

# Epileptic Seizure Classification Using Logistic Regression Approach

Isaac Angelo M. Dioses\*, Generaldo D. Maylem, MDFPNA\*\*, 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](mailto:jdiosesjr@gmail.com)

\*\*\*\*\*MAPUA University, School of Information Technology, Makati Campus, Philippines  
Email:debbiemeer30@gmail.com

\*\*\*\*\*

## Abstract:

People with epilepsy often have unprovoked, unexpected seizures that render them helpless and increase their risk of suffocation, fainting, or injury from automobile accidents. Machine Learning and AI nowadays are very useful. Researchers are using these technologies and techniques in medicine, etc. This research has effectively developed a machine learning model for logistic regression in this work, which outperforms the model that was previously employed in another investigation. Using a Logistic Regression classifier in MATLAB with 4500 observations per second and the default binary GLM LR hyperparameter, the datasets are split into 5-fold cross-validation with an accuracy of 72.5% with 667 features and 2216 examples. This is very effective compared to the previous study, which used a 33% data split. Additionally, the confusion model's AUC curve of 0.7237 and logistic regression results indicate it is extremely promising.

**Keywords** —Machine Learning, Artificial Intelligence, Logistic Regression, Brain, Epilepsy, AI Medicine

\*\*\*\*\*

## I. INTRODUCTION

Untreated epilepsy can be fatal, despite being a common neurological illness. All ages are affected by this illness. A seizure requires a sophisticated chemical shift in the brain's nerve cells. These chemical alterations occur in the positive and negative ion-containing nerve cells, which produce electrical signals [1]. Individuals who have epilepsy experience unprovoked, unexpected seizures that

leave them defenseless and leave them at risk of asphyxia, fainting, or harm from car crashes [2,3]. Every shaking event has the potential to cause bodily harm and can linger for short or extended periods of time. Because epilepsy seizures frequently reoccur without warning or apparent underlying cause, those who have the condition can find it difficult to socialize to varied degrees. Most epilepsy cases have an enigmatic cause [4]. While certain forms of illness are directly linked to genetic

changes, most cases lack clear explanations. Some cases may be caused by brain injury, infections, or tumors [5]. A transient aberrant electrical discharge that happens in the cerebral networks and lasts for less than a few minutes is the cause of an epileptic seizure (ES). In addition to being unpredictable in terms of intensity and length, ES attacks are difficult to forecast [6]. Many people with epilepsy get better to the point that they no longer require medication, therefore not all occurrences of the condition are permanent. In over 70% of cases, medication can be used to control seizures [7]. An especially useful diagnostic technique for examining the functional architecture of the brain during an ES attack is electroencephalography (EEG). With EEG, epilepsy prediction and treatment have been extensively researched. The non-Gaussian and non-stationary EEG signals are used to assess the electrical activity in the brain, which is then utilized to diagnose the many types of brain illnesses. The division of normal from pathological brain function is facilitated by the examination of EEG readings. Longer EEG recordings need to be examined to accurately predict epilepsy [8]. The four distinct states of epileptic seizures are: the preictal state, which occurs prior to the onset of the seizure; the ictal state, which begins with the seizure and ends with an attack; the postictal state, which follows the ictal state; and the interictal state, which follows the postictal state of the first seizure and ends prior to the onset of the preictal state of a subsequent seizure [4]. To address this issue, several research have used prediction techniques based on machine learning (ML) over the years. Deep learning (DL) is a sophisticated machine learning technique that uses a multi-layer hierarchical design to process data to identify patterns more precisely from enormous datasets [4,6,9,8,10,11]. Artificial Intelligence are widely used in these days, from agriculture, to aquamarines and medical fields [24,25,30,31,32,33,34]

## **II. RELATED LITERATURE**

Current seizure detection approaches rely on manually designed feature extraction methods from EEG signals, including time domain, frequency

domain, time-frequency domain, and nonlinear signal analysis [12,13,14,15]. Most of the methods make use of very simple algorithms, such as linear and non-linear SVMs, binary linear classifiers, k-nearest neighbours, linear Bayes classifiers, and neural networks in some cases. Numerous methods employ algorithms tailored to individual patients [4]. In one study, the Short-Time Fourier Transform was used to transform raw EEG data into a matrix that could be fed into a 2D CNN. The frequency and time axes would be included in the resulting two-dimensional matrix in this manner. Their goal was to forecast seizures that would occur five to thirty-five minutes in advance. They were able to get a sensitivity of roughly 0.81 on two datasets and 0.75 on the third [16].

