An Algorithm for Health Informatics Based on Extreme Machine Learning

SHARAD GARG[1], CHANDRA SHEKHAR YADAV[2], AREEBA KAZIM[3]
[1, 2, 3] *Computer Science and Engineering Department, Noida Institute of Engineering and Technology, Greater Noida, India*

Abstract— Health-care organisations may foresee patterns in a patient's medical condition and behaviour by using data mining, which entails examining several options and establishing connections between apparently unrelated bits of information. The volume and variety of raw data generated by healthcareinstitutions make it difficult to make sense of everything. Data must be collected and stored in an organised manner, as well as integrated, in order to develop a unified medical information system. In health, data mining permits the examination of a diverse set ofdata models that are unavailable or obscured by conventional analytical techniques. The objective of this research is to take a diabetic health dataset and analyse it using machine learning techniques to increase diabetesprediction accuracy.

Indexed Terms-- dataset, health, machine learning

I. INTRODUCTION

Using self-learning neural networks, deep learning programmes can find, recognise, and classify malignant tumours in photos. Deep learning, for example, is fast gaining favour in radiology and medical imaging as a subset of sophisticated machine learning that models thehuman brain's functioning.

Machine learning algorithms can now detect anomalies in images that aren't visible to the naked eye, assisting in disease diagnosis and treatment.

Healthcare will continue to be revolutionised by breakthroughs in machine learning. Predictive analytics for predicting breast cancer recurrence using medical data and photographs is currently being developed, as is a diagnostic tool for diabetic retinopathy. On a wide scale, The area of health informatics (HI) aspires to build a vast network of connections between seemingly unrelated concepts. Typically, a healthcare dataset is found to be insufficient & noisy; As a result, in the realm of software engineering, data reading from dataset linkage has historically failed. Machinelearning (ML) is a rapidly increasing field of computer science because of its capacity tostore enormous amounts of information. Several machine learning (ML) technologies could be utilized to examine information & generate insights that could messi assist employees and clinicians enhance the quality of their work; however, there is currently no developer-friendly solution available. Diabetes is one of the world's worst illnesses. Diabetes may be caused by a numberof factors, including obesity, hyperglycemia, and other disorders. It affects the insulin hormone, causing crabs' metabolisms to become irregular and blood sugar levels to fall. Diabetes is caused by an inadequate amount ofinsulin produced by the body. Diabetes affects around 422 million people globally, with the majority living in low- and middle-income nations, according to the World Health Organization (WHO). By 2030, this figure may have risen to 490 billion. Diabetes, on the other hand, is common in many nations, including Canada, China & India. India's genuine diabetes population is 40 million, owing to the country's population expansion tomore than 100 million persons. Diabetes is a leading cause of mortality globally. Early detection & treatment of disorders such as diabetes may help decrease their progression and extend a person's life. To do this, this research studies diabetes prediction using a number of risk factors for the condition. Thisis accomplished by using the Pima Indian Diabetes Dataset and anticipating diabetes onset using a variety of Machine Learning classification and ensemble methodologies. Machine Learning is a technique for training computers and other devices to do certain tasks.

Various Machine Learning Techniques deliver beneficial results in the accumulationof information by using obtained datasets to develop diverse classification and ensemble models.

II. IMPLEMENTATION

A possible method for improving accuracy isto use an extreme machine learning algorithm to informatics of health prediction. The extreme learning machine (ELM) is a cutting- edge learning method that employs feedforward neural networks with a single hidden layer. ELM is a risk-minimization approach that may be learned in a single iteration and is based on empirical risk minimization theory. Many rounds and local reductions are avoided with this method.

To improve illness prediction in health-caresystems and reduce the time it takes toanticipate disease, the author of the reference base work [1] employs software engineering and machine learning methodologies. As a result of the growing number of patients, there are not enough hospitals or beds to accommodate them. By utilising software and machine learning algorithms, this problem of predicting disease in less time can be solved. SEMLHI is the title of the author's intended paper (where SE refers to software & ML refers to machine learning & HI refers to health data). SEMLHI, according to the author, is made up of four components.

Health Informatics Data: Health Informatics Data: Machine Learning models must be built utilising datasets in order to predict anydisease. This sort of information can corrupt machine learning expectation precision. To resolve this issue, the author is applying PREPROCESSING to medical care information to eliminate all absent and invalid qualities and afterward changing non-numeric information over to numeric information utilizing Python SKLEARN PREPROCESSING classes. This dataset frequently has extra columns or attributes, which the author removes using a dimensionality reduction technique known as PCA. PCA (principal component analysis) eliminates superfluous features from a dataset & retains just those that are critical for accurate prediction.

