International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 3, May-June 2024
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Ensemble Bagged Trees for Epilepsy Classification

Generaldo D. Maylem, MDFPNA*,Isaac Angelo M. Dioses**,Genica Lynne C Maylem, MDFPNA ***, Jesusimo L. Dioses Jr.****, Danielle Bagaforo Meer*****
*MAPUA University, School of Information Technology, Makati Campus, Philippines
Email: iamdioses@mymail.mapua.edu.ph)
** Isabela State University, College of Medicine, San Fabian Echague Isabela.
Email:gdmaylem@yahoo.com)
*** Isabela State University, College of Medicine, San Fabian Echague Isabela.
Email:gcmaylem@yahoo.com)
**** Isabela State University, College of Computing Studies, Information and Communication Technologies, San Fabian Echague Isabela.
Email:jdiosesjr@gmail.com
*****MAPUA University, School of Information Technology, Makati Campus, Philippines
Email:debbiemeer30@gmail.com

--************************----------------------------------

Abstract:
Epilepsy is a persistent non-communicable brain condition. Epilepsy can be hereditary or result from an acquired brain injury, such as a stroke or trauma. In addition to strange actions, symptoms, and sensations, a person experiencing a seizure may possibly lose consciousness. According to estimates from the World Health Organization (WHO), 50 million people worldwide experience epileptic seizures; most of these cases are asymptomatic. The EEG signals are deconstructed using multiscale principal analysis, empirical mode decomposition, and feature extraction and selection methods that need statistical characteristics. Numerous industries use artificial intelligence, including aquaculture, agriculture, and medicine. Artificial intelligence approaches have enabled tackling the problem by introducing new technologies. Using Ensembled Bagged Trees, we can create a model that can classify epileptic seizures with 83.7% Accuracy, AUC of 89% indicates a surprising result using MATLAB's Ensembled Bagged Trees.

Keywords — **EEG, Machine Learning, Trees, Bagged Trees, Confusion Matrix, Epilepsy**

--************************----------------------------------

I. INTRODUCTION

A chronic, non-communicable brain disorder is epilepsy. Epilepsy can be brought on by an acquired brain problem, such as a stroke or trauma, or it can be inherited. When someone is having a seizure, they may even lose consciousness in addition to displaying odd behaviours, symptoms, and sensations. The World Health Organization (WHO) estimated that 50 million people worldwide suffer from epileptic seizures; most of them are asymptomatic. This is based on the organization's most recent evaluation.A "seizure" is a paroxysmal disruption of brain activity brought on by an excessive, hypersynchronous neuronal firing. "Epileptic seizure" is a term used to differentiate between a seizure brought on by aberrant firing of neurons and a non-epileptic event, like a psychogenic seizure. Recurrent, spontaneous seizures is called "epilepsy" [1].Recurrent, spontaneous seizures is called "epilepsy." There are several causes of epilepsy, and they are all related to underlying brain dysfunction [2].An epileptic seizure (ES) is caused by a transitory aberrant electrical discharge that occurs in the cerebral networks and lasts for less than a few minutes. ES attacks are not only hard to predict in terms of duration and strength, but they are also unpredictable [3].Not all cases of epilepsy are irreversible because many people with the condition improve to the point where they are no longer in need of medicines. Medication can manage seizures in more than 70% of instances [4].Later, in 1929, Hans Berger placed electrodes connected to a galvanometer on the skull to demonstrate EEG impulses' existence [18]. In addition to feature extraction and selection techniques that need statistical features, multiscale principal analysis and empirical mode decomposition are employed to break down the EEG signals [12].Artificial intelligence is applied in many industries, including medicine, agriculture, and aquaculture. With the advent of new technology, that problem can be solved using artificial intelligence techniques.

