A Novel Software Engineering Approach toward Using Machine Learning for Improving the Efficiency of Health Systems

Diksha[1], Rakesh Kumar[2]

[1]Research Scholar, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana
[2]Professor, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana

ABSTRACT

Nowadays, from medical care businesses, enormous volume of data is produced. It is important to gather, store, and process this data to find from it the useful information and use it to take decisions. Blood glucose levels in diabetes patients are abnormally high, this is a sign of disease. Diabetes mellitus is a condition that may be fatal since it raises a person's likelihood of developing cardiovascular disease, kidney disease, and nerve damage. In this paper, it will examine how to utilize machine learning to anticipate diabetes. The prediction will predict whether a patient has diabetes or not. Here an extreme machine learning algorithm is used for this work. It also measures the accuracy of extreme machine learning algorithms. These algorithms can compare with other algorithms like KNN, SVM, Naive Bayes algorithm, Radom Forest, Logistic regression, etc, and also works with python machine learning models like Pandas, Seaborn, Matplotlib, NumPy, etc.

Keywords- Diabetes Detection; Extreme Machine Learning; Health Machine Learning.

INTRODUCTION

Type 1 diabetes occurs when the pancreas, a gland, cannot make enough insulin, while Type 2 diabetes occurs when the body's cells cannot use insulin. During digestion, glucose enters our bloodstream. The circulation sends insulin to cells to take glucose and convert it to energy. Beta cells produce insulin. Glucose remains in circulation because the pancreas doesn't create enough insulin in type1 diabetes. Thus, blood glucose/sugar levels increase dangerously high, which is harmful. High blood sugar may produce acute hunger, frequent urination, and excessive thirst. So, in order to solve this problem, the researcher's instance to make the use of data analytic for the early prediction of its disease and it has been observed that in a number of applications the data analytic has been used successfully by making the use of technique like machine learning, predictive analysis etc. Tech devices generate vast volumes of data. Academic data has proliferated worldwide. To efficiently manage, analyse, and visualise data, new software tools and methods must be found, developed, and implemented due to their complexity, volume, and heterogeneity.These results show how this subject has evolved and how big data papers are increasing. Since 2012, big data has grown exponentially, attracting more academics. Humanity has advanced in materials science, computer science, biotechnology, proteomics, and genomics. Individualized diagnosis and treatment are becoming the norm because of AI and big data. Deep learning, a form of machine learning, aims to construct computer systems that can discover patterns in training data and apply them to prediction and classification tasks on fresh data. Machine learning uses statistics, data mining, and optimization to develop models. Representational learning, a key topic in machine learning, automatically represents data. Deep learning methods are used for multi-level representations. The diabetes dataset will include patient data including glucose and insulin levels. Tabular data in a spreadsheet. Support vector machine, logistic regression, and other approaches might be evaluated for improvement. The test's outcome depends on the illness's detection or prognosis approach. Extreme machine learning might greatly improve system accuracy.

.

LITERATURE SURVEY

Machine learning is a subfield of computer science that studies the ways in which machines and other kinds of software may exhibit behaviors that are similar to those of humans. The operation of artificial intelligence and the process behind it could have been examined. The use of an artificial intelligence model has been done in a range of fitness-related domains, including radiography, the development of drugs, dermatology, and other professions. Detection of DR (Diabetes Retinopathy) using an artificial intelligence system can be traced combined with a fundus picture obtained from a smartphone. A retrospective study using mydriatic smartphone-based retinal images and eye art Artificial intelligence software can be shown a high sensitivity of 95.8% for the detection of diabetic retinopathy at any level of severity, and a sensitivity of over 99% for the detection of RDR (Relative Dose Response Test) . Both of these tests are used to determine whether or not a patient has diabetes (Sight Threatening diabetic retinopathy). The benefits of artificial intelligence include the development of more effective medical treatment methods and the reduction in the number of errors that occur during diabetes diagnosis. The machine learning model for diabetes prediction uses an artificial neural network model that is helpful for medical treatment (P. Ongsulee, Mzwil,2021).

The dataset issue could have been solved by machine learned used a variety of different algorithms. All different kinds of algorithms were quick to emphasize the fact that there was no one method that was most widely accepted as the best way to solve a problem. These kinds of algorithms were heavily relied on the nature of the issue that one was attempting to solve, as well as the number of variables that a certain form of the model was able to handle most effectively. The following was a brief overview of some of the most frequently used algorithms in machine learning. The use of machine learning to the management of medical devices has helped several sectors of the healthcare industry achieve higher levels of patient safety and quality of diagnosis and treatment. The amount of processing time required for the production of various classifiers was measured in milliseconds. The maximum accuracy possible was achieved by the expert system that was based on the decision tree classifier. The yearly cost of medical equipment maintenance and administration at a healthcare institution was around one percent of the overall budget, according to static (Mienye(2011), Rani(2016)).

Patients were only included in the PID dataset if they were female and were at least 21 years old. The dataset consisted of information on 768 patients and the 9 distinctive characteristics that was associated with each of them. WEKA was an open-source machine learning and data mining software application, was used by Masses for the purpose of doing performance analysis on the diabetes dataset. The data preparation, regression, visualisation, and feature selection capabilities offered by WEKA went at an additional cost. Python, which was utilised for the coding, was the programming language that was used to create the Neural Network (Mukhina, Anna A(2020)).

Implementation

A technique of extreme machine learning that can be used to health informatics prediction is a potential approach for improving accuracy that can be implemented. This can be done via the use of health informatics. The excessive learning machine, often known as ELM, is an innovative feedforward neural network learning technique. It requires a large amount of input data and only has one hidden layer in its architecture. In contrast to other approaches, the empirical risk minimization hypothesis may be learned in only a single iteration using ELM. This makes the empirical risk minimization hypothesis particularly advantageous. By using this method, it is possible to prevent a certain number of rounds as well as a local decrease.

A flow diagram of the procedure that was discussed before may be seen in Fig. 1.

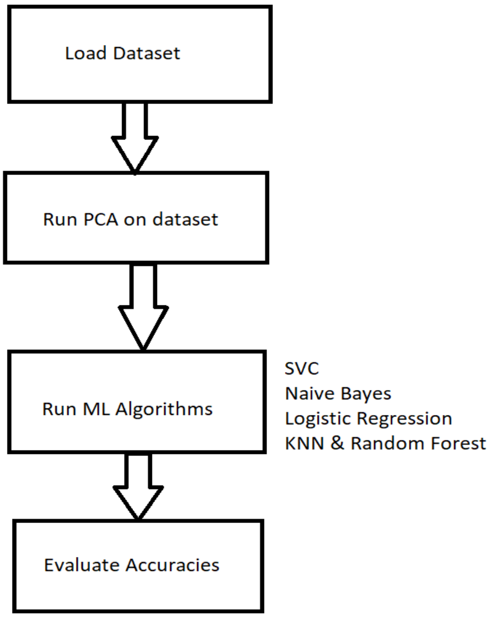

Fig. 1: Flow Chart

The author suggests combining Software Engineering with Machine Learning methods as a means of enhancing disease prediction in healthcare systems. In addition, the author suggests, in the reference base paper [1], shortening the amount of time required to anticipate disease as a consequence of worries such as a shortage of suitable hospitals or beds to accommodate an increasing number of patients. The author believes that by combining the capabilities of the software and machine learning, it will be possible to get better results and find a solution to the challenge of illness prediction in a more expedient manner. In his work, the author puts out the idea that may be abbreviated as SEMLHI (where SE stands for software, ML for machine learning, and HI for health information). According to the author, SEMLHI may be broken down into four distinct parts.

1) Data on Health Informatics: To make accurate predictions about diseases, Machine Learning models need to be constructed from datasets. However, datasets frequently contain missing data, null values, and non-numeric data, all of which have the potential to lower the accuracy of ML prediction. The author uses PREPROCESSING to get rid of all of the missing and null values in the healthcare data, and then uses Python to transform the non-numerical input into numerical data. This allows the author to successfully solve the problem. PREPRO SKLEARN PCA is a technique for reducing dimensionality that may be used to get rid of unneeded columns or characteristics in a dataset.

2. Run PCA on Dataset: The author of this tutorial makes use of many different machine learning approaches, including Multinomial Naive Bayes, Linear SVC, Logistic Regression, Random Forest, and Kernel Neural Networks, amongst others, throughout the presentation. This strategy will train itself using datasets that are accessible to the public before developing a training model. After that, the trained model will be used to new test data to make predictions. You can train a computer to learn and anticipate without the need for any aid from a human being if you use the methods that were just discussed.

3) Machine Algorithm Model: Once the author has developed the models outlined above, he or she may utilise the model to predict whether or not the patient lab results are positive or negative.

4) Software: Developers can validate the dependability of the aforementioned modules by doing software quality checks, software verification methods, and UNIT TESTING.

As a case study, the extreme machine approach was used for the data set pertaining to health. Extreme Learning Machine is the name given to a novel teaching method that was created in the 1990s for use with single hidden layer feedforward neural networks (ELM). When compared to the conventional neural network learning method, the standard neural network learning technique, this approach avoids the slow training speed and the over-fitting issues associated with the conventional neural network learning technique. Extreme learning machines consist of one or more layers of hidden nodes, each of which has its own set of parameters that are learned via a process of trial and error. They find use in many different applications, such as regression, classification, sparse approximation, grouping, feature learning, and compression, to name a few. Deep learning often makes use of a specific kind of neural network known as a multi-layered perceptron, or MLP for short. In certain communities, it is also often referred to as a multi-layer perceptron network (MLPN) (MLP). An MLP, which is also sometimes referred to as a "vanilla" neural network, is noticeably simpler than the sophisticated models that are used today. On the other hand, the techniques that it developed as a pioneer paved the way for the future construction of neural networks that are far more complex. The structure of a multi-layered perceptron is comparable to that of the human brain in that it consists of an interconnected network of neurons that can exchange information with one another. A specific number value has been assigned for each neuron. To make the network easier to understand, it is possible to divide it into three primary layers.

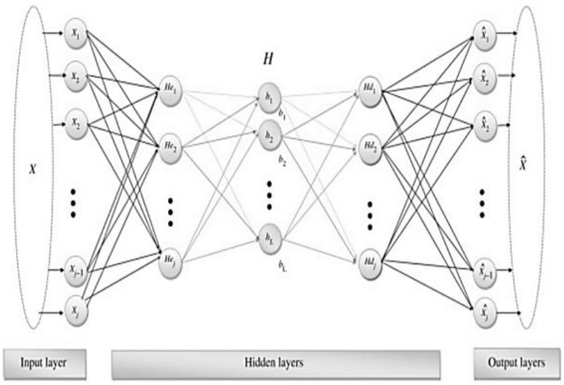

Fig. 2: ELM Layers

RESULTS

The next screen, Fig. 3, illustrates the user interface that will be used for the execution of the project. You may use the buttons to upload datasets, run algorithms, pre-process data, and do accuracy checks on your results.

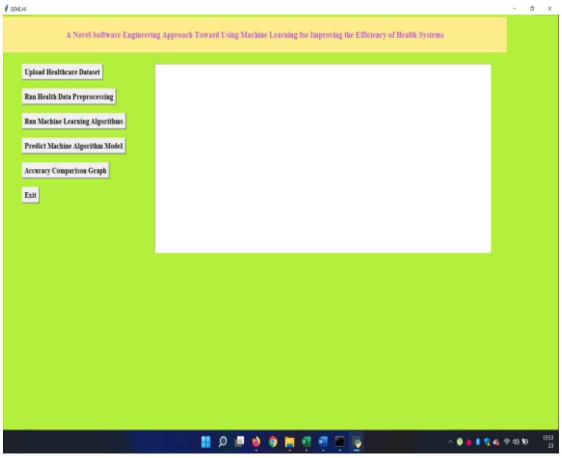

Fig. 3: Graphical User Interface

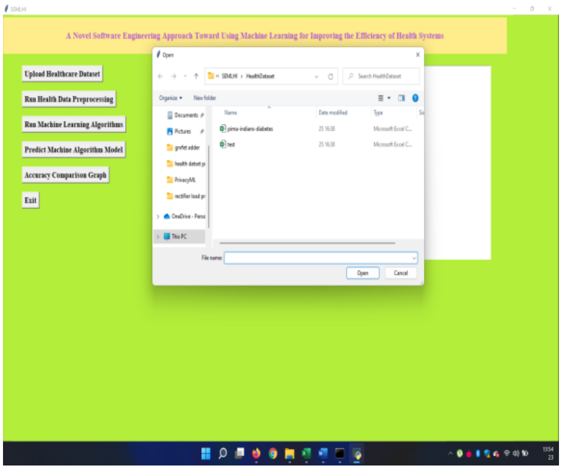

Fig. 4: Dataset Loading Screen

After it has been published, a dataset may be read in either the Excel or the CSV format, as seen in screen image 4. This choice is determined by the platform. The screen that can be seen below displays the message that can be read to show that the dataset has been loaded successfully. The act of putting information into a variable of the array type is referred to as loading.

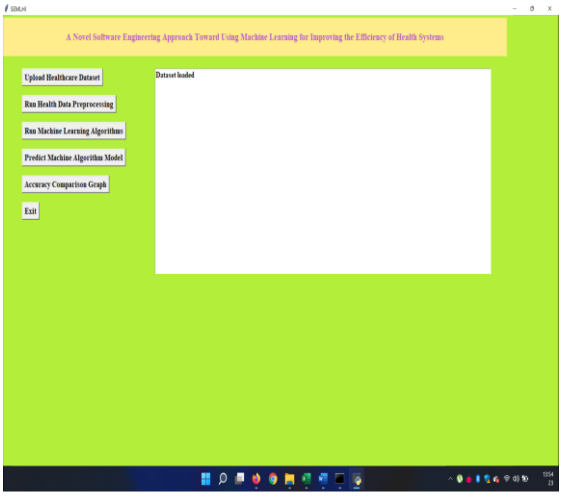

Fig. 5: Dataset is Loaded

The dataset is loaded into the website shown above, and the dataset is shown on the screen shown below, where the final label 'Class' contains the values 0 and 1, where 0 indicates the absence of sickness and 1 indicates the existence of illness in lab values.

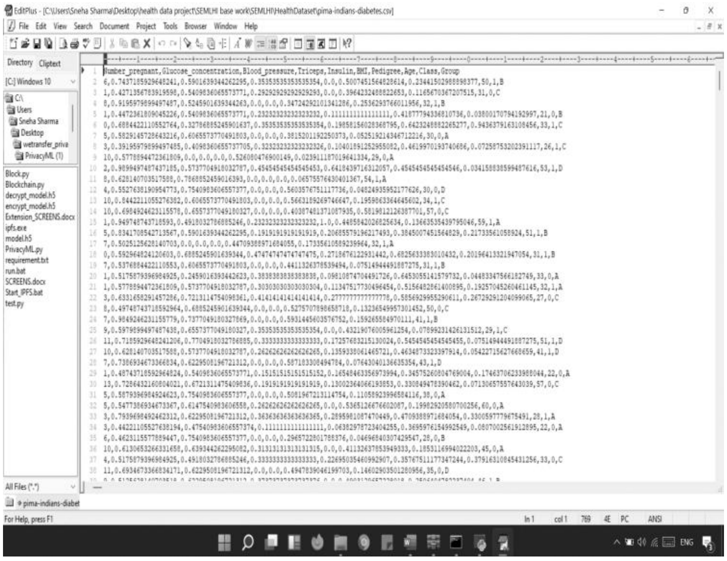

Fig. 6: View of Dataset

ML method will train using the above research proposal values and the Class Value (which includes 0 or 1), and then construct a model using the above lab report values and the Class Value (which includes 0 or 1), as shown in the above dataset screen. ML method will train using the above lab report values and the Class Value (which includes 0 or 1). In order to determine the class label, a train model was developed and then applied to the test data shown below. Because there is no Class label field in the following sample dataset, for instance, machine learning will make an educated assumption based simply on the laboratory findings. Take a look at the test results that are listed down below.

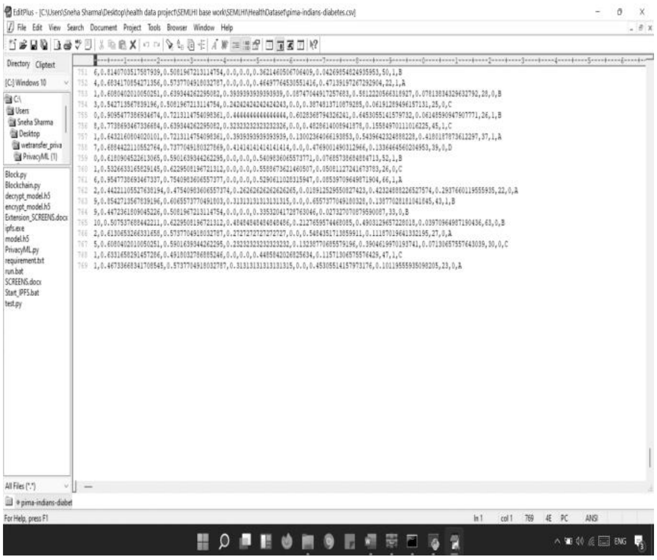

Fig. 7: Dataset test used

The data from the test lab report that was presented earlier do not include a Class classification. The next stage, which is seen on the screen that follows, is pre-processing. This phase includes applying principal component analysis (PCA) to the dataset. Principal component analysis (PCA) is a method of statistical analysis that employs an orthogonal transformation in order to change a set of correlated variables into a set of uncorrelated variables. This is accomplished by transforming the variables into a different set. Principal Component Analysis is what is meant to be abbreviated as "PCA." PCA is the technique that is used the most often for exploratory data analysis and visualisation in machine learning, and it is also the instrument that is used the most frequently in data science.

