using Logistics Regression. Comput. Intell. Neurosci. 2022.
13. Causey, J.L., Guan, Y., Dong, W., Walker, K., Qualls, J.A., Prior, F., Huang, X., 2019. Lung cancer screening with low-dose CT scans using a deep learning approach.
14. Chayal, N.M., Patel, N.P., 2021. Review of Machine Learning and Data Mining Methods to Predict Different Cyberattacks, Lecture Notes on Data Engineering and Communications Technologies.
15. Cheran, S.C., Gargano, G., 2005. Computer aided diagnosis for lung CT using artificial life models. Proc. - Seventh Int. Symp. Symb. Numer. Algorithms Sci. Comput. SYNASC 2005 2005, 329–332.
16. Dekhil, O., Naglah, A., Shaban, M., Ghazal, M., Taher, F., Elbaz, A., 2019. Deep Learning Based Method for Computer Aided Diagnosis of Diabetic Retinopathy, in: IST 2019 - IEEE International Conference on Imaging Systems and Techniques, Proceedings. Institute of Electrical and Electronics Engineers Inc.
17. Fatima, S.A., Hussain, N., Balouch, A., Rustam, I., Saleem, M., Asif, M., 2020. IoT enabled Smart Monitoring of Coronavirus empowered with Fuzzy Inference System. Int. J. Adv. Res. Ideas Innov. Technol. 6, 188–194.
18. Gai, K., Guo, J., Zhu, L., Yu, S., 2020. Blockchain Meets Cloud Computing: A Survey. IEEE Commun. Surv. Tutorials 22, 2009–2030.
19. Ghazal, T.M., Noreen, S., Said, R.A., Khan, M.A., Siddiqui, S.Y., Abbas, S., Aftab, S., Ahmad, M., 2022. Energy demand forecasting using fused machine learning approaches. Intell. Autom. Soft Comput. 31, 539–553.
20. Ghoshal, B., Tucker, A., 2020. Estimating Uncertainty and Interpretability in Deep Learning for Coronavirus (COVID-19) Detection 1–14.
21. Ihnaini, B., Khan, M. A., Khan, T.A., Abbas, S., Daoud, M.S., Ahmad, M., Khan, Muhammad Adnan, 2021. A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data

Fusion Based on Deep Ensemble Learning. Comput. Intell. Neurosci. 2021.
22. Iqbal, K., Khan, M.A., Abbas, S., Hasan, Z., 2019. Time complexity analysis of GA-based variants uplink MC-CDMA system. SN Appl. Sci. 1, 1–8.
23. Kadir, T., Gleeson, F., 2018. Lung cancer prediction using machine learning and advanced imaging techniques. Transl. Lung Cancer Res. 7, 304–312.
24. Khan, M.F., Ghazal, T.M., Said, R.A., Fatima, A., Abbas, S., Khan, M. A., Issa, G.F., Ahmad, M., Khan, Muhammad Adnan, 2021. An iomt-enabled smart healthcare model to monitor elderly people using machine learning technique. Comput. Intell. Neurosci. 2021.
25. Khan, Z., 2022. Used Car Price Evaluation using three Different Variants of Linear Regression 1, 40–49.
26. Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R., Rouf, N., Mohi Ud Din, M., 2020. Machine learning based approaches for detecting COVID-19 using clinical text data. Int. J. Inf. Technol. 12, 731–739.
27. Kyzalas, S., Nygård, L., Fischer, B.M., Edmund, J.M., Vogelius, I.R., 2019. EP-1939 Repeatability of FDG PET/CT based radiomic features using wavelet and Laplacian of Gaussian filters. Radiother. Oncol. 133, S1056–S1057.
28. Ma, F., Sun, T., Liu, L., Jing, H., 2020. Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. Futur. Gener. Comput. Syst. 111, 17–26.
29. Majeed, R.F., AB. M. Saeed, S., Abdulmajeed Abdilkarim, D., Mohammed Sidqi, H., 2020. Skin Tumors Diagnosis Utilizing Case Based Reasoning and The Expert System. Kurdistan J. Appl. Res. 5, 96–114.
30. Muneer, S., Rasool, M.A., 2022. A systematic review : Explainable Artificial Intelligence (XAI) based disease prediction 1, 1–6.
31. Muneer, S., Raza, H., 2022. An IoMT enabled smart healthcare model to monitor elderly people using Explainable Artificial Intelligence (EAI) 1, 16–22.
32. Naseer, I., 2022. Removal of the Noise And Blurriness using Global & Local Image Enhancement Equalization Techniques 1.
33. Saleem, M., Abbas, S., Ghazal, T.M., Adnan Khan, M., Sahawneh, N., Ahmad, M., 2022. Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. Egypt. Informatics J.
34. Saleem, M., Khan, M.A., Abbas, S., Asif, M., Hassan, M., Malik, J.A., 2019. Intelligent FSO Link for Communication in Natural Disasters empowered with Fuzzy Inference System. 1st Int. Conf. Electr. Commun. Comput. Eng. ICECCE 2019 1–6.
35. Sharma, P., Jindal, R., Borah, M.D., 2021. Blockchain Technology for Cloud Storage. ACM Comput. Surv. 53, 1–32.
36. Siddiqui, S.Y., Haider, A., Ghazal, T.M., Khan, M.A., Naseer, I., Abbas, S., Rahman, M., Khan, J.A., Ahmad, M., Hasan, M.K., Mohammed, A., Ateeq, K., 2021. IoMT Cloud-Based Intelligent Prediction of Breast Cancer Stages Empowered with Deep Learning. IEEE Access 9, 146478–146491.
37. Tabassum, N., Ditta, A., Alyas, T., Abbas, S., Alquhayz, H., Mian, N.A., Khan, M.A., 2021. Prediction of Cloud Ranking in a Hyperconverged Cloud Ecosystem Using Machine Learning. Comput. Mater. Contin. 67, 3129–3141.
38. Taher M. Ghazal, et al., n.d. A review on security threats, vulnerabilities, and counter measures of 5G enabled Internet-of-Medical-Things. Inst. Eng. Technol. 16, 421–432.
39. Ubaid, M., Arfa, U., Muhammad, H., Muhammad, A., Farooq, S., Saleem, M., 2022. Intelligent Intrusion Detection System for Apache Web Server Empowered with Machine Learning Approaches 1.