II. Related Literature

A survey of scalable technology for predictive analysis and early screening for diseases associated with a lifestyle is conducted. Investigates several types of AI using massive amounts of data in medical technology [5-6]. Recent research has examined how machine learning and deep learning approaches can be effectively applied to improve medical breakthroughs such as predicting Skin Cancer Diseases, breast cancer and malignant cells etc.The research by Usman et al. [11] suggested a method for extracting characteristics from EEG signals that uses empirical mode decomposition (EMD) and extracted time and frequency domain information. Here, the CHB-MIT dataset is employed. The average forecast time on the scalp for every individual in the sample is 23.60 minutes, and the model has yielded a higher true-positive rate of 92.23%.Potapov's research has shown that signal preprocessing can remove information that is essential for classification while still extracting useful features and suppressing noise. Original data must be considered for the signal classification approach since the classification accuracy influences the original data [13].To distinguish between epileptic signals and those that are not, the authors of this work [14] used fuzzy logic in conjunction with a genetic algorithm. To determine the precise risk for seizure detection, they devised several techniques and binary genetic algorithms to quantify the risk component.

A study conducted by Raghu Et. Al [15], the authors employed CNN and transfer learning to classify seven types of seizures with non-seizure EEG using a corpus of EEG data from Temple University Hospital. The purpose of this study is to perform a multiclass categorization. The signal was first converted into a spectrogram before being sent as input to CNN. Numerous DL pre-trained networks were used to select the optimal network for the investigation. The two best models for categorization accuracy employed in this study were 88.30% (Inceptionv3) and 82.85% (Googlenet).

EEG data were classified into ictal and interictal categories in a different investigation. When trying to comprehend brain output, the main issue here is the non-stationary and non-linear nature of the EEG signal. To develop a method for epileptic seizure identification on the CHB-MIT dataset that combines fuzzy-based and traditional machine learning approaches, the focus is on feature extraction from seizure EEG recordings [16].

Another study detects epilepsy by manually observing EEG data, which makes the transition to an automated diagnosis system both challenging and straightforward. Utilized is the Bonn EEG Dataset. With a 94.7% accuracy rate, the authors suggested using a least squares support vector machine (LSSVM) as a superior method for solving linear equations [17].

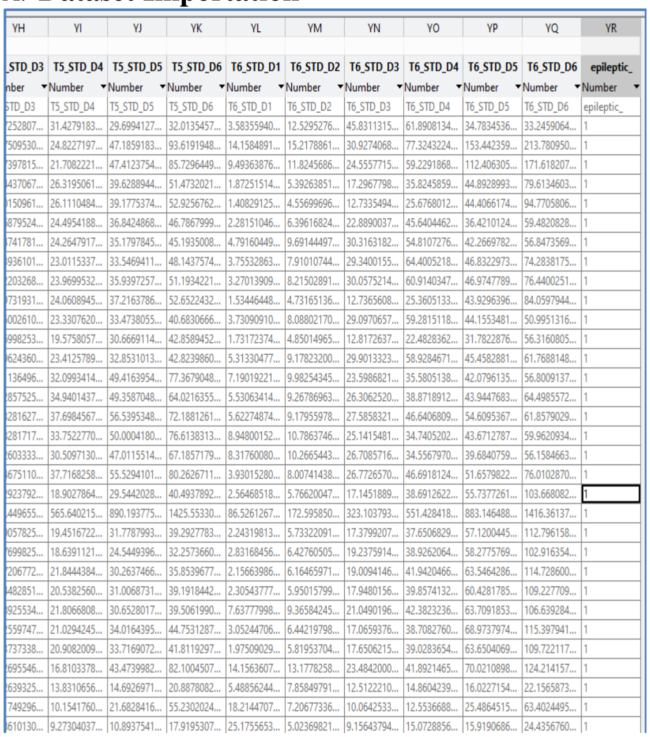
Ensemble Tree

A collection of classifiers built using a specific algorithm is called an ensemble. The predictions of each classifier in the ensemble are combined to classify each new example. According to Breiman, these predictions can be combined by adopting more complicated combinations or by taking the majority vote (for classification tasks) or the average (for regression tasks) [26,27,28]. Decision trees and related ensembles are a common set of techniques for classifying medical data in decision support systems. They are continuously being created by scholars [30]. Tree-based ensemble learning methods perform better than other ML techniques, according to Merghadi et al. [29].

III. Methodology

The datasets used in this study came from Kaggle; it will be cross validated for findings integrity by one of our Neurologist writers and comprises 2217 datasets and recordings of various EEG readings.The Image Processing Toolbox in MATLAB provides a wide range of industry-standard graphical tools and methods for processing, analyzing, creating algorithms, and visualizing medical images [31].In recent years, the MathWorks software MATLAB has become more and more well-liked as a quick development tool. Its many Toolboxes, sturdy design, and user-friendliness make it a popular tool in many domains, including medical image processing. MATLAB R2024a will be used in this work to facilitate machine learning.

