SAJSET
ISSN: 2992-3581

Savannah Journal of Science and Engineering Technology
Volume: 01; Issue: 05; Pages: 257-262; 2023.
www.sajsetjournal.com.ng
DOI: xx.xxxx.x/sajset-01-2023-0053

SAJSET
ISSN: 2992-3581

A model for predicting household cooking energy sources using multi-layer perceptron technique

[1,2]Tabra, M. S., [1]Ahmadu, A. S. and [1]Baha, B. Y.
[1]*Department of Computer Science, Modibbo Adama University, Yola, Nigeria.*
[2]*Department of Computer Science, Gombe State University, Gombe, Nigeria.*

Paper History
Received: 20th August, 2023
Accepted: 30th August, 2023
Published: Sept, 2023

Corresponding author
Tabra, M. S.
munattabra@gmail.com

Abstract:
Cooking with dirty energy is one of the causes of CO_2 emission, and adversely affects climate change. In Nigeria, most households are using dirty sources of energy for cooking such as firewood, charcoal and kerosene due to availability and cheapness. The aim of the research is to develop a Machine Learning (ML) model to determine households' cooking energy sources. In this study, a Multi-layer Perceptron (MLP) model was developed to classify households' cooking energy sources in Nigeria. Secondary dataset obtained from the National Bureau of Statistics (NBS) database for Nigerian General Household Panel Survey (NGHPS) was used. Dataset were pre-processed, and Forward Sequential Feature Selection technique was used to select the most relevant features. A dataset consisting a total of 4,980 households was used to develop MLP Architecture by varying the number of hidden layers (1 and 2 layers), and the number of units were selected in multiples of 10 and 5 for 1 and 2 layers respectively. The best accuracy of 96.89% was obtained using 1 –hidden layer at (60) units in the hidden layer, and 97.09% at (70, 70) units using 2-hidden layer architecture. In general, the study shows the significance of using 2-hidden layers architecture and the influence from the use of variations in the number of units in the hidden layers to find the optimal number of units for the problem under question. The study shows that given some specific household's characteristics, the household's Major Cooking Fuel Source in Nigeria can be determine with an accuracy of 97%, precision of 95%, recall and F1 Score of 94% each. In future, this study can be improved by exploring other Machine Learning Techniques for better performance.

Keywords: *Cooking Energy Sources, Cooking Fuel, Multi-layer Perceptron, Hyperparameters*

1. Introduction

Climate change poses a significant global challenge. The use of solid fuels for cooking presents severe health risks, particularly for women and children [1]. Household fuel choices also contribute to climate change through greenhouse gas emissions [2]. The smoke produced from burning wood fuel is a major factor in acute respiratory infections (ARI) among young children. Moreover, in many developing countries, reliance on biomass for cooking and heating exposes women and children to high levels of indoor air pollution [1]. A study conducted in the Commonwealth of Independent States (CIS) region, including Kazakhstan and the Kyrgyz Republic, among other countries, revealed a positive relationship between CO_2 emissions (and fossil fuel consumption) and mortality from cardiovascular disease, diabetes mellitus, cancer, and chronic respiratory disease [3].

Nigeria, with a population of over 200 million people, is one of the most densely populated countries. It possesses abundant natural resources, including various energy sources used for household cooking such as Liquefied Petroleum Gas (LPG), electricity, firewood, charcoal, animal dung, and kerosene. These energy sources can be classified as either clean or dirty [4]. Clean energy sources include LPG and electricity, while firewood, charcoal, and animal dung fall under dirty energy sources. In Nigeria, firewood is commonly used for cooking, and kerosene is used for lighting [5]. The demand for dirty energy sources outweighs that of clean energy sources in the country due to poor electricity supply and high LPG costs [6]. This situation places a significant burden on the environment due to the country's large population [4].

The United Nations, in its 17 Sustainable Development Goals (SDG), includes Item 7, aiming to ensure access to affordable, reliable, sustainable, and modern energy for all by 2030 (Resolution of the General Assembly on 25 September 2015) [3]. To address the transition from dirty to clean energy, it is crucial to understand the energy sources used by households across the country. In recent years, machine learning (ML) has emerged as a powerful tool for technological advancement [7]. Numerous studies have utilized machine learning algorithms to develop models for predicting household energy usage and demand. For instance, a study investigated the predictors of charcoal and firewood use for cooking using the 2014 Uganda census dataset [8]. They employed a multinomial logistic regression model to identify factors influencing household charcoal and firewood use compared to electricity. Additionally, an SVM-based data-driven model was developed to predict lighting energy consumption in office buildings in Philadelphia, PA, with promising results [9].