Convolutional neural networks tailored for ultra-low power implantable microcontrollers were used in a different investigation to identify seizures early. They were able to identify the seizure six seconds ahead of time, with a median sensitivity of 0.96 and a median ROC-AUC of 0.89 across 24 individuals [17].

Three patients' more than one thousand seizures were studied by researchers. To create a model based on the knowledge they had acquired, they investigated theoretical ideas regarding epileptic seizures and how they could be predicted. They reduced the amount of data by using effective preprocessing techniques, and they also employed the Short-Time Fourier Transform to transform the time-series iEEG data into an image-like format that convolutional neural networks could use as inputs. Cooperative multi-scale CNNs are used in the proposed seizure prediction system to automatically learn features from iEEG data. With an average sensitivity of roughly 0.88 and ROC-AUC of 0.84, the model outperformed earlier methods that attempted to predict seizures five minutes in advance [18,19].

The authors of a different study recommend applying deep learning to recognize seizures in pediatric patients. The ictal and interictal brain state signals are classified using the CHB-MIT dataset to apply the supervised classifier. neural network-connected two-dimensional deep convolution auto-encoder [20].

The model presented in this study uses sampling approaches such as down sampling, random sampling, and the synthetic minority oversampling methodology to address the problems of data imbalance, low accuracy, and categorization. The heterogeneous deep ensemble model, which the authors proposed, has an F-measure value of 0.91 and an accuracy score of 0.93 [21]. Over the past 20 years in particular, the use of logistic regression (LR) analysis as a statistical technique in medical research has increased [22]. When the goal of the research methodology is to determine whether an event happened rather than when it did, longitudinal research (LR) is employed (time course information is not used). It is frequently employed in research in the health sciences because it is especially suitable for models involving illness condition (diseased or healthy) and decision making (yes or no). More sophisticated versions, known as polychotomous or multinomial logistic regression, can handle scenarios in which the predicted variable falls into more than two categories [23].

### II.A. LOGISTIC REGRESSION

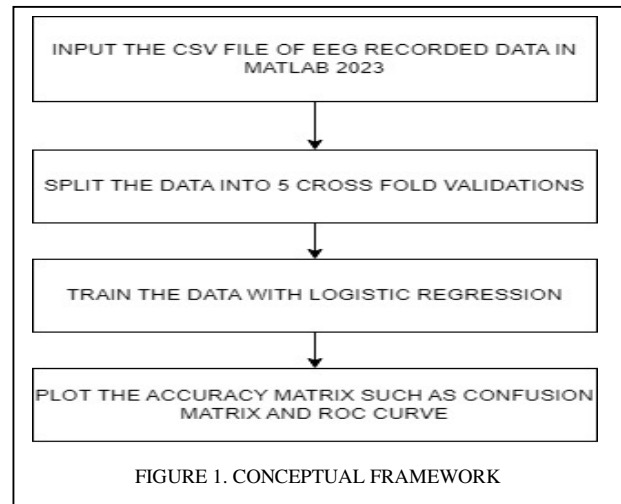
Over the past 20 years in particular, the use of logistic regression (LR) analysis as a statistical technique in medical research has increased. When the goal of the research methodology is to determine whether an event happened rather than when it did, longitudinal research (LR) is employed (time course information is not used). It is frequently employed in research in the health sciences because it is especially suitable for models involving illness condition (diseased or healthy) and decision making (yes or no). More sophisticated versions, known as polychotomous or multinomial logistic regression, can handle scenarios in which the predicted variable falls into more than two categories. Medical research is using Logistic Regression algorithm and more. A study of the factors that predict whether an intervention will result in an improvement or not, or the presence or absence of a disease in connection to various factors, are examples of how logistic regression is used in medicine. Medical research is using LR more and more [24,25]. A study of the factors that predict whether an intervention will result in an

improvement or not, or the presence or absence of a disease in connection to various factors, are examples of how logistic regression is used in medicine [26,27,28].