A. Dataset Importation



YH	YI	YJ	YK	YL	YM	YN	YO	YP	YQ	YR
_STD_D3	T5_STD_D4	T5_STD_D5	T5_STD_D6	T6_STD_D1	T6_STD_D2	T6_STD_D3	T6_STD_D4	T6_STD_D5	T6_STD_D6	epileptic_
nber	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
STD_D3	T5_STD_D4	T5_STD_D5	T5_STD_D6	T6_STD_D1	T6_STD_D2	T6_STD_D3	T6_STD_D4	T6_STD_D5	T6_STD_D6	epileptic_
252807...	31.4279183...	29.6994127...	32.0135457...	3.58355940...	12.5295276...	45.8311315...	61.8908134...	34.7834536...	33.2459064...	1
509530...	24.8227197...	47.1859183...	93.6191948...	14.1584891...	15.2178861...	30.9274068...	77.3243224...	153.442359...	213.780950...	1
397815...	21.7082221...	47.4123754...	85.7296449...	9.49363876...	11.8245686...	24.5557715...	59.2291868...	112.406305...	171.618207...	1
437067...	26.3195061...	39.6288944...	51.4732021...	1.87251514...	5.39263851...	17.2967798...	35.8245859...	44.8928993...	79.6134603...	1
150961...	26.1110484...	39.1775374...	52.9256762...	1.40829125...	4.55699696...	12.7335494...	25.6768012...	44.4066174...	94.7705806...	1
879524...	24.4954188...	36.8424868...	46.7867999...	2.28151046...	6.39616824...	22.8890037...	45.6404462...	36.4210124...	59.4820828...	1
741781...	24.2647917...	35.1797845...	45.1935008...	4.79160449...	9.69144497...	30.3163182...	54.8107276...	42.2669782...	56.8473569...	1
936101...	23.0115337...	33.5469411...	48.1437574...	3.75532863...	7.91010744...	29.3400155...	64.4005218...	46.8322973...	74.2838175...	1
203268...	23.9699532...	35.9397257...	51.1934221...	3.27013909...	8.21502891...	30.0575214...	60.9140347...	46.9747789...	76.4400251...	1
731931...	24.0608945...	37.2163786...	52.6522432...	1.53446448...	4.73165136...	12.7365608...	25.3605133...	43.9296396...	84.0597944...	1
002610...	23.3307620...	33.4738055...	40.6830666...	3.73090910...	8.08802170...	29.0970657...	59.2815118...	44.1553481...	50.9951316...	1
998253...	19.5758057...	30.6669114...	42.8589452...	1.73172374...	4.85014965...	12.8172637...	22.4828362...	31.7822876...	56.3160805...	1
624360...	23.4125789...	32.8531013...	42.8239860...	5.31330477...	9.17823200...	29.9013323...	58.9284671...	45.4582881...	61.7688148...	1
136496...	32.0993414...	49.4163954...	77.3679048...	7.19019221...	9.98254345...	23.5986821...	35.5805138...	42.0796135...	56.8009137...	1
857525...	34.9401437...	49.3587048...	64.0216355...	5.53063414...	9.26786963...	26.3062520...	38.8718912...	43.9447683...	64.4985572...	1
281627...	37.6984567...	56.5395348...	72.1881261...	5.62274874...	9.17955978...	27.5858321...	46.6406809...	54.6095367...	61.8579029...	1
281717...	33.7522770...	50.0004180...	76.6138313...	8.94800152...	10.7863746...	25.1415481...	34.7405202...	43.6712787...	59.9620934...	1
603333...	30.5097130...	47.0115514...	67.1857179...	8.31760080...	10.2665443...	26.7085716...	34.5567970...	39.6840759...	56.1584663...	1
675110...	37.7168258...	55.5294101...	80.2626711...	3.93015280...	8.00741438...	26.7726570...	46.6918124...	51.6579822...	76.0102870...	1
923792...	18.9027864...	29.5442028...	40.4937892...	2.56468518...	5.76620047...	17.1451889...	38.6912622...	55.7377261...	103.668082...	1
449655...	565.640215...	890.193775...	1425.55330...	86.5261267...	172.595850...	323.103793...	551.428418...	883.146488...	1416.36137...	1
057825...	19.4516722...	31.7787993...	39.2927783...	2.24319813...	5.73322091...	17.3799207...	37.6506829...	57.1200445...	112.796158...	1
699825...	18.6391121...	24.5449396...	32.2573660...	2.83168456...	6.42760505...	19.2375914...	38.9262064...	58.2775769...	102.916354...	1
206772...	21.8444384...	30.2637466...	35.8539677...	2.15663986...	6.16465971...	19.0094146...	41.9420466...	63.5464286...	114.728600...	1
482851...	20.5382560...	31.0068731...	39.1918442...	2.30543777...	5.95015799...	17.9480156...	39.8574132...	60.4281785...	109.227709...	1
925534...	21.8066808...	30.6528017...	39.5061990...	7.63777998...	9.36584245...	21.0490196...	42.3823236...	63.7091853...	106.639284...	1
559747...	21.0294245...	34.0164395...	44.7531287...	3.05244706...	6.44219798...	17.0659376...	38.7082760...	68.9737974...	115.397941...	1
737338...	20.9082009...	33.7169072...	41.8119297...	1.97509029...	5.81953704...	17.6506215...	39.0283654...	63.6504069...	109.722117...	1
695546...	16.8103378...	43.4739982...	82.1004507...	14.1563607...	13.1778258...	23.4842000...	41.8921465...	70.0210898...	124.214157...	1
639325...	13.8310656...	14.6926971...	20.8878082...	5.48856244...	7.85849791...	12.5122210...	14.8604239...	16.0227154...	22.1565873...	1
749296...	10.1541760...	21.6828416...	55.2302024...	18.2144707...	7.20677336...	10.0642533...	12.5536688...	25.4864515...	63.4024495...	1
610130...	9.27304037...	10.8937541...	17.9195307...	25.1755653...	5.02369821...	9.15643794...	15.0728856...	15.9190686...	24.4356760...	1