Author employs a variety of machine learning methods in this lesson, including Linear SVC, Logistic Regression, Random Forest & KNN are all examples of multinomial Naive Bayes. Random Forest, Logistic Regression, and KNN are all examples of multinomial Naive Bayes & then apply the train model to fresh test data to do prediction. By using the aforementioned methods, you may teach a computer to learn and anticipate without the assistance of a person.

Model of a Machine Algorithm: After developing the models described above, Fresh test data could be used by the author to predict whether a patient's lab results will be positive or negative.

Software: Using software quality assurance, unit testing, and software verification, developers utilise this module to test the dependability of the above modules.

To execute this approach, the author recommends work that uses a variety of dataset sizes and classification, clustering, and regression. The author uses the Palestine Hospital dataset, which is not available on the internet and which the author does not publish,thus the author uses the INDIAN DIABETES dataset instead.

The extreme machine method is used to a health dataset in this paper. With only one hidden layer, the Extreme Learning Machine is a breakthrough feedforward neural network learning technique (ELM). When compared totraditional neural network learning methods,this methodology avoids the slow training speed and over-fitting concerns. Extreme learning machines have a single layer ofhidden nodes, or multiple layers of hiddennodes, from which the hidden nodes'parameters are learned & They can have asingle layer of concealed nodes or numerous levels. They are employed in procedures likeas regression, classification, grouping, compression, sparse approximatio & feature learning. The parameters of hidden nodes arelearned from a single layer of hidden nodes, ormultiple layers of hidden nodes, in extreme learning machines. A multi-layer perceptron isanother name for it (MLP). An MLP, oftenknown as a "vanilla" neural network, is a far more basic form of today's more complicatedmodels. However, The approaches it pioneered cleared the

door for the development of muchmore powerful neural networks. A multi-layered perceptron is composed of interconnected neurons that interact with oneanother, much as the human brain does. Eachneuron has a numerical value assigned to it. The network may be divided into three majorlayers. ELM is a single-hidden-layer feed-forward neural network (SLFN). The SLFN'sefficiency should be sufficient to allow foradditional learning when using data like threshold level, weight, and activation functionto construct a system.

The ELM (extreme learning machine) is a useful tool for swiftly training single-layer feed-forward neural networks (SLFNs). In recent years, Combining ELM with autoencoders to extract features fromunlabeled data has emerged as an intriguing new method, thanks to the progress of unsupervised learning methodologies.

In batch and sequential learning, the extreme learning machine (ELM) is frequently utilised, &because of its quick and efficient learning pace, rapid convergence, great generalisation capacity, and ease of implementation, incremental learning is popula. During the evolution of the standard ELM, numerous improved ELM algorithms have been developed, and the ELM's applicability has broadened from supervised to semi-supervised to unsupervised learning, among other things. The picture illustrates the concept of extreme learning by depicting inputs, outputs, and several hidden layers.

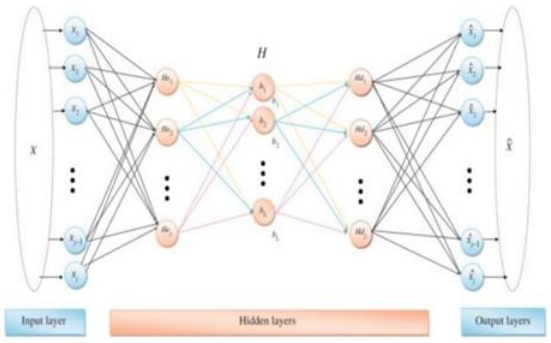

Figure 1: ELM Layers

III. RESULTS

Figure 2: Running all ML Algorithms

The extreme machine learning result is displayed on the screen above, as well as the machine learning method's maximumpercentage of efficiency.

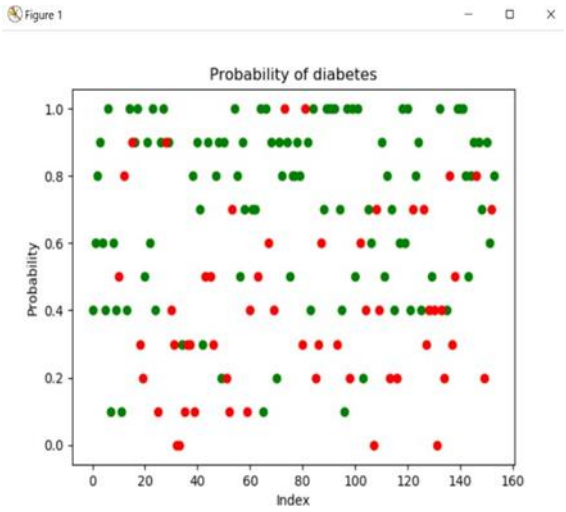

Figure 3: Probability Chart

The probability of diabetes is presented in the above screen, with red indicating yes and green indicating no.

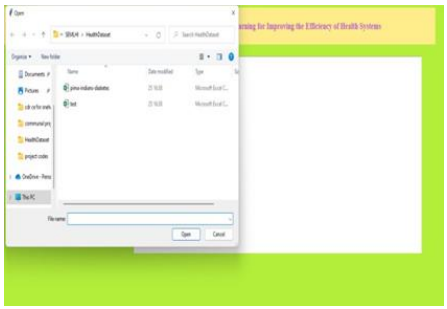

Figure 4: Loading Test Data

When the test method is run, the screen shown above appears. The test data must be entered for the test operations.

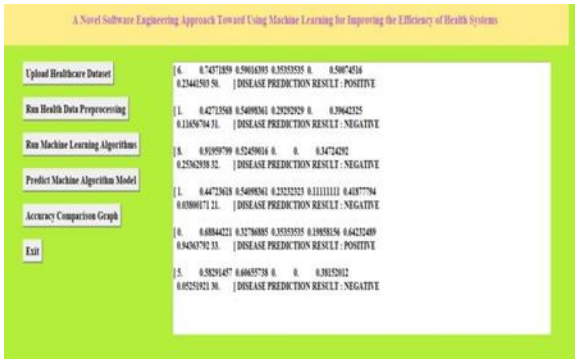

Figure 5: Running Test results

The diabetes positive and negative prediction result for the test loaded data is displayed on the screen above.

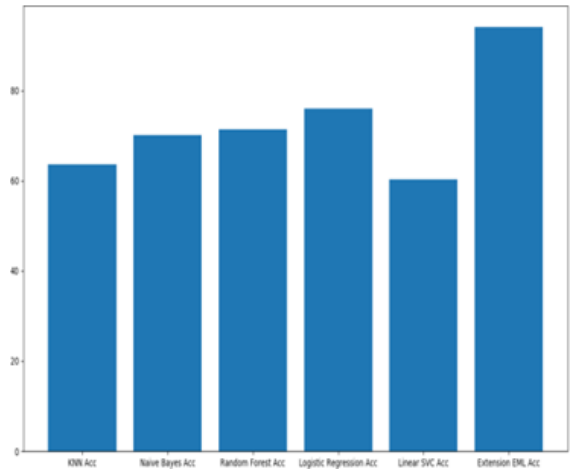

Figure 6: Accuracy results

The comparison of the extreme learning algorithm with other algorithms may be seen on the screen above. It demonstrates that the extreme learning algorithm performs best for diabetes prediction on the hospital dataset.

IV. CONCLUSION

The purpose of this work is to demonstrate how an extreme machine learning approachcan be used to significantly increase the accuracy of diabetes prediction utilising a health dataset. Due to the extreme learning machine's (ELM) quick convergence, high generalisation ability, and simplicity of building, Batch learning, sequential learning, and incremental learning are all examples of how it's used. ELM (extreme learningmachine) is a highly efficient and effective learning algorithm for feed forward neural networks with only one hidden layer. In compared to other conventional neural network approaches, it is less prone to overfitting and requires less training time. With only one iteration of the learning process,ELM is effective, according to empirical risk minimization theory.

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] 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.

[16] Kumar, Naresh & Das, Nripendra & Gupta, Deepali & Gupta, Kamali & Bindra, Jatin. (2021). Efficient Automated Disease Diagnosis Using Machine Learning Models. Journal of Healthcare Engineering. 2021. 1-13 10.1155/2021/9983652.

[17] Arvanitis, Theodoros & White, Sean & Harrison, Stuart & Chaplin, Rupert & Despotou, George. (2021). A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications. 10.1101/2021.02.11.21250741.

[18] Kaissis, Georgios & Ziller, Alexander & Passerat- Palmbach, Jonathan & Ryffel, Théo & Usynin, Dmitrii & Trask, Andrew & Lima, Ionésio & Mancuso, Jason & Jungmann, Friederike & Steinborn, Marc-Matthias & Saleh, Andreas & Makowski, Marcus & Rueckert, Daniel & Braren, Rickmer. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence. 3. 1- 12 10.1038/s42256-021-00337-8.

[19] Ojo, Adebola & Olanrewaju, Ahmed. (2019). Knowledge Discovery in Medical Database using Machine Learning Techniques. International Journal of Computer Applications. 178. 14-21.

[20] Rahman, Mohammed & Akhter, Nahida & Saleh, Sabbir. (2017). Health Data Mining: Machine Learning Approach. 3. 89-96.