



ENSEMBLE MACHINE LEARNING CLASSIFIERS: A PERFORMANCE REVIEW ON HEALTHCARE DATA

Li Zhenyu

Department of Computer Science, School of Information Technology, South China University of Technology,
Guangzhou, China

Abstract: The machine learning methodology consists of two stages: the training stage, during which an algorithmic classification program is taught to assign labels to data, and the testing stage, during which the algorithmic classification program is put through its paces. Data classification, also known as supervised learning, is a type of data processing in which data is divided into predetermined categories.

Two major ensemble machine learning classifiers, ReLU and Sigmoid, are the focus of our investigation. ELM is a novel approach to regression and classification issues that originates from single-hidden-layer feedforward neural networks. Classifiers such as ELM, SVM, KNN, Logistic regression, and others will be used in this work to analyze medical datasets with the help of activation functions. Diabetes, lung cancer, and brain tumor data sets are included. For some activation functions, including ReLU and sigmoid, ELM is optimized for classification performance. Using a publicly available, free online dataset of patients with mild to severe disease difficulties, we will employ ELM's Sigmoid, Relu, and other classifiers to analyze data on 900 to 1000 patients in the medical field.

Keywords: Extreme Learning Machine, Sigmoid, ReLU, SVM, KNN, J48, Logistic Regression, Artificial Intelligence, Machine Learning, and Deep Learning.

I. INTRODUCTION

The Extreme learning machine (ELM) from single hidden-layer feedforward neural networks is a new area for regression and classification problems. In this thesis, ELM will be applied to medical datasets using activation functions [1]. These medical datasets include Diabetes, Lung cancer, and brain cancer. The healthy and unhealthy features are given as inputs to the ELM classifier [2]. Furthermore, ELM is designed to achieve the best classification performance for certain activation functions such as logarithmic, sinusoid, and tangent sigmoid. We body is composed of trillions of cells. Throughout our life, we will grow and usually divide as needed. Cancer occurs if the whole transition goes catastrophically wrong, and our bodies continue to produce cells while old or malformed cells fail to die as they should. It makes our body's natural functioning challenging. According to research and studies, cancer is more than one killer disease affecting 1 in 3 patients worldwide. However, there are two categories of cancer worldwide: Hematologic (blood) cancers and Brain tumors [3]. These diseases are similar but may differ in how they expand, spread, and respond to medications. The reason for using machine learning methods in medicine is to diagnose diseases more quickly and automatically [2]. While achieving this

aim, the most critical problem that can be encountered is to find a machine learning method that gives the fastest and most accurate results and apply it to these medical data [3]. To solve this problem, the ELM classifier method, it has been discovered for fast and accurate results in the studies in the literature, has been proposed [1]. Amongst the most powerful data extraction issues is classification. The source is a database of records, each with various properties. Numeric features have numeric fields, whereas categorical features have non-numeric fields. There is also a third attribute called the class label. The goal of this classifier is to create a modeling approach that can be used to forecast future unlabeled records of the class label. Typically, classification accuracy is determined by calculating the proportion of rows classified. That fall within an appropriate class. That should probably also be determined. Performance analysis of different activation functions such as logarithmic Sigmoid, Sinusoid, and tangent Sigmoid of an ELM classifier will be performed to improve the fast recognition accuracy performance on several other medical data from the UCI database and Kaggle dataset. The best performing activation function will be selected. In the final phase.

1. Problem Statement

The medical line is diverse and analytical, and taking a specialization will narrow our project into specific diseases and patients. Our bodies are composed of billions of cells that divide and develop as appropriate during our lifetimes. When cells become aberrant or aged, they usually die—diseases like cancer start. Our bodies are composed of billions of cells that divide and develop as appropriate during our lifetimes.

When anything goes wrong during this process, our cells continue to make new cells while the old or abnormal ones do not die as they should. Cancer cells may push away normal cells as they proliferate. It creates our body's functioning challenging. Many people can effectively treat cancer. More people than ever before are living whole lives following cancer treatment.

2. Purpose

The goal is to develop time-efficient algorithms for diagnosing patients early in treatment, maintenance, checks, and emergencies. With the advancement in technology and processing power, fewer resources are needed for improved convergence [8]. Doctors are attentive to patient data and information, which has transformed the medical industry in terms of speedy disease identification, accurate results, and reliable diagnostics.