Figure 1. EEG Dataset Importation using MATLAB.

Figure 1 shows that the EEG Dataset was imported, and it has 667 manual features extracted that help to classify whether that patient has epilepsy or not.

B. Classification Learner

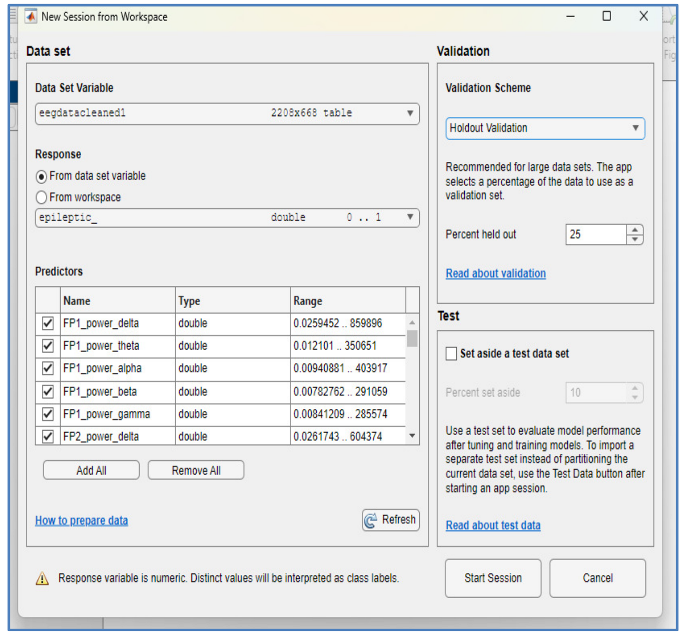


Figure 2. Classification Learner Session preprocessing.

Figure 2 shows that the holdout validation was 25% and 75% for training with 2217 samples, and it shows the first 6 predictor features.

IV. Results and Discussion

After importing and preprocessing the datasets, we trained them using the Ensemble Bagged Trees Algorithm. The model achieved an accuracy of 83.7%, as shown in Figure 3, processing 3300 observations per second during training, which took 35.7 seconds. The model utilized a maximum of 2207 splits and included 30 learners in the ensemble.As illustrated in Figure 5, the confusion matrix of the trained model is quite promising. It shows the following results: True Positives (TP): 252, False Positives (FP): 45, False Negatives (FN): 45, True Negatives (TN): 210. These values indicate that the model correctly identified 252 positive cases and 210 negative cases, while it incorrectly identified 45 positive cases as negative and 45 negative cases as positive

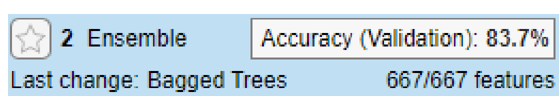


Figure 4. Accuracy Validation of 83.7%

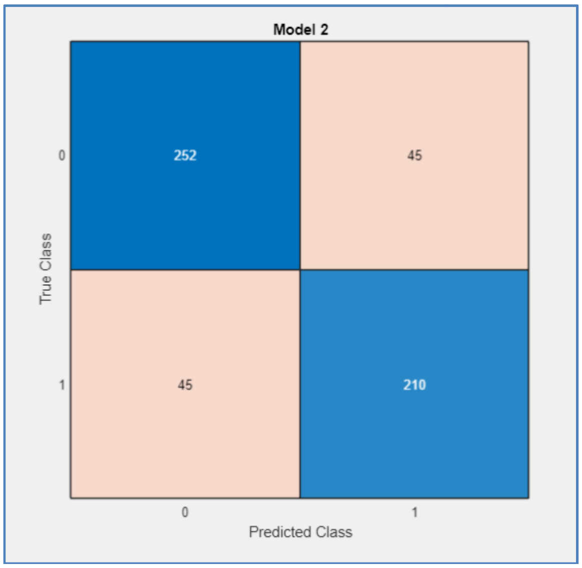


Figure 5 Confusion Matrix

TABLE I
ACCURACY FORMULA

ACCURACY	$\frac{\text{TRUE POSITIVE + TRUE NEGATIVE}}{\text{(TP + TN+ FP + FN)}}$

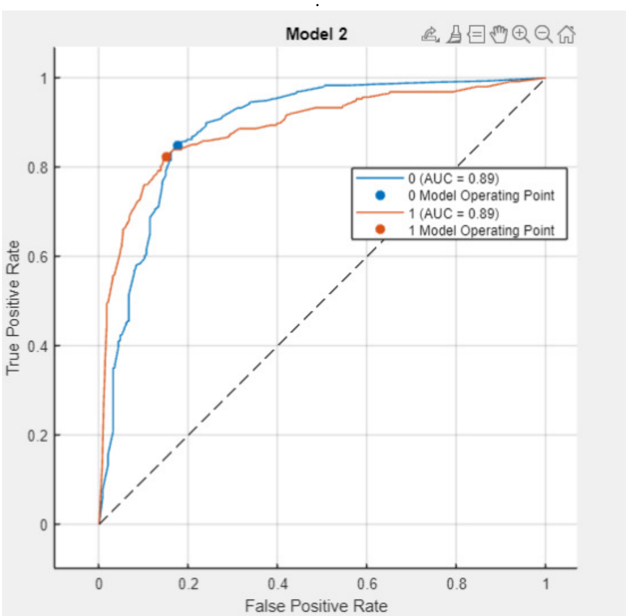


Figure 6 AUC Curve

In Figure 6, the ROC curve of the model is depicted, demonstrating its performance across different threshold values. The Area Under the ROC Curve (AUC) is a crucial metric for evaluating the model's ability to distinguish between classes. Our model achieved an AUC of 0.89, which signifies a high level of performance, as the AUC value ranges from 0.5 (no discrimination) to 1 (perfect discrimination).Table 1 provides a detailed breakdown of how the accuracy metrics were calculated. It shows the contributions of true positives, false positives, true negatives, and false negatives to the overall accuracy, reflecting the model's performance in various aspects. The accuracy, confusion matrix, and AUC collectively demonstrate the Ensemble Bagged Trees Algorithm's robustness and effectiveness in this application.