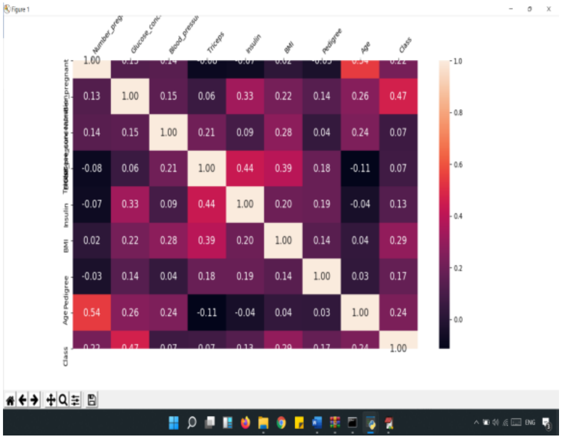

Fig. 8: Principal Component Analysis Output

The titles of the columns are shown at the top of the graph. The only column values that are significant, and the only ones that will be used for machine learning training, are those that are positive. The information obtained from the examinations and practises has been subjected to pre-processing, and the records have been tallied on the screen down below.

The screen that you see above will display whenever the test procedure is carried out. It is essential to enter the test results into the relevant processes for the tests.

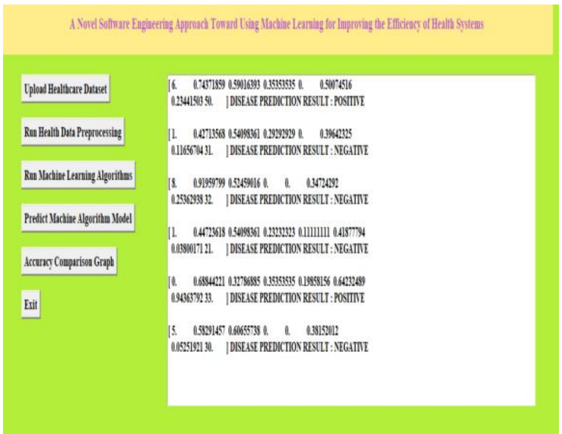

Fig. 9: Test Result Output

On the screen that can be seen above, the positive and negative diabetes prediction results for the data that was put into the test have been shown. And output graph shown below

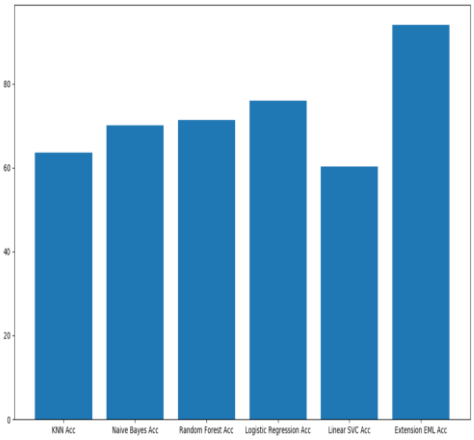

Fig. 10: Output Comparison for Accuracy

CONCLUSION

In this work, an extreme form of machine learning is used to a health dataset in an effort to increase the accuracy of diabetes prediction. Applications for batch learning, sequential learning, and incremental learning often make use of the extreme learning machine (ELM). This is due to the ELM's quick convergence, excellent generalisation capabilities, and ease of implementation. This is the case for a number of quite different reasons. Extreme Learning Machine (ELM) is a learning approach that is both productive and effective in terms of its efficacy and efficiency, and it was designed specifically for use with single hidden layer feed forward neural networks. ELM's primary use is in image recognition. When compared to traditional neural network methodologies, there is a decreased likelihood of overfitting difficulties occurring, and it takes less time to become skilled. In addition to this, fewer data are required for training purposes as compared to other systems.