Moreover, researchers conducted a survey to compare the accuracy of machine learning-supported regression models with conventional regression models for energy consumption forecasting [10].

Furthermore, a model based on Support Vector Regression was proposed to predict household electricity consumption under multiple behavioural intervention strategies, aiding decision-making for different households [11]. The studies of [12] proposed a hybrid machine learning model combining multi-layered perceptron, support vector regression, and cat boost to forecast energy consumption using load data from renewable and non-renewable energy sources.

Additionally, a neural network-based model was compared with a conditional demand analysis (CDA) method for modelling residential end-use energy consumption in the Canadian residential sector [13], the study found that neural networks outperformed CDA, especially in assessing the impact of socioeconomic factors on energy consumption patterns. Lastly, MLP was used to classify pneumonia based on X-Ray pictures, yielding an accuracy of 82.8%, along with precision, recall, and f1-score all at 82.8% [14].

1.1. Energy sources

Energy sources are natural resources that can directly or indirectly provide heat, light, or power through conversion or transformation. These sources are categorized as clean or dirty energy sources. Clean energy sources encompass LPG and electricity produced from renewable sources like hydro, solar, wind, and other constantly replenished sources. These energy options have a minimal presence of harmful carbon-dioxide (CO_2) emissions, which are known to have a detrimental effect on the environment [4]. On the contrary, dirty energy refers to energy production methods that contribute to climate change and harm communities, particularly in the global south. Many dirty energy projects involve the extraction or burning of fossil fuels (such as oil, gas, and coal) for electricity generation, leading to the release of carbon dioxide into the atmosphere, a major cause of climate change [15].

1.2. Multi-layer Perceptron

The Multi-layer Perceptron (MLP) is one of the most widely used Neural Network architectures today, initially introduced by Rumelhart and McClelland [16]. The term "Neural Networks" originates from their early development as models inspired by the human brain, where each unit represents a neuron, and the connections represent synapses. In early models, neurons fired when the total signal surpassed a certain threshold. However, neural networks were later recognized as valuable tools for nonlinear statistical modelling [17]. In an MLP, each unit performs a biased weighted sum of its inputs and passes this activation level through a transfer function to produce the output. The units are organized in a layered feed-forward topology as shown in Figure 1. The network can be seen as an input-output model, with the weights and thresholds (biases) serving as free parameters of the model. Such networks can model functions of high complexity, with the number of layers and units in each layer determining the complexity of the function. Key design considerations for Multilayer Perceptron include determining the number of hidden layers and the number of units in those layers [14, 18]. The optimal number of hidden units to use is not clearly defined. One common starting point is to use one hidden layer, with the number of units equal to half the sum of the number of input and output units [16].

2. Research Methods

2.1 Dataset

In the current study, a secondary data was acquired from the NBS database. The data was collected through a nationally representative Nigerian General Household Panel Survey (NGHPS) conducted by the National Bureau of Statistics in Nigeria.

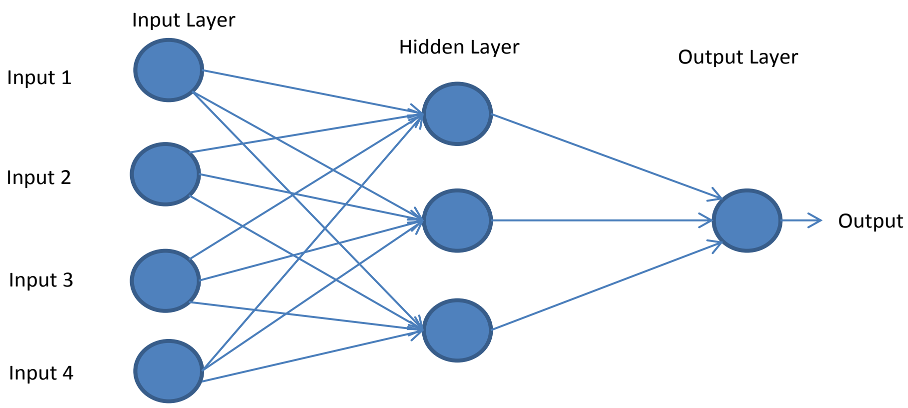

Figure 1: Multi-layer Perceptron Architecture Example

The data was collected through face-to-face interviews with household members and covered a wide range of topics related to household demographics, socio-economic status and fuel expenditure practice. The households' fuel expenditure practice determines their major cooking fuel. The dataset used consists of the records of 4,980 households.

2.2 Data Visualization

The dataset contains information about various household characteristics, such as the size of the household, its geographic location, and the socioeconomic status of its members. Based on the household fuel expenditure practice, Figure 2 shows the distribution of major cooking fuel used by the households in our dataset.