V.CONCLUSIONS

Using our pre-processed datasets, we trained and evaluated the Ensemble Bagged Trees Algorithm in this study. The results we obtained are impressive and show the algorithm's effectiveness and dependability. Our efforts have yielded several important conclusions that highlight the advantages and possibilities of our machine learning methodology.First, the model performed admirably, with an accuracy of 83.7%. This high degree of precision indicates that the Ensemble Bagged Trees Algorithm successfully identified the underlying patterns in the data, allowing it to provide accurate predictions. In addition, the model trained quite quickly, analyzing 3300 observations per second and finishing the training in just 35.7 seconds. This efficiency demonstrates the algorithm's applicability for large-scale applications where time is crucial and highlights its computational efficacy.The ROC curve study further validated the model's robustness, which showed an Area Under the Curve (AUC) of 0.89. The model's excellent discriminative capacity is demonstrated by its high AUC value, which also shows how well it can discriminate between positive and negative classes at various threshold values. AUC is a crucial measure of model performance and getting a number near 0.9 here emphasizes how well the Ensemble Bagged Trees Algorithm works in this situation.As a result of research study, we have concluded that the Ensemble Bagged Trees Algorithm is a very reliable and accurate model. The robustness and practical deployment possibilities of the model are demonstrated by the high AUC value and favorable results obtained from the confusion matrix. This work demonstrates the great potential of ensemble approaches in machine learning and opens up new avenues for investigation and improvement in further studies.

REFERENCES

[1] C. E. Stafstrom and L. Carmant, “Seizures and Epilepsy: An overview for neuroscientists,” Cold Spring Harbor Perspectives in Medicine, vol. 5, no. 6, p. a022426, Jun. 2015, doi: 10.1101/cshperspect.a022426.

[2] V. S. S. Wong, “The causes of epilepsy: common and uncommon causes in adults and children,” JAMA, vol. 307, no. 18, p. 1979, May 2012, doi: 10.1001/jama.2012.4478.

[3] K. Rasheed et al., “Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A review,” IEEE Reviews in Biomedical Engineering, vol. 14, pp. 139–155, Jan. 2021, doi: 10.1109/rbme.2020.3008792.

[4] M. J. Eadie, “Shortcomings in the current treatment of epilepsy,” Expert Review of Neurotherapeutics, vol. 12, no. 12, pp. 1419–1427, Dec. 2012, doi: 10.1586/ern.12.129.

[5] P. R. Kunekar, M. Gupta, and B. Agarwal, “Detection and Analysis of Life Style based Diseases in Early Phase of Life: A Survey,” in Communications in computer and information science, 2019, pp. 55–69. doi: 10.1007/978-981-13-8300-7_6.

[6] P. Kunekar, “The effects of artificial intelligence and medical technology on the life of human,” Mumbaiunivercity, Jul. 2022, [Online]. Available: https://www.academia.edu/63348031/The_Effects_of_Artificial_Intelligence_and_Medical_Technology_on_the_Life_of_Human

[7] Recent research has examined how machine learning and deep learning approaches can be effectively applied to improve medical breakthroughs.

[8] B. Arora, P. Jadhav, R. B. Sulaiman, A. Kareem, P. Kunekar, and B. Pant, “Integrating artificial intelligence and deep learning for enhanced medical innovation,” IC3I, Dec. 2022, doi: 10.1109/ic3i56241.2022.10073054.

[9] P. Kunekar, R. Vaishnav, E. Kalaivani, D. Gangodkar, C. Kaur, and J. A. Dhanraj, “Applications of Machine Learning Techniques in Detecting Skin Cancer,” IC3I, Dec. 2022, doi: 10.1109/ic3i56241.2022.10072834.

[10] I. S. Chakrapani, N. Tyagi, S. Tyagi, P. Kunekar, D. L. Padmaja, and K. Pant, “Applications of deep learning (DL) techniques in detecting breast cancer and malignant cells,” IC3I, Dec. 2022, doi: 10.1109/ic3i56241.2022.10072720.

[11] S. M. Usman, M. Usman, and S. Fong, “Epileptic seizures prediction using machine learning methods,” Computational and Mathematical Methods in Medicine, vol. 2017, pp. 1–10, Jan. 2017, doi: 10.1155/2017/9074759.

[12] M. Li, W. Chen, and T. Zhang, “Automatic epilepsy detection using wavelet-based nonlinear analysis and optimized SVM,” Biocybernetics and Biomedical Engineering, vol. 36, no. 4, pp. 708–718, Jan. 2016, doi: 10.1016/j.bbe.2016.07.004.

[13] P. Potapov, “On the loss of information in PCA of spectrum-images,” Ultramicroscopy, vol. 182, pp. 191–194, Nov. 2017, doi: 10.1016/j.ultramic.2017.06.023.

[14] R. Harikumar and B. S. Narayanan, “Fuzzy techniques for classification of epilepsy risk level from EEG signals,” IEEE, Jun. 2004, doi: 10.1109/tencon.2003.1273316.

[15] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, “EEG based multi-class seizure type classification using convolutional neural network and transfer learning,” Neural Networks, vol. 124, pp. 202–212, Apr. 2020, doi: 10.1016/j.neunet.2020.01.017.

[16] N. Aayesha, M. B. Qureshi, M. Afzaal, M. S. Qureshi, and M. Fayaz, “Machine learning-based EEG signals classification model for epileptic seizure detection,” Multimedia Tools and Applications, vol. 80, no. 12, pp. 17849–17877, Feb. 2021, doi: 10.1007/s11042-021-10597-6.

[17] D. A. Torse and R. Khanai, “Classification of epileptic seizures using ensemble empirical mode decomposition and least squares support vector machine,” IEEE, Jan. 2021, doi: 10.1109/iccci50826.2021.9402307.

[18] H. Berger, “Über das Elektrenkephalogramm des Menschen,” Archiv Für Psychiatrie Und Nervenkrankheiten, vol. 87, no. 1, pp. 527–570, Dec. 1929, doi: 10.1007/bf01797193.

[19] N. Srivastava, “A logistic regression model for predicting the occurrence of intense geomagnetic storms,” Annales Geophysicae, vol. 23, no. 9, pp. 2969–2974, Nov. 2005, doi: 10.5194/angeo-23-2969-2005.

[20] P. Reed and Y. Wu, “Logistic regression for risk factor modelling in stuttering research,” Journal of Fluency Disorders, vol. 38, no. 2, pp. 88–101, Jun. 2013, doi: 10.1016/j.jfludis.2012.09.003.

[21] J. L. Dioses, R. P. Medina, A. C. Fajardo, and A. A. Hernandez, “Performance of Classification Models in Japanese Quail Egg Sexing,” IEEE, Mar. 2021, doi: 10.1109/cspa52141.2021.9377275.

[22] M. A. A. Ozaeta, A. C. Fajardo, F. P. Brazas, and J. A. M. Cantal, “Deep Learning Approach for Seagrass Species Classification,” IEEE, Jan. 2024, doi: 10.1109/gecost60902.2024.10474987.

[23] S. M. Rasdas, A. C. Fajardo, and J. S. Limbago, “Predicting Abundance of Fish Species Populations in Manila Bay, Philippines Based on Ensemble Learning Approach,” IEEE, Jun. 2023, doi: 10.1109/jcsse58229.2023.10201953.

[24] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, “Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors,” IEEE, May 2018, doi: 10.1109/icird.2018.8376326.

[25] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, “An image processing technique for coffee black beans identification,” IEEE, May 2018, doi: 10.1109/icird.2018.8376325.

[26] L. Breiman, “Bagging predictors,” Machine Learning, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/bf00058655.

[27] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, “On combining classifiers,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226–239, Mar. 1998, doi: 10.1109/34.667881.

[28] N. T. K. Ho, J. J. Hull, and S. N. Srihari, “Decision combination in multiple classifier systems,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 1, pp. 66–75, Jan. 1994, doi: 10.1109/34.273716.

[29] A. Merghadi et al., “Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance,” Earth-science Reviews, vol. 207, p. 103225, Aug. 2020, doi: 10.1016/j.earscirev.2020.103225.

[30] M. Budnik and B. Krawczyk, “On optimal settings of classification tree ensembles for medical decision support,” Health Informatics Journal, vol. 19, no. 1, pp. 3–15, Mar. 2013, doi: 10.1177/1460458212446096.

[31] "Medical Image Processing in MATLAB", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.4, Issue 11, page no.547-552, November-2017, Available :http://www.jetir.org/papers/JETIR1711093.pdf