Machine Learning predicts data accurately, saving time, energy, resources, and lives. It takes data as input and outputs processed data. Many classifiers and algorithms must be compared to determine which has high quality and accuracy for decision-making. In this research, we assess ELM and another classifier to select the optimal diagnosis utilizing lung cancer, diabetes, and brain tumor datasets.

II. RELATED WORK

The reason for using machine learning methods in medicine is to diagnose diseases quickly and automatically. While achieving this aim, the most critical problem that can be encountered is to find a machine learning method for the fastest and most accurate results and apply it to these medical datasets. To solve this problem, the Extreme Learning Machine (ELM) classifier method, Sigmoid, ReLU, and others, which give fast and accurate results in the studies in literature, have been proposed. Many other researchers and publications have contributed their quota to the success of comparing ELM and different classifiers, and among the list are as follows:

Wenchao Yu, Fuzhen Zhuang, He Qing, and Zhongzhi Shi (2015). Learning deep representations via extreme learning machines. We compare our results to those of other algorithms on various real-world regression problems. First, we compare our results to AVG-ELM, an ensemble that averages all individual responses to produce the final prediction [9], [14].

A team of researchers created and enhanced the ELM algorithm. Many specialists are capable of creating various types of ELM algorithms. P Saratchandran, N Sundararajan, NY Liang, GB Huang, and N Sundararajan. An

online sequential ELM approach that may train data blocks of fixed or non-fixed sizes with an incremental value was created as a fast and accurate online sequential learning strategy for feedforward networks. As a result, the ELM method can train a huge amount of data instances [10], [11].

III. MATERIAL AND METHOD USED

Machine learning models, primarily supervised learning, are the applied technique. Deep learning is a form of machine learning that employs three or more neural network layers. These neural networks attempt, with varying degrees of success, to imitate human brain activity, allowing them to "learn" from huge volumes of data. While a single-layer neural network can still make approximations, several hidden layers will assist optimize and adapt improve accuracy. The dataset concerning this research, we considered a medical dataset a readily available and critical challenge that affects all over the world; these datasets are:

i. *Diabetes*

Each of these datasets is unique and has a different attribute. The diabetic dataset is obtained at the This dataset's original data came from the National Institute of Diabetes and Digestive and Kidney Diseases. The purpose is to predict if a patient has Diabetes based on diagnostic measures, which are easily accessible at kaggle website with the following properties. These examples were chosen from a larger database according to particular criteria. All of the patients at this clinic are women over the age of 21. [12].

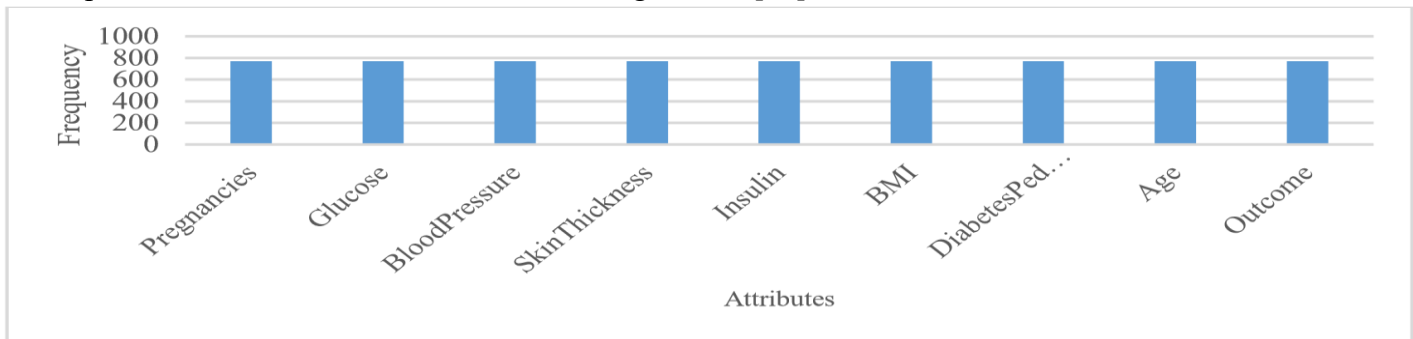


Fig. 1 Analysis of the Diabetics Dataset

If an individual develops Diabetes in line with World Health Organization criteria (that is, until the 2-hour i.e. pre glucose concentration is at least 200 mg/dl at each scan assessment or if discovered through routine care delivery), it diagnostics, binary-valued variable under research. ii. **Lung cancer:**

Cancer is a disease in which the cells of the body multiply uncontrolled. Lung cancer develops when cancer spreads to the lungs. This illness is the leading cause of cancer death [13]. Following skin cancer, lung cancer is the second most frequent cancer across both men and women in the United States. Lung cancer rates are down nationally after rising for decades as fewer people smoke and lung cancer therapies improve. Tobacco use is the most common lung cancer risk factor. Other kinds of cigarettes (like piping and cigarettes), second-hand tobacco, and exposure to contaminants such as asbestos or gases at work or home all can result in lung cancer, as can inherited familial causes. [15]. This dataset can be assessed on the open-source dataset repository at UCI site.

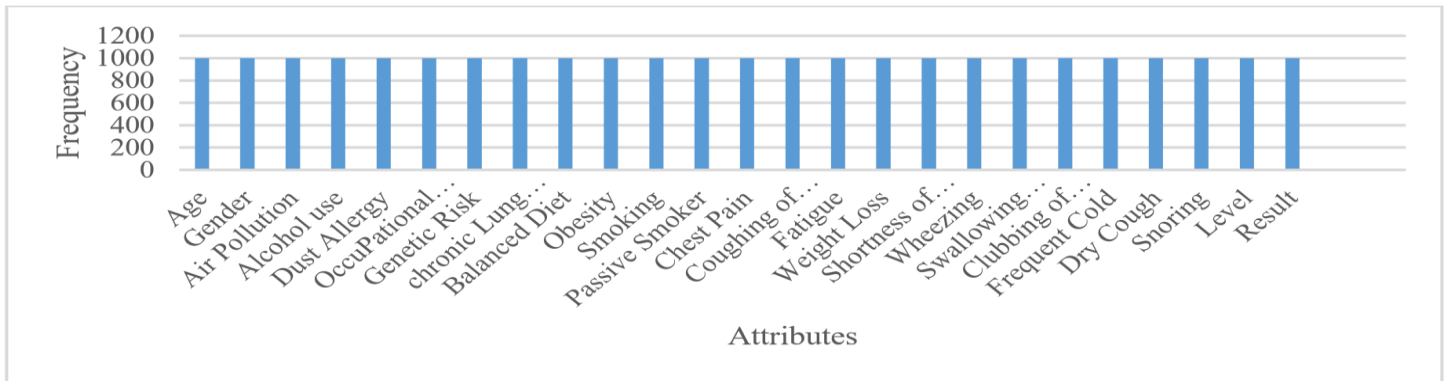


Fig. 2. Analysis of the Lung Cancer Dataset.

iii. *Brain Tumor:*

A central nervous system cancer is a malignant development of cells in the brain. Several brain tumors are malignant (cancerous), but others are not (non-malignant, non-cancerous, or benign). A tumor can form in the brain or in other neurons in the central nervous system (CNS), like the spinal or the nervous system. Most physical functions are controlled by the brain, including awareness, movement, feelings, ideas, speech, and memory. A cancer diagnosis can affect the brain's ability to function normally as well as effectively [16].

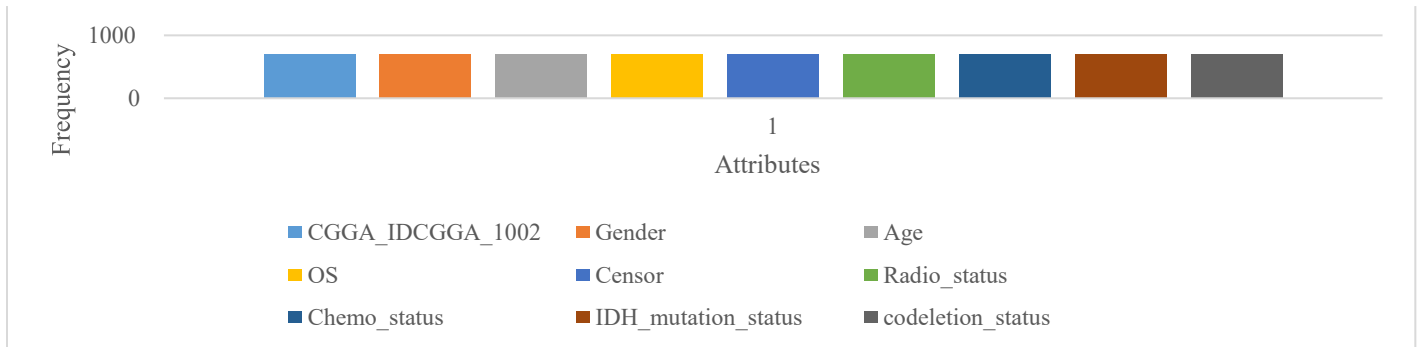


Fig. 3. Analysis of the Brain Tumor Dataset

i. *Feature Extraction and Selection*

Deep Convolutional Neural Networks Kernel filters apply to a dataset at all available offsets in the convolution layer. A filter comprises up of a layers with interconnecting weight, which takes input from the preceding layer's recessed portion, and a convolution layer, which functions as a learning algorithm and produces the output by convolving the last layer's local receptive fields with fixed-size kernel filters. So when the attributes were retrieved, their precise arrangement is less critical because their comparative relation towards other features has been generally maintained. The size of each output map has been defined. This layer's primary aim is to do extraction of features by searching across the whole input nodes, providing significant shift-invariance [17]. Furthermore, by combining using the same kernel across the input layer, it gives the benefit of weight sharing. Because it tends to offer outcomes towards the limits of its range, Sigmoid produces clear divisions. It is also worth emphasizing that its influence might be seen as a likelihood. That sigmoid function defines as follows:

$$\alpha(u) = 1/(1 + e^{(-u)}) \quad (1)$$

Because only little modifications are performed in each training update, a NN utilizing it can quickly become stranded in low gradient areas, causing the learning process to plateau. Rectified Linear Unit (ReLU) is commonly employed in NNs because it enables the efficient construction of deep neural networks [14]. If the state is negative, ReLU produces 0; else, the value of the condition is reverted. [18]:

$$\alpha(u) = \begin{cases} 0 & u < 0 \\ u & u \geq 0 \end{cases} \quad (2)$$

One shortcoming of ReLU is that any negative numbers soon yield 0, decreasing the model's ability to successfully develop out from input because all neurons with negative pre-processing step "turn down.", which outlines the NN training approach and the back-propagation algorithm. To get around the difficulty, many variants were created, including Leaky ReLU [18], Parametric ReLU [17], and Exponential Linear Unit [18].

TABLE I. Dataset used in this study

Dataset	Category	No. of cases
Diabetes	Medical	769
Lung cancer	Medical	1001
Brain Tumor	Medical	703
Total		2473

After extraction of features with DLN, data are transferred to the ELM algorithm. ELM uses only one layer of hidden neurons, with weights and biases determined at random. The generalized least squares approach is used to determine output weights, as shown below analytically.

$$\gamma = H^{-1}T \quad (3)$$

Here is the output layer's weight, H^{-1} is now the pseudo inverse of the output of the hidden layer, while T is a target matrix. There is no need to update the weights and biases in the suggested network on a regular basis; it automatically initialized randomly and offer higher quality performance. [19-20].

IV. EXPERIMENTAL OUTCOMES

The results of the experiments are presented in this section. The findings are a result of using the proposed model ELM. We offer detailed information about the findings from the experiment.

i. Implementation details

This section presents the setup of the presented model. This work uses Anaconda Navigator Application using Python programming environment to code the proposed ensemble deep transfer-learning model. Furthermore, using publically available datasets, ELM (ReLU and Sigmoid) and other classifiers were included in the used Python environment. With windows 10.1 32Gb RAM, 512 SSD, Core i7 9th gen, Anaconda Navigator Application using Python 3.1 and Nvidia graphics display.

ii. Accuracy Evaluation

The precision might be above average. The ELM algorithm offers a high level of statistical power. The accuracy can be affected by; the activation function used., whether or not the data is normalized, the nature of the characteristics; and the overall database distribution.

In the context of artificial neural networks, the ReLU activation function is defined as the positive component of its argument. This activation function first appeared in the perspective of graphical image enhancement in hierarchical neural network models towards the late 1960s. It was later revealed that it has considerable biological and mathematical motivations. In comparison to the commonly used activation functions prior to 2011, it was revealed in 2011 to facilitate greater development of deeper networks. [21].

Accuracy is the proportion of rectified cases to the total number of issues. It is expressed by Equation 3, with TP stands for True Positive, TN is for True Negative, FP stands for False Positive, and FN stands for False Negative. A true positive (TP) outcome is one in which the model accurately foretells a positive class. A TN is also a result in which the model correctly identifies any negative type. The FP outcome causes the method to forecast the positive class inaccurately. Furthermore, a FN causes the model to indicate the negative category inaccurately.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Table II. shows the Accuracy and Compilation time of the Diabetes dataset using the ELM (ReLU) function and other classifiers. That the ELM (ReLU) methodology was being verified using six different standard classifiers,

including SVM, K-neighbor Classifier, Logistic Regression, Bagging Classifier, Nave Bayes, and SGD Classifier. For the training of the dataset, 70% of the data is utilized for training, and the remaining 30% is used for testing. Performance indicators including accuracy and time have been studied and reported on., scientifically showing that ELM is 100% efficient and effective.