In recent study, Logistic regression approach for detecting epileptic seizures in EEG signals: the model employed test spilt, which accounted for 33% of the dataset for validation. The model's validation accuracy is 63.9%, and its training accuracy is 66.92%. False-positive rate: 32.83%, false-negative rate: 0.0%, true-negative value percentage: 67.17%, true-positive value percentage: 0.0%. that is quite low. In this Research, we will use logistic regression with 5 cross fold validation to compare the previous study results and assume that it will improve the accuracy of detecting epileptic seizures.

### III. METHODOLOGY

In the figure below, it shows the conceptual framework of this research.



#### A. Datasets

The datasets were obtained in Kaggle [29] it consists of 2217 datasets and records of different EEG readings, and it will be cross validated by one of our Neurologist authors for the integrity of results.

#### B. MATLAB Software

MATLAB's Image Processing Toolbox offers an extensive collection of industry-standard graphical tools and techniques for image processing, analysis,

creation of algorithms, and visualization [35]. The MathWorks product MATLAB has gained popularity as a tool for rapid development in recent years. Its numerous Toolboxes, robust interface, and ease of use make it a preferred tool across numerous fields, including medical image processing [36]. In this study we will use MATLABR2023B for machine learning process.

**C. Importing Datasets**

Figure 2. EEG Dataset Importation using MATLAB.

In Figure 2, shows the some keypoints values extracted from brain using EEG such as C3, C4, P3, O4 etc. with a total of 667 features extracted from brain,

**D. Classification Learner**

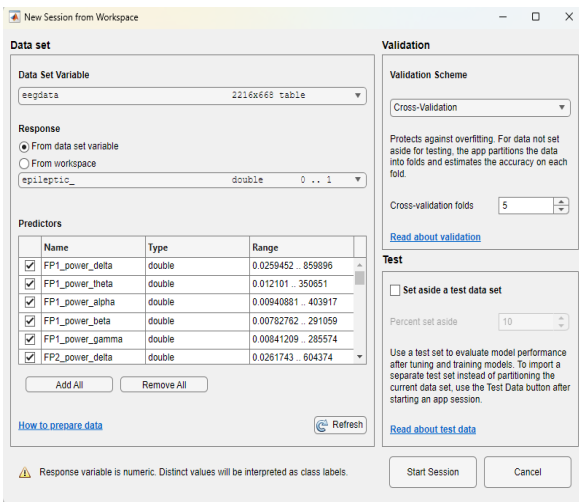


Figure 3. Data Preprocessing

In Figure 3. It shows the preparation of datasets to learn in MATLAB classification learner. It splits the data into cross validation folds of 5 with 667 features with 2216 samples.

**IV. RESULTS AND DISCUSSION**

After importing datasets and preprocessing, we trained our datasets to Binary GLM Logistic Regression Algorithm with 72.5% Accuracy surpassing the previous study with 66.92% with 4500 observations processed per second with training time of 85 seconds with preset binary GLM Logistic Regression hyperparameters.

As shown in figure 4, the confusion matrix of the trained model is quite promising, with the true positive of 861, false positive of 315, false negative of 295, and true negative of 725. Meanwhile in figure 5, the roc curve of the model has AUC or Area Under the ROC Curve which under the whole ROC curve in two dimensions is measured has 0.7237 AUC. In table 1 it shows how accuracy matrix was calculated.