According to the concept of empirical risk reduction, the empirical likelihood maximisation technique only needs one iteration of the learning process in order to be effective. Because of this, it is not necessary to do local minimization and go through a series of iterations in order to determine which value is the lowest. As a direct consequence of this, it has found use in a wide range of businesses and organisations, including the financial services industry, the healthcare industry, the manufacturing industry, and many others. As a result of using a variety of strategies all through the customary ELM development process, more effective ELM algorithms have been produced. In addition to the standard supervised learning applications of the conventional ELM, it now supports semi-supervised and unsupervised learning. This is in addition to the traditional supervised learning applications.

REFERENCES

[1]. M. Moreb, T. A. Mohammed and O. Bayat, "A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems," in IEEE Access, vol. 8, pp. 23169-23178, 2020, doi: 10.1109/ACCESS.2020.2970178.

[2]. Beam, Andrew &Kohane, Isaac. (2018). Big Data and Machine Learning in Health Care. JAMA. 319. 10.1001/jama.2017.18391.

[3]. Ferdous, Munira& Debnath, Jui& Chakraborty, Narayan. (2020). Machine Learning Algorithms in Healthcare: A Literature Survey. 1-6. 10.1109/ICCCNT49239.2020.9225642.

[4]. Rahman, Atta & Sultan, Kiran & Naseer, Iftikhar & Majeed, Rizwan & Musleh, Dhiaa & Gollapalli, Mohammed &Chaabani, Sghaier& Ibrahim, Nehad& Siddiqui, Shahan & Khan, Muhammad. (2021). Supervised Machine Learning-based Prediction of COVID-19. Computers, Materials and Continua. 69. 21-34. 10.32604/cmc.2021.013453.

[5]. Patel, Yash. (2021). Machine Learning in HealthCare.

[6]. Shams, Mahmoud &Elzeki, Omar &Abdelfatah, Mohamed & Abou El-Magd, Lobna& Darwish, Ashraf &Hassanien, Aboul Ella. (2021). Impact of COVID-19 Pandemic on Diet Prediction and Patient Health Based on Support Vector Machine. 10.1007/978-3-030-69717-4_7.

[7]. Harikumar& Mustafa, Malik & Sanchez, D. &Sajja, Guna &Gour, Sanjeev &Naved, Mohd&Jawarneh, Malik. (2021). IMPACT OF MACHINE learning ON Management, healthcare AND AGRICULTURE. Materials Today: Proceedings. 10.1016/j.matpr.2021.07.042.

[8]. Mienye, Domor& Sun, Yanxia& Wang, Zenghui. (2021). Improved Machine Learning Algorithms with Application to Medical Diagnosis.

[9]. Fetaji, Bekim&Fetaji, M. &Ebibi, Mirlinda& Ali, Maaruf. (2021). Predicting Diabetes Using Diabetes Datasets and Machine Learning Algorithms: Comparison and Analysis. 10.1007/978-3-030-90016-8_13.

[10]. Khan, Bilal & Naseem, Rashid & Shah, Muhammad Arif& Wakil, Karzan& Khan, Atif & Uddin, M. Irfan & Mahmoud, Marwan. (2021). Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques. Journal of Healthcare Engineering. 2021. 1-16. 10.1155/2021/8899263.

[11]. Lalmuanawma, Samuel & Hussain, Jamal. (2020). Applications of Machine Learning and Artificial Intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos Solitons & Fractals. 10.1016/j.chaos.2020.110059.

[12]. Khan, Bilal & Naseem, Rashid & Shah, Muhammad Arif& Wakil, Karzan& Khan, Atif & Uddin, M. Irfan & Mahmoud, Marwan. (2021). Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques. Journal of Healthcare Engineering. 2021. 1-16. 10.1155/2021/8899263.

[13]. Scott, Ian & Carter, Stacey &Coiera, Enrico. (2021). Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health & Care Informatics. 28. e100251. 10.1136/bmjhci-2020-100251.

[14]. Mijwil, Maad& Salem, Israa&Abttan, Rana. (2021). Utilisation of Machine Learning Techniques in Testing and Training of Different Medical Datasets. Asian Journal of Computer and Information Systems. 9. 29-34. 10.24203/ajcis.v9i4.6765.

[15]. [15] Zhao, Rui & Yan, Ruqiang& Chen, Zhenghua& Mao, Kezhi& Wang, Peng & Gao, Robert. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing. 115. 10.1016/j.ymssp.2018.05.050.