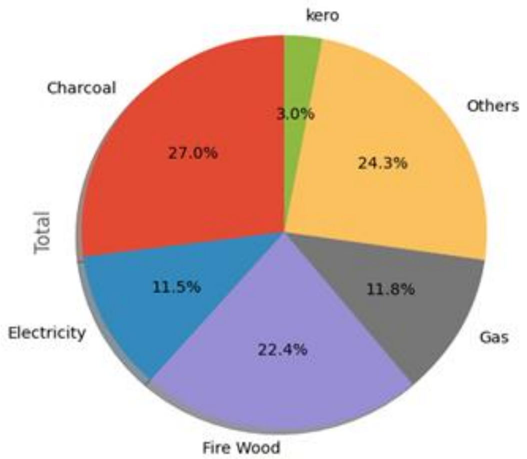

Figure 2: Distribution of Major Cooking Fuel used by the Households

2.3 Research Architecture

The study architecture is shown in Figure 3 representing both the training and testing phases. The training steps consist of data Preprocessing where imputation techniques were used to fill the missing values and Normalization was used to normalize the dataset within a scale range (0,1); feature selection was conducted using Forward Sequential Feature Selection Technique to select the most relevant features in the dataset; Later the selected set of features were used to train the MLP classifier. The testing phase is the process of matching the actual target value with the value predicted by the trained classifier.

2.4 Performance Measure

Confusion Matrix was used to evaluate the performances of the models on the test dataset using the metrics Accuracy, Precision, Recall and F1 Score given by equation 1 to 4:

$$Accuracy = (TP+TN)/(PN) \quad (1)$$

$$Precision = TP/(TP+FP) \quad (2)$$

$$TPR = TP/P \text{ or } TP/(TP+FN) \quad (3)$$

$$F1\,Score = 2 * (Precision * Recall)/(Precision + Recall) \quad (4)$$

2.5. Feature Selection

The selected Features using Forward Sequential Feature Selection are Zone, State, Local Government Area (LGA), Sector, Enumeration Area (EA), Size, Age of Household Head (HHHAge), Age of Pouse (SpouseAge), Sex of Household Head (HHHSex). Others include Education Level of Household Head (EduHHH), Education of Spouse (EduSpouse), Occupation of Household Head (OccuHHH), Occupation of Spouse (OccuSpouse), Total Earning from Occupation (EarningOccu), Monetary Assistance from Abroad (MonAssistAbroad), Expenditure on Charcoal (CharcoalAmount), Expenditure on Gas (GasAmount), Expenditure on Electricity (ElectAmount), Expenditure on Firewood (FWoodAmount) and Expenditure on Kerosene (KeroAmount) were considered totalling 20 features in the dataset.

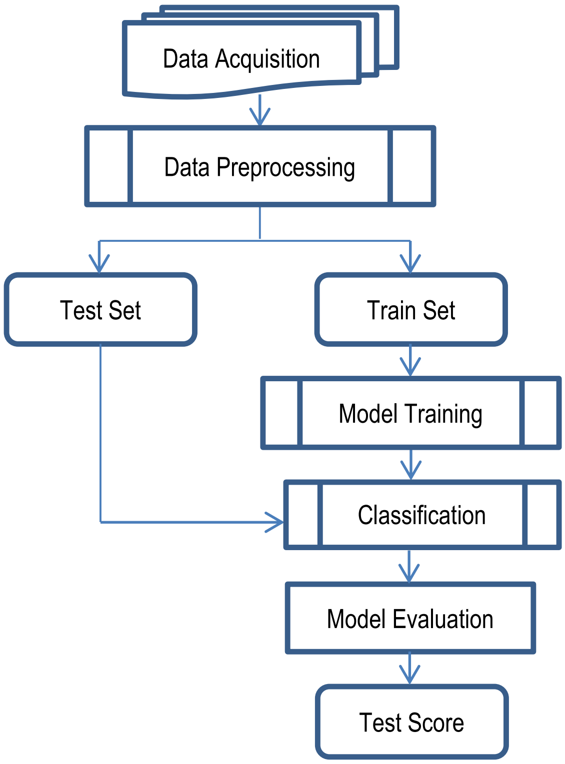

Figure 3: Research Architecture

3. Result and Discussion

3.1 Result

The classification process using a multilayer perceptron was tested with variations in the number of Hidden layers and number of Units in the Hidden layers to determine the optimal level of accuracy on the multilayer perceptron in classifying household major cooking energy sources. The Variations in the Number of Hidden layers are [1; 2]. For one (1) hidden layer, the variation in the number of units is 1; 10; 20; 30; 40; 50; 60; 70; 80; 90; 100; 110; 120; 130; 140; 150. For Two (2) Hