Received: 15 January 2024 | Revised: 22 February 2024 | Accepted: 28 March 2024 | Published online: 3 April 2024

RESEARCH ARTICLE

Medinformatics
2024, Vol. 1(2) 73–78
DOI: 10.47852/bonviewMEDIN42022470

BON VIEW PUBLISHING

Elastic Net – MLP – SMOTE (EMS)-Based Model for Enhancing Stroke Prediction

Hussam Mezher Merdas[1,*]

[1]*Department of Computer Science, University of Kerbala, Iraq*

Abstract: A stroke is a sudden disruption in the blood supply to the brain, affecting one or more blood vessels that nourish the brain. This results in a disturbance or deficiency in the brain's oxygen supply, causing damage or impairment to brain cells. In some cases, determining the timing and severity of a stroke can be challenging. This study proposes an EMS (Elastic Net – MLP – SMOTE) model built on artificial intelligence, specifically utilizing two machine learning algorithms, Elastic Net and multilayer perceptron (MLP) by using Synthetic Minority Over-sampling Technique (SMOTE). The Elastic Net algorithm was employed for feature selection to identify crucial features, followed by prediction using the MLP algorithm. The Elastic Net algorithm was used due to its incorporation of both $L2$ and $L1$ regularization, providing good results in discerning influential features in model performance. The MLP algorithm was employed for its reliance on deep learning techniques, which yield promising results in such cases. This algorithm classified data from a comprehensive dataset containing essential features related to stroke. SMOTE is used to increase the performance of the model. Notably, no previous research study has integrated these three techniques together (Elastic Net – MLP – SMOTE). EMS achieved a prediction accuracy of 95% and MSE = 0.05. This model facilitates predicting the occurrence of stroke by relying on the patient's historical data, mitigating the sudden onset of this serious disease.

Keywords: stroke, machine learning, Elastic Net, multilayer perceptron, SMOTE

1. Introduction

A stroke is a serious life-threatening condition that occurs when blood flow to the brain stops or when there is a decrease in the amount of blood flowing to the brain [1]. It is crucial to address this condition immediately to minimize brain damage and prevent further complications. Early detection of stroke symptoms is crucial and may save the patient from the risk of developing this condition, which can lead to death or serious complications. The severity of cases varies from one patient to another, as does the recovery process. Some patients recover quickly after a short period of illness, while others may take longer. A patient having a stroke does not only mean that he is affected but also affects the psychological state of his relatives, friends, and co-workers. It can also affect any person, regardless of his age, physical condition, or health [2].

There are two types of stroke: ischemic and hemorrhagic, ischemic occurs when a blood clot or plaque buildup blocks or narrows an artery supplying blood to the brain while hemorrhagic occurs when a weakened blood vessel in the brain bursts and bleeds into the surrounding brain tissue [3]. A stroke can be mild or very severe. Ischemic strokes are the most common type and occur as a result of stopping blood flow to some parts of the brain. What increases the chance of stroke is if the patient has had a previous stroke. Heart disease also plays a role in stroke, and age also plays a role in this. Other important causes are high blood pressure, high cholesterol, and an anxious and unstable lifestyle, as well as smoking and drinking alcohol [4].

One of the most important symptoms of a stroke is paralysis on one side of the body and numbness on both sides or numbness of the face. It may also cause difficulty speaking and walking, and in some severe cases, a stroke can lead to coma. The first 24 h are crucial, so the injury and its symptoms can be treated more effectively. Some patients quickly recover from the symptoms after a period that may be longer or shorter, but others continue to suffer from problems including difficulty walking, lack of concentration, and others. Physical therapy and regular and scientific exercise help reduce the symptoms of this disease and speed recovery. The patient's environment also plays an essential role in improving his psychological state to restore his balance and improve his mood to reduce the effects of the injury and help speed up recovery [5].

To assist doctors and support patients, artificial intelligence was introduced into the field of healthcare. Machine learning (ML) algorithms were employed for this purpose. ML algorithms are used in predicting various fields, including medical and industrial applications. The results obtained can be excellent, average, or weak, depending on the algorithm type and data processing methods used. Classification and regression are crucial branches of ML. Classification is employed when data can be categorized into only two classes, indicating discrete data. On the other hand, regression models are used when dealing with continuous data with varying and unexpected values [6].

In this study, the classification branch was utilized to categorize data into two cases: individuals prone to stroke and those not prone to stroke. A dataset from a medical facility was preprocessed, involving tasks such as cleaning, handling human errors, and

*Corresponding author: Hussam Mezher Merdas, Department of Computer Science, University of Kerbala, Iraq. Email: hussam.m@s.uokerbala.edu.iq

addressing missing values. The Elastic Net algorithm was applied to identify features significantly influencing predictions. Subsequently, a multilayer perceptron (MLP) algorithm was employed to generate the necessary predictions. The accuracy of the obtained predictions was then measured. Then apply the SMOTE technique to improve the results. This study represents a step forward in predicting the likelihood of individuals suffering from the serious condition of stroke based on historical information. This study can form an important basis for doctors and patients to help detect the possibility of a patient having a sudden stroke.

The prediction of stroke occurrence has been explored by several studies listed below, yet there is no prior study that relied on performing feature selection for data using Elastic Net and utilized MLP for predictive purposes:

- The first research was proposed by Bathla and Kumar [7], as it employs several types of ML techniques to predict the occurrence of strokes by relying on various variables such as "high blood pressure, body mass index, heart disease, blood glucose levels, smoking status, previous strokes, and age." The proposed model gave positive results, according to the researchers in this study, as ten different classifiers were trained and the model then combined them using a mixed voting approach to achieve excellent accuracy. The model gave a prediction accuracy of up to 97%, and this indicates that the mixed voting classifier is better than the individual classifiers. Given the above, this model is considered an ideal model for predicting stroke, according to the researchers in this study. The proposed model helps doctors and patients by providing early detection and future prediction of the possibility of stroke occurring.
- This model was proposed by Ferdib-Al-Islam and Ghosh [8] and is based on two basic steps. First, a random forest regression algorithm was employed to find and replace missing values before starting the classification process. In the second step, the model was run using Automated Hyperparameter Optimization using a deep neural network to predict the possibility of stroke using a medical dataset. A dataset containing 43,400 records of patients whose data had been previously registered was used. The dataset contains 783 cases of patients suffering from stroke. The results using the proposed model showed that the model gave a false negative rate of only 19.1%, which is good when compared to other traditional methods, according to the researchers. The false positive rate, accuracy, and sensitivity given by this study are 33.1%, 71.6%, and 67.4%, respectively. According to the researchers themselves, the approach proposed in this study effectively reduced the false negative rate well, which means successfully reducing the misdiagnosis rate for predicting stroke.
- This study was proposed by Lin et al. [9], and the model was based on the use of the Taiwan Stroke Registry for stroke patients in 2006. The accuracy of three ML models (support vector machine, random forest, and artificial neural network) was employed and evaluated in this study. In addition, a hybrid artificial neural network with 10-fold cross-validation and 10-layer cross-validation evaluation was employed. ML techniques in the proposed model showed an area under the curve (AUC) exceeding 0.94 for both ischemic and hemorrhagic stroke types by using initially available data. After that, new data were added, which improved the predictive ability to 0.97 AUC. In this study, the researchers examined 206 clinical variables for many patients. The researchers selected 17 important features for ischemic strokes and 22 features for hemorrhagic strokes. The error analysis used in this study revealed that most prediction errors arose from the most severe stroke cases. This study concluded that ML algorithms trained using large and different datasets gave good prediction results. The results obtained confirm that the data added later are of great importance in obtaining good prediction results.
- This model was proposed by Tazin et al. [10] and is based on using four different ML models and employing them to predict the probability of a patient's stroke occurring in the future. The research relied on employing a set of physiological criteria and using several learning algorithms, including logistic regression (LR), decision tree classification, random forest classification, and voting classifier. These four algorithms were used separately, one by one, to train four different models for accurate stroke prediction. According to the researchers, Random Forest proved to be the most accurate algorithm among the proposed algorithms, as this algorithm gave an accuracy of approximately 96 percent. The study relied on using an available dataset to train and test the model. The researchers indicated that the accuracy rate of the models used in this study is higher than some previous studies, and they indicated that the reason for the good prediction was the result of the excellent comparison between the proposed models.
- In this study, Govindarajan et al. [11] proposed a model that classifies stroke in terms of its occurrence or non-occurrence. This model integrates text-mining tools and artificial intelligence techniques. ML algorithms are used as important and useful techniques in several fields of data management, medicine, etc., through good and appropriate training. The study extracts patients' symptoms from case sheets and then trains the system to future-proof the occurrence of stroke using the resulting data. The researchers collected data from 507 patients at Sogam Multispecialty Hospital in Kumbakonam, Tamil Nadu, India. Next, case papers were extracted using labeling and maximum entropy methodology. Researchers identified important, critical, and useful features from the data obtained and then used them to classify stroke. Several algorithms were employed, including "artificial neural networks, support vector machines, boosting, bagging, and random forests." The model gave good results and indicated the superiority of artificial neural networks trained with the stochastic gradient descent algorithm over other models, as the accuracy obtained was approximately 95% and a standard deviation of 14.69.
- In this study, Ahmed et al. [12] proposed a strategy based on comparing several distributed ML algorithms to predict the occurrence of a stroke for some people in the future. A dataset of several patients with stroke was used. The researchers pointed out that this proposed strategy was conducted based on a big data platform based on Apache Spark. The Apache Spark platform is one of the most widely used big data platforms, as the MLlib library was used to deal with a large dataset. MLlib is a Spark-compatible API for providing AI algorithms. In this study, four types of artificial intelligence algorithms were used. Decision trees, support vector machines, random forest classifiers, and LR algorithms were employed to design a model that predicts stroke. Finally, hyperparameter tuning and cross-validation were employed using ML algorithms. To get better results. To calculate the accuracy of the proposed model, several measures were used, including precision, recall, and $F1$. The results obtained from the proposed model showed that the random forest classifier achieved the highest accuracy of 90%.

2. Research Methodology

2.1. Preprocessing

In this study, an open-source dataset sourced from Kaggle specifically related to stroke was utilized (https://www.kaggle.com/code/docxian/stroke-prediction/input). This dataset consists of 5110

Table 1. Dataset used

ID	Gender	Age	Hypertension	Heart disease	Ever married	Work type	Residence type	Avg glucose level	BMI	Smoking status	Stroke
9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	Formerly smoked	1
51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	Never smoked	1
31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	Never smoked	1
60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	Smokes	1
1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	Never smoked	1

rows and 12 columns (ID, gender, age, hypertension, heart disease, ever married, work type, residence type, avg glucose level, BMI, smoking status, stroke) as shown in Table 1.

Preprocessing steps were applied to the dataset, including handling missing values by imputing them with the mean values. Additionally, categorical values were converted into numerical ones. Ultimately, a processed dataset suitable for classification was obtained. As illustrated in Figure 1, outlining the model's steps.

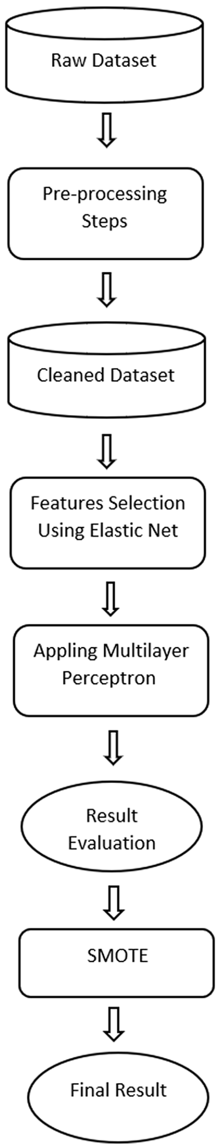

Figure 1. Steps of proposed model

It is noted that the model was fed a relatively good but unarranged dataset. Therefore, it was arranged and organized and the necessary steps were taken to make it suitable for use. Missing values were compensated for by using the average for the values in the same feature. Unnecessary fields such as (ID) have also been ignored. And other necessary treatments have been conducted. The new dataset was then entered into the Elastic Net algorithm to select the desired features, as will be explained in this study. Then, the MLP algorithm was applied correctly and appropriately to make accurate predictions. After that, the SMOTE technique was used to balance the results coming out of MLP and make them more realistic.

2.2. Features selection using Elastic Net

The Elastic Net is a type of regularized linear regression that combines two other techniques: Ridge Regression and Lasso Regression. It relies on two penalty functions, namely $L1$ penalty (lasso) and $L2$ penalty (ridge), to address constraints associated with individual regularization methods [13].

1) ***L*1 Penalty (Lasso):** This penalty encourages sparsity in the model by adding the absolute values of the coefficients to the loss function. It tends to shrink some coefficients to exactly zero, effectively selecting a subset of features and providing a form of feature selection.
2) ***L*2 Penalty (Ridge):** This penalty adds the squared values of the coefficients to the loss function. It helps prevent multicollinearity by suppressing large values of related coefficients.

The Elastic Net combines these penalties by adding both $L1$ and $L2$ penalties to the linear regression loss function. The algorithm includes two control parameters, alpha and lambda, which govern the strength of the penalties [14]:

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda P_\alpha(\beta)\right) \tag{1}$$

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha\left|\beta_j\right|\right) \tag{2}$$

Where:

- n is the observation number.
- y_i is the response at observation i.
- x_i is data, a vector of values at observation i.
- λ is a positive regularization parameter corresponding to one value of Lambda.
- $\beta 0$ and β are parameters.

When $\alpha = 1$ that means Elastic Net is the same as lasso. Elastic Net approaches ridge regression when α shrinks toward 0. $P\alpha(\beta)$ (the penalty term) interpolates between the $L1$ norm of β and the squared $L2$ norm of β for other values of α. This combination allows the Elastic Net to benefit from the feature selection property of Lasso while dealing with correlated variables and preventing their complete exclusion. The Elastic Net is particularly useful in feature selection when dealing with datasets containing a large number of features or when there is linear dependence among predictors. The $L1$ penalty in the Elastic Net tends to push some coefficients to zero, effectively achieving automatic feature selection. Finally, the Elastic Net strikes a balance between the ridge and lasso techniques, making it a versatile tool for linear regression tasks, especially when feature selection is crucial, and there is a need to effectively handle correlated predictors [15].

2.3. Applying MLP

The MLP is a type of artificial neural network that contains multiple layers, including an input layer that receives the features of the input data, one or more hidden layers (intermediate layers that transform the input through weighted connections and apply activation functions), and an output layer that produces the final output or prediction. It is widely used for classification in ML [16].

MLP makes the training process by several steps:

1) **Forward Propagation**: The input data are fed forward through the network. Each neuron in the hidden layers applies a weighted sum of inputs and passes the result through an activation function.
2) **Loss Calculation**: The output is compared to the actual target, and a loss function measures the prediction error.
3) **Backpropagation**: The error is propagated backward through the network. The weights are adjusted using optimization algorithms like gradient descent to minimize the loss.
4) **Iteration**: Steps 1–3 are repeated iteratively until the model converges to an optimal set of weights.

Common activation functions that are used with MLP include sigmoid, hyperbolic tangent, and rectified linear units. These functions introduce non-linearity, enabling the network to learn complex relationships in the data [17]. MLPs are powerful for classification tasks due to their ability to capture intricate patterns in data. They excel in scenarios with non-linear decision boundaries. The final layer often employs a SoftMax activation function for multi-class classification, producing probability distributions over classes [18].

MLP has more advantages as flexibility which means that MLPs can model complex relationships in data [19], and versatility which means MLP applies to various types of data and tasks, and feature learning (Automatically learning relevant features from the input). In conclusion, the MLP is a versatile neural network architecture commonly used for classification tasks, offering the flexibility to model complex relationships in data [20].

3. Results

EMS achieved a classification accuracy of up to 95%. Additionally, the mean squared error (MSE) metric was utilized to measure the model's error ratio using the equation below [21]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (Y_i - Y_i')^2 \tag{3}$$

where n is the observation number, Y_i is the observed value and Y'_i is the predicted value. The proposed model in this study yielded an error rate estimated at only 0.05.

The prediction accuracy was measured using a confusion matrix. This matrix contains four values, where the main diagonal values are the ones intended to be maximized. The upper-left value represents true negatives predicted correctly, while the lower-right value represents true positives predicted correctly. The upper-right value represents false positives predicted incorrectly, and the lower-left value represents false negatives predicted incorrectly. Accuracy can be derived by dividing the sum of the main diagonal values by the total matrix values [22]. The diagram below (Figure 2(A)) illustrates the confusion matrix. It reveals an issue where the model did not predict any true negative values, indicating excellent accuracy (95%) but a complete failure to predict a specific class related to stroke occurrence. This indicates that the proposed model suffers from a data imbalance issue. This means there may be a significant difference in the number of samples between different classes. Typically, this disparity is evident between the less common and the more common classes.

When there is data imbalance, the model may be prone to bias towards the more common class, tending to predict the more common class without adequately considering the less common classes. To address this issue, the **SMOTE** (Synthetic Minority Over-sampling Technique) technique has been employed. SMOTE is a technique used to address the class imbalance problem, especially in binary classification tasks [23]. The model becomes biased toward the majority class when one class significantly outnumbers the other, leading to suboptimal performance. This technique works by generating synthetic samples for the minority class [24]. It does this by creating synthetic instances along the line segments connecting existing minority class instances. This helps balance the class distribution and provides the model with more representative examples of the minority class. SMOTE aims to improve the model's ability to generalize and make better predictions for the

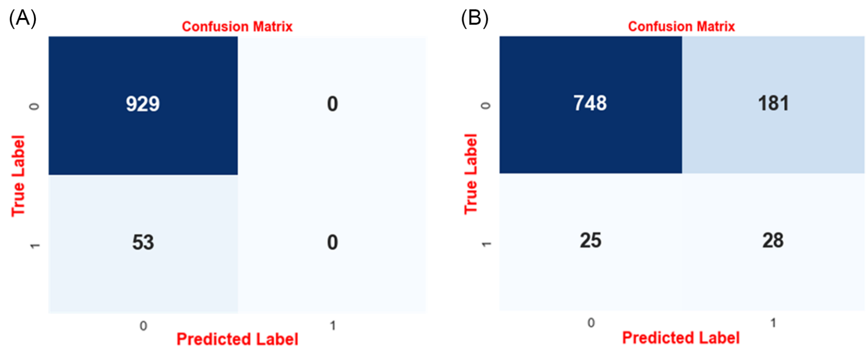

Figure 2. (A) Confusion matrix before using SMOTE. (B) Confusion matrix after using SMOTE

minority class by introducing synthetic samples. It is commonly applied before training a model to mitigate the impact of class imbalance on classification performance [25].

Applying this technique in the proposed model significantly improved the results. The proposed model was able to predict numerous classified values in each row individually, resulting in less sensitivity and bias. However, despite these improvements, the accuracy decreased to 79%. Future researchers may enhance the model for greater accuracy. Nevertheless, the SMOTE technique made the model more balanced and achieved better overall performance as shown in Figure 2(B).

This study can be further developed by other researchers in the future through the use of other types of SMOTE, such as BorderlineSMOTE, ADASYN, or MSMOTE.

4. Conclusion

Stroke is considered a serious medical condition that requires caution. This study proposed a model based on two important algorithms, Elastic Net and MLP. This study proposed a model called EMS that incorporates three crucial techniques: Elastic Net, MLP, and SMOTE. The use of SMOTE significantly enhanced the model's performance and made it more balanced. The proposed model yielded good results. This study serves as an initial step for subsequent research, allowing researchers to build upon it to enhance prediction accuracy. Additionally, utilizing other metrics for classification accuracy measurement could further refine the model, making it an ideal tool for warning patients about the risk of stroke.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by the author.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

References

[1] Almarzooq, A. H., Alshahrani, N. F., Al-Hariri, N. A., Alobaid, J. M., Alshurafa, Z. H., Abdulmajid, T. J., . . . , & Alotaibi, I. M. (2020). Evaluation of acute ischemic stroke management approach in the emergency department: Literature review. *Archives of Pharmacy Practice*, *11*(4), 26–31.

[2] Satterfield, B. A., Bhatt, D. L., & Gersh, B. J. (2022). Cardiac involvement in the long-term implications of COVID-19. *Nature Reviews Cardiology*, *19*(5), 332–341. https://doi.org/10.1038/s41569-021-00631-3

[3] Alhatemi, R. A. J., & Savaş, S. (2024). A weighted ensemble approach with multiple pre-trained deep learning models for classification of stroke. *Medinformatics*, *1*(1), 10–19. https://doi.org/10.47852/bonviewMEDIN32021963

[4] Rajora, M., Rathod, M., & Naik, N. S. (2021). Stroke prediction using machine learning in a distributed environment. In *Proceedings of the 17th International Conference on Distributed Computing and Internet Technology*, 238–252. https://doi.org/10.1007/978-3-030-65621-8_15

[5] He, S., Chen, Q., Jing, Z., Gu, L., & Luo, K. (2022). Avellis syndrome with ipsilateral prosopalgia, glossopharyngeal neuralgia, and central post-stroke pain: A case report and literature review. *Medicine*, *101*(39), e30669. https://doi.org/10.1097/MD.0000000000030669

[6] Han, S., Mannan, N., Stein, D. C., Pattipati, K. R., & Bollas, G. M. (2021). Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems. *Journal of Manufacturing Systems*, *61*, 45–53. https://doi.org/10.1016/j.jmsy.2021.08.004

[7] Bathla, P., & Kumar, R. (2023). A hybrid system to predict brain stroke using a combined feature selection and classifier. *Intelligent Medicine*. https://doi.org/10.1016/j.imed.2023.06.002

[8] Ferdib-Al-Islam, & Ghosh, M. (2021). An enhanced stroke prediction scheme using SMOTE and machine learning techniques. In *2021 12th International Conference on Computing Communication and Networking Technologies*, 1–6. https://doi.org/10.1109/ICCCNT51525.2021.9579648

[9] Lin, C. H., Hsu, K. C., Johnson, K. R., Fann, Y. C., Tsai, C. H., Sun, Y., . . . , & Hsu, C. Y. (2020). Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry. *Computer Methods and Programs in Biomedicine*, *190*, 105381. https://doi.org/10.1016/j.cmpb.2020.105381

[10] Tazin, T., Alam, M. N., Dola, N. N., Bari, M. S., Bourouis, S., & Khan, M. M. (2021). Stroke disease detection and prediction using robust learning approaches. *Journal of Healthcare Engineering*, *2021*, 7633381. https://doi.org/10.1155/2021/7633381

[11] Govindarajan, P., Soundarapandian, R. K., Gandomi, A. H., Patan, R., Jayaraman, P., & Manikandan, R. (2020). Classification of stroke disease using machine learning algorithms. *Neural Computing and Applications*, *32*(3), 817–828. https://doi.org/10.1007/s00521-019-04041-y

[12] Ahmed, H., Ghany, S. F. A., Youn, E. M. G., Omran, N. F., & Ali, A. A. (2019). Stroke prediction using distributed machine learning based on Apache Spark. *International Journal of Advanced Science and Technology*, *28*(15), 89–97. http://sersc.org/journals/index.php/IJAST/article/view/1553

[13] Altelbany, S. (2021). Evaluation of ridge, elastic net and Lasso Regression Methods in precedence of multicollinearity problem: A simulation study. *Journal of Applied Economics and Business Studies*, *5*(1), 131–142. https://doi.org/10.34260/jaebs.517

[14] Jenul, A., Schrunner, S., Liland, K. H., Indahl, U. G., Futsæther, C. M., & Tomic, O. (2021). RENT—Repeated elastic net technique for feature selection. *IEEE Access*, *9*, 152333–152346. https://doi.org/10.1109/ACCESS.2021.3126429

[15] Sloboda, B. W., Pearson, D., & Etherton, M. (2023). An application of the LASSO and elastic net regression to assess poverty and economic freedom on ECOWAS countries. *Mathematical Biosciences and Engineering*, *20*(7), 12154–12168. https://doi.org/10.3934/mbe.2023541

[16] Sahu, R. K., Müller, J., Park, J., Varadharajan, C., Arora, B., Faybishenko, B., & Agarwal, D. (2020). Impact of input feature selection on groundwater level prediction from a multi-layer perceptron neural network. *Frontiers in Water*, *2*, 573034. https://doi.org/10.3389/frwa.2020.573034

[17] Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive

survey and benchmark. *Neurocomputing*, *503*, 92–108. https://doi.org/10.1016/j.neucom.2022.06.111

[18] Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W., Fajardo-Flores, S. B., . . . , & Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Genomics*, *22*, 19. https://doi.org/10.1186/s12864-020-07319-x

[19] Fernández, J., Corbetta, M., Kulkarni, C. S., Chiachío, J., & Chiachío, M. (2024). Training of physics-informed Bayesian neural networks with ABC-SS for prognostic of Li-ion batteries. *Computers in Industry*, *155*, 104058. https://doi.org/10.1016/j.compind.2023.104058

[20] Fouladfar, M. H., Soppelsa, A., Nagpal, H., Fedrizzi, R., & Franchini, G. (2023). Adaptive thermal load prediction in residential buildings using artificial neural networks. *Journal of Building Engineering*, *77*, 107464. https://doi.org/10.1016/j.jobe.2023.107464

[21] Al-Amri, R. M., Hadi, A. A., Mousa, A. H., Hasan, H. F., & Kadhim, M. S. (2023). The development of a deep learning model for predicting stock prices. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *31*(3), 208–219. https://doi.org/10.37934/araset.31.3.208219

[22] Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label classifier performance evaluation with confusion matrix. *Computer Science & Information Technology*, *10*, 1–14. https://doi.org/10.5121/csit.2020.100801

[23] Alfrhan, A. A., Alhusain, R. H., & Khan, R. U. (2020). SMOTE: Class imbalance problem in intrusion detection system. In *2020 International Conference on Computing and Information Technology*, 1–5. https://doi.org/10.1109/ICCIT-144147971.2020.9213728

[24] Elreedy, D., Atiya, A. F., & Kamalov, F. (2023). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*. https://doi.org/10.1007/s10994-022-06296-4

[25] Adi Pratama, F. R., & Oktora, S. I. (2023). Synthetic minority over-sampling technique (SMOTE) for handling imbalanced data in poverty classification. *Statistical Journal of the IAOS*, *39*(1), 233–239. http://doi.org/10.3233/SJI-220080

How to Cite: Merdas, H. M. (2024). Elastic Net – MLP – SMOTE (EMS)-Based Model for Enhancing Stroke Prediction. *Medinformatics*, *1*(2), 73–78. https://doi.org/10.47852/bonviewMEDIN42022470