TABLE II. Accuracy and Compilation time of Diabetes dataset using ELM (ReLU) function and other classifiers

Method	Dataset	Accuracy %	Compile Time (Seconds)
ELM ReLU	Diabetes	1	0.02s
Support Vector Machine	Diabetes	0.74	0.033s
K-Neighbour Classifier	Diabetes	0.714	0.021s
Logistic Regression	Diabetes	0.746	0.22s
Bagging Classifier	Diabetes	0.753	0.28s
Naïve Bayes	Diabetes	0.77	0.27s
SGD Classifier	Diabetes	0.61	0.23s

Table II. shows the accuracy and compilation time of the Brain Tumor dataset using the ELM (ReLU) function and other classifiers; which ReLU function has almost 92% accuracy and a compilation time of 0.455 seconds which is still the best among its different standard classifiers.

TABLE III. Accuracy and Compilation time of Brain Tumor dataset using ELM (ReLU) function and other classifiers

Method	Dataset	Accuracy%	Compile Time
ELM Relu	Brain Tumor	0.915	0.455s
Support Vector Machine	Brain Tumor	0.755	1.01s
K-Neighbour Classifier	Brain Tumor	0.784	0.49s
Logistic Regression	Brain Tumor	0.782	0.494s
Bagging Classifier	Brain Tumor	0.9	0.486s
Naïve Bayes	Brain Tumor	0.773	0.499s
SGD Classifier	Brain Tumor	0.384	0.494s

Table III. shows the accuracy and compilation time of the Brain Tumor dataset using the ELM (ReLU) function and other classifiers; which Relu function has 91% accuracy and a compilation time of 0.464 seconds which is still the best among its different standard classifiers.

TABLE IV. Accuracy and Compilation time of Lung Cancer dataset using ELM (ReLU) function and other classifiers

Method	Dataset	Accuracy%	Compile Time
ELM Relu	Lung Cancer	0.91	0.464s
Support Vector Machine	Lung Cancer	0.634	0.623s
K-Neighbour Classifier	Lung Cancer	0.642	0.623s
Logistic Regression	Lung Cancer	0.641	0.664s
Bagging Classifier	Lung Cancer	0.677	0.462s
Naïve Bayes	Lung Cancer	0.574	0.435s
SGD Classifier	Lung Cancer	0.634	0.799s

iii. Comparative Analysis

We compare other ELM like sigmoid and observe the difference, if there is any, among the ELM functions with a comparative difference.

Table IV. shows the Accuracy and Compilation time of the Diabetes dataset using the ELM (Sigmoid) function and other classifiers. However, the Sigmoid function shows an excellent value for accuracy and time, to which there is no comparison.

Table V. Accuracy and Compilation time of Diabetes dataset using ELM (Sigmoid) function and other classifiers.

Method	Dataset	Accuracy	Compile Time
ELM Sigmoid	Diabetes	0.959	0.02s
Support Vector Machine	Diabetes	0.74	0.033s
K-Neighbour Classifier	Diabetes	0.714	0.021s
Logistic Regression	Diabetes	0.746	0.22s
Bagging Classifier	Diabetes	0.753	0.28s
Naïve Bayes	Diabetes	0.77	0.27s
SGD Classifier	Diabetes	0.61	0.23s

iv. Comparison based on time consumption

Time consumption analysis is displayed in table V. displays a time consumption comparison with the other common classifiers and the ELM Sigmoid function. Because of ELM's great performance, both ReLU and Sigmoid use extremely minimal time. In this study, the time required to process such a health illness disease is paramount to the diagnoses and lifesaving in the medical field while using ELM's weights and biases and hidden neurons at 5. Figure 4. shows a detailed display of the time constraint.

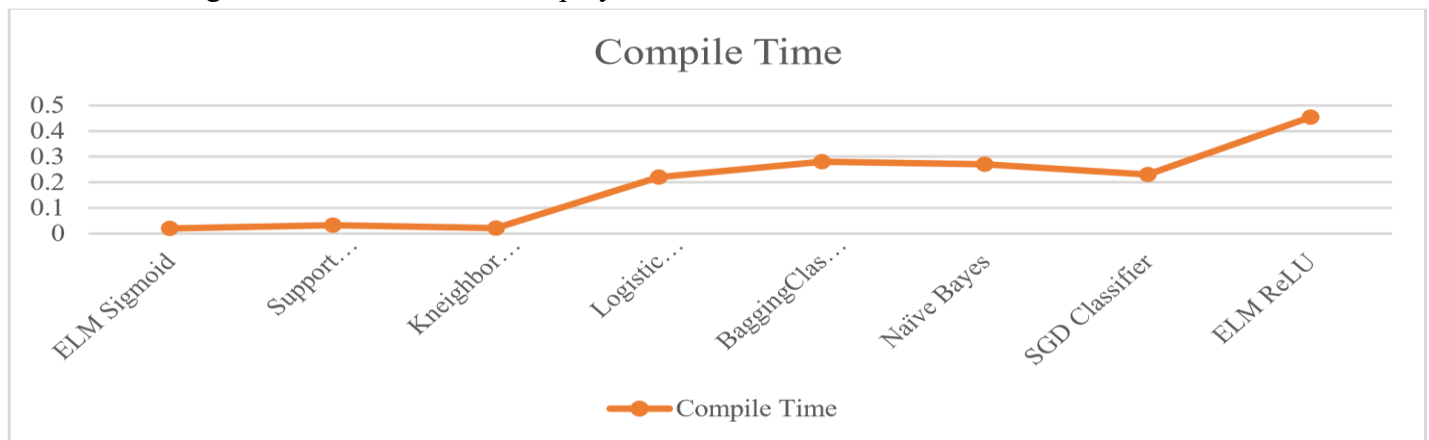


Fig. 4. Time Comparison of ELM and Other Classifiers

v. Comparison based on the accuracy

The accuracy of the dataset analysis is displayed in table V. Comparing accuracy with the other standard classifiers and the ELM Sigmoid function shows that ELM achieved 100% accuracy, which is excellent. The accuracy of both ReLU and Sigmoid is perfect because of the high performance of ELM. Furthermore, ELM allows fine-tuning the neurons to give exceptional results. Figure 5. gives display output of the result.

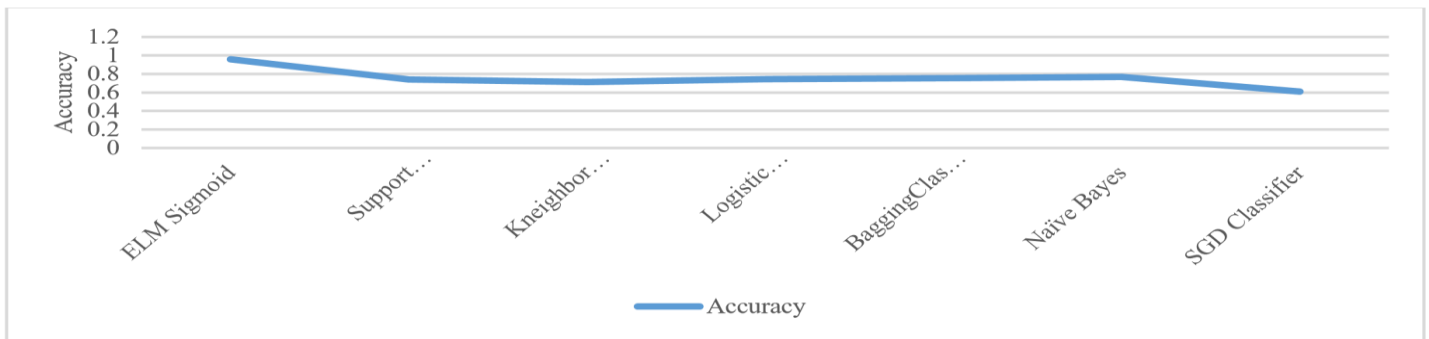


Fig. 5. Accuracy of ELM (Sigmoid) with other classifiers

vi. Comparison based on ELM functions

Among the two ELM functions we used, each one is time fast and achieves high accuracy; however, the ReLU function has better and higher accuracy than the Sigmoid function in Diabetes, Brain Tumor, and Lung cancer datasets. Furthermore, Sigmoid has a more significant advantage regarding timely completion because it is faster in time compilation. Anyhow compared to other classifiers, they are speedy and accurate. Figure 6. shows the accuracy of the ELM function and Table VI. displays the time difference between the two ELM functions.