TABLE I  
 ACCURACY FORMULA

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

**A. Accuracy Matrix**

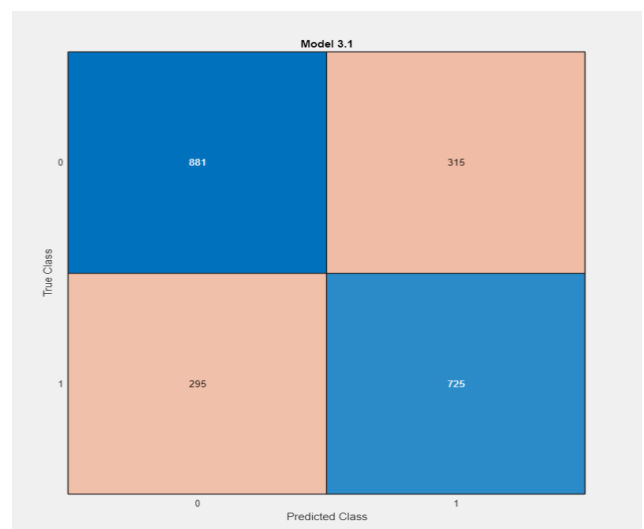


Figure 4. Confusion Matrix

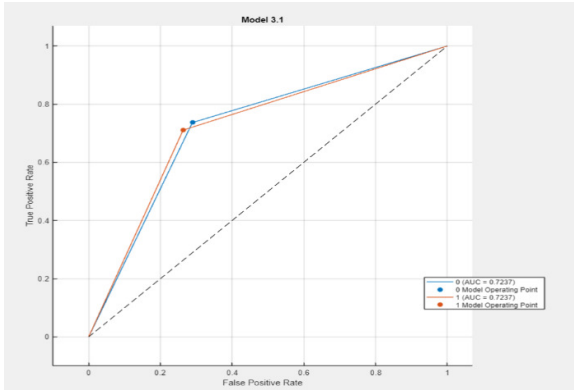


Figure 5. Roc Curve of the trained model

## CONCLUSIONS

Even though epilepsy is a common neurological condition, it can be fatal if left untreated. This disease affects people of all ages. A complex chemical change in the nerve cells of the brain is necessary for a seizure. In this study, we successfully created a Logistic regression machine learning model that surpasses the previous model that was used in past study. With the accuracy of 72.5% with 667 features and 2216 examples Using Logistic Regression classifier in MATLAB with 4500 observations per seconds and default binary GLM LR hyperparameter splitting the datasets into 5-fold cross validation which is very effective versus the last study which they use 33% data split. It also shows the confusion model is quite promising in logistic regression and AUC curve of 0.7237. Overall, the model that was developed in this research successfully outperforms the past studies.

This study is just a preliminary research and foundation for the EEG results using logistic regression. Moreover, it is recommended to collect more datasets, validate more, and use image processing techniques as well as other binary classification techniques and DL models such as YOLO, Faster RCNN, Mobilenet SSD v2 and applied it in real life.

## ACKNOWLEDGMENT

One of the authors, Isaac Angelo Diones, wants to acknowledge and give thanks to Atty. Janet Abuel and her family. Dr. Elizabeth Alonzo, Dr. Teresita Molano, and my mentors for endless support.

## REFERENCES

- [1] H. U. Amin, M. Z. Yusoff, and R. F. Ahmad, "A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques," *Biomedical Signal Processing and Control*, vol. 56, p. 101707, Feb. 2020, doi: 10.1016/j.bspc.2019.101707
- [2] T. Yan, Y. Geng, J. Wu, and C. Li, "Interactions between multisensory inputs with voluntary spatial attention," *NeuroReport/Neuroreport*, vol. 26, no. 11, pp. 605–612, Aug. 2015, doi: 10.1097/wnr.0000000000000368.
- [3] A. Y. Mutlu, "Detection of epileptic dysfunctions in EEG signals using Hilbert vibration decomposition," *Biomedical Signal Processing and Control*, vol. 40, pp. 33–40, Feb. 2018, doi: 10.1016/j.bspc.2017.08.023.
- [4] P. P. S. S. M., and S. Kattepara, "Machine Learning Models for Epileptic Seizure Prediction," *IEEE*, Apr. 2023, doi: 10.1109/icit57646.2023.10134350.
- [5] "World Health Organization (WHO)", *Epilepsy*, February 2018.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, p. 051001, Aug. 2019, doi: 10.1088/1741-2552/ab260c.
- [10] G. Li, C. H. Lee, J. J. Jung, Y. C. Youn, and D. Camacho, "Deep learning for EEG data analytics: A survey," *Concurrency and Computation*, vol. 32, no. 18, Feb. 2019, doi: 10.1002/cpe.5199.
- [11] M. Zhou et al., "Epileptic seizure detection based on EEG signals and CNN," *Frontiers in Neuroinformatics*, vol. 12, Dec. 2018, doi: 10.3389/fninf.2018.00095.
- [12] T. Yan et al., "Increased local connectivity of brain functional networks during facial processing in schizophrenia: evidence from EEG data," *Oncotarget*, vol. 8, no. 63, pp. 107312–107322, Sep. 2017, doi: 10.18632/oncotarget.20598.
- [13] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowledge-based Systems*, vol. 45, pp. 147–165, Jun. 2013, doi: 10.1016/j.knosys.2013.02.014.

- [14] G. Pei et al., "Effects of an Integrated Neurofeedback System with Dry Electrodes: EEG Acquisition and Cognition Assessment," *Sensors*, vol. 18, no. 10, p. 3396, Oct. 2018, doi: 10.3390/s18103396.
- [15] M. De Lucia, J. Fritschy, P. Dayan, and D. S. Holder, "A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis," *Medical & Biological Engineering & Computing*, vol. 46, no. 3, pp. 263–272, Dec. 2007, doi: 10.1007/s11517-007-0289-4.
- [16] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, and O. Kavehei, "A Generalised Seizure Prediction with Convolutional Neural Networks for Intracranial and Scalp Electroencephalogram Data Analysis," arXiv.org, Jul. 06, 2017. <https://arxiv.org/abs/1707.01976>.
- [17] M. Hugle et al., "Early Seizure Detection with an Energy-Efficient Convolutional Neural Network on an Implantable Microcontroller," *IEEE*, Jul. 2018, doi: 10.1109/ijcnn.2018.8489493.
- [18] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," arXiv (Cornell University), Jan. 2016, doi: 10.48550/arxiv.1609.02907.
- [19] R. Hussein, M. O. Ahmed, R. Ward, Z. J. Wang, L. Kuhlmann, and Y. Guo, "Human intracranial EEG quantitative analysis and automatic feature learning for epileptic seizure prediction," arXiv (Cornell University), Jan. 2019, doi: 10.48550/arxiv.1904.03603.
- [20] A. M. Abdelhameed and M. A. Bayoumi, "A deep learning approach for automatic seizure detection in children with epilepsy," *Frontiers in Computational Neuroscience*, vol. 15, Apr. 2021, doi: 10.3389/fncom.2021.650050.
- [21] M. A. a R. Dr. J. V. F. Sarkar Dr. Rashal, "Multi-Feature learning model for epilepsy classification supervised by a highly robust heterogeneous deep ensemble," *turcomat.org*, Mar. 2022, doi: 10.17762/turcomat.v13i2.12203.
- [22] T. Oommen, L. G. Baise, and R. M. Vogel, "Sampling bias and class imbalance in maximum-likelihood logistic regression," *Mathematical Geosciences*, vol. 43, no. 1, pp. 99–120, Oct. 2010, doi: 10.1007/s11004-010-9311-8.
- [23] E. Boateng and F. Oduro, "Predicting Microfinance Credit Default: A Study of NSOaTreman Rural Bank, Ghana," *Journal of Advances in Mathematics and Computer Science*, vol. 26, no. 1, pp. 1–9, Jan. 2018, doi: 10.9734/jamcs/2018/33569.
- [24] 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.
- [25] 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.
- [26] K. S. Khan, P. F. W. Chien, and L. S. Dwarakanath, "Logistic regression models in obstetrics and gynecology literature," *Obstetrics and Gynecology (New York. 1953. Online)/Obstetrics and Gynecology*, vol. 93, no. 6, pp. 1014–1020, Jun. 1999, doi: 10.1097/00006250-199906000-00024.
- [27] Y. Kim, S. Kwon, and S. H. Song, "Multiclass sparse logistic regression for classification of multiple cancer types using gene expression data," *Computational Statistics & Data Analysis*, vol. 51, no. 3, pp. 1643–1655, Dec. 2006, doi: 10.1016/j.csda.2006.06.007.
- [28] P. Howell and S. Davis, "Predicting Persistence of and Recovery from Stuttering by the Teenage Years Based on Information Gathered at Age 8 Years," *Journal of Developmental and Behavioral Pediatrics/Journal of Developmental & Behavioral Pediatrics*, vol. 32, no. 3, pp. 196–205, Apr. 2011, doi: 10.1097/dbp.0b013e31820fd4a9.
- [29] "Epilepsy detection using EEG signals," Kaggle, May 19, 2021. <https://www.kaggle.com/datasets/oussamabenhassine/epilepsy-detection-using-eeeg-signals/data>
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] "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>
- [36] M. Bister, C. Yap, K. Ng, and C. Tok, "Increasing the speed of medical image processing in MatLab@," *Bijj*, vol. 3, no. 1, Jan. 2007, doi: 10.2349/bijj.3.1.e9.