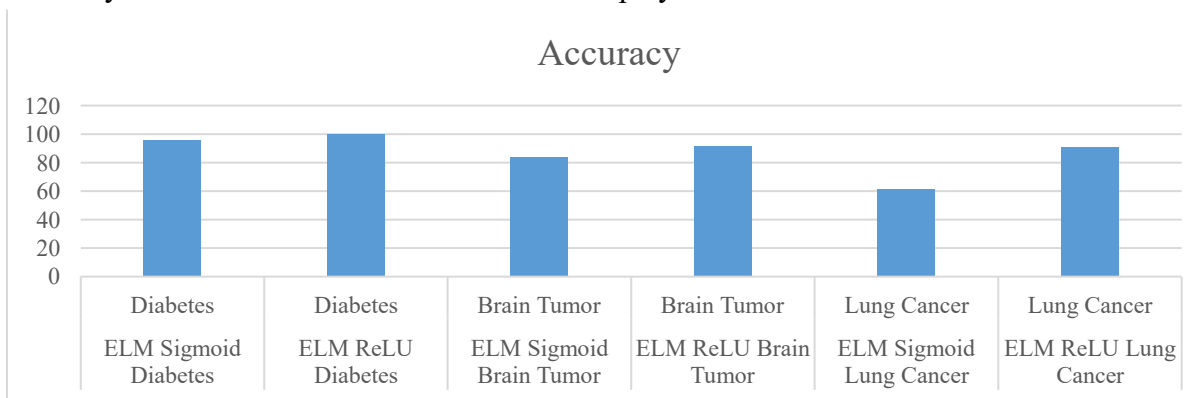


Fig. 6. Comparison of ELM Functions Using Medical dataset

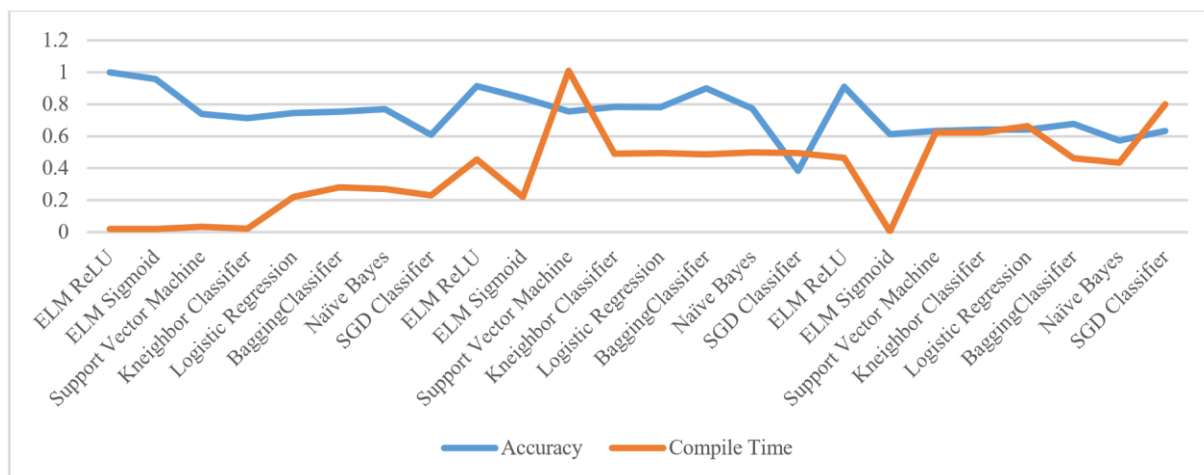


Fig. 7. Result of ELM functions and other classifiers using the medical dataset

vii. A comparison of time consumption and classifier accuracy

Finally, the study's overall objective is to bring an experimental result to ELM functions accuracy and time fastness comparable to other classifiers using medical datasets, and we have achieved our objective. Table VI.

shows the completion of the experiment conducted along with its respective values for each function. Fig 8. displayed the good characteristics feature of time and accuracy for the attributes, showing a good performance.

V. CONCLUSION

This research verified an extreme learning machine, a simple and efficient learning strategy for single-hidden layer feedforward neural networks. The proposed ELM differs from gradient-based learning for feedforward neural networks in three ways:

ELM learns quickly. Most simulations can complete ELM learning in milliseconds. Recently, most conventional machine learning couldn't pass a performance criterion. Even for basic applications, feedforward networks rarely require long training times using traditional learning techniques.

2. Conventional gradient-based supervised learning may experience a low learning rate, overfitting, and local minima. To prevent these concerns, weight decay and early stopping approaches may be necessary. Without such modest obstacles, ELM produces plausible answers. ELM learning seems simpler than other feedforward neural network algorithms.

SLFNs with different kernel functions are outside the initial scope and will be studied further. Future comparisons between ELM and backpropagation brain activation functions may be possible. In feedforward neural networks, multiple hidden layers can be used with backpropagation, whereas ELM is only used in SLFN. SLFNs can imitate any weighting and classification function. The ELM algorithm is highly adaptable.

REFERENCES

- Li, Bin & Li, Yibin & Rong, Xuwen. (2013). The extreme learning machine learning algorithm with tunable activation function. *Neural Computing and Applications*. 22. 10.1007/s00521-012-0858-9.
- Maciejewska, Monika & Szczurek, Andrzej & Ochromowicz, Ł. (2009). The characteristics of a “stop-flow” mode of sensor array operation using data with the best classification performance. *Sensors and Actuators B-chemical - SENSOR ACTUATOR B-CHEM*. 141. 417-423. 10.1016/j.snb.2009.07.018.
- Damasceno, Daniela & Teodosio, Cristina & van den Bossche, Wouter Jose Luis & van Dongen, J.J.M. & Orfao, Alberto. (2019). Distribution of subsets of blood monocytic cells throughout life. *Journal of Allergy and Clinical Immunology*. 144. 10.1016/j.jaci.2019.02.030.
- Yeh, wei-chih & Le, Trung-Hieu. (2021). Training neural network. 10.1016/B978-0-323-90198-7.00007-0.
- The C4.5 algorithm J48 is used in numerous domains for data classification
- Man, Zhihong & Lee, Kevin & Wang, Dianhui & Cao, Zhenwei & Miao, Chunyan. (2011). A modified ELM algorithm for single-hidden layer feedforward neural networks with linear nodes. *Proceedings of the 2011 6th IEEE Conference on Industrial Electronics and Applications, ICIEA 2011*. 2524-2529. 10.1109/ICIEA.2011.5976017.
- Yu, Wenchao & Zhuang, Fuzhen & He, Qing & Shi, Zhongzhi. (2015). Learning deep representations via extreme learning machines. *Neurocomputing*. 149. 308-315. 10.1016/j.neucom.2014.03.077.
- Yingu, Gao & Shengli, Xie. (2005). An improved form of the ELMS algorithm. *Systems Engineering and Electronics, Journal of*. 16. 433-436.

Liang NY, Huang GB, Saratchandran P, Sundararajan N. A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw. 2006 Nov;17(6):1411-23. doi: 10.1109/TNN.2006.880583. PMID: 17131657.

<https://www.kaggle.com/datasets/mathchi/diabetes-data-set>

Wada, A & Tateishi, R & Sugiyama, T & Matsuda, M & Hattori, M. (1965). [Lung cancer occurring in chronic pulmonary diseases]. Gan no rinsho. Japan journal of cancer clinics. 11. 637-42.

Khellal, H. Ma and Q. Fei, "Ensemble of Extreme Learning Machines for Regression," 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), 2018, pp. 1052-1057, doi: 10.1109/DDCLS.2018.8515915.

Márquez-Medina, Diego & Figueiredo, Ana & Ferreira, Catarina & Ulisses, Pedro. (2015). Lung Cancer Overview.

Rasmussen, Holger. (1987). Brain Tumors Can Be Subtle. The Western journal of medicine. 147. 473.

Xu, Yuesheng & Zhang, Haizhang. (2022). Convergence of deep convolutional neural networks. Neural networks: the official journal of the International Neural Network Society. 153. 553-563. 10.1016/j.neunet.2022.06.031.

Peycelon, R & Correard, R. (1970). [Surgical treatment of sigmoiditis. Results. Indications. Apropos of 45 cases]. Annales de chirurgie. 24. 1159-64.

Chen, Sitan & Klivans, Adam & Meka, Raghu. (2021). Efficiently Learning Any One Hidden Layer ReLU Network from Queries.

Aguirre, Diego & Fuentes, Olac. (2019). Improving Weight Initialization of ReLU and Output Layers. 10.1007/9783-030-30484-3_15.

Feng, Lang & Liu, Qianhui & Tang, Huajin & Ma, De & Pan, Gang. (2022). Multi-Level Firing with Spiking DS-ResNet: Enabling Better and Deeper Directly-Trained Spiking Neural Networks. 2446-2452. 10.24963/ijcai.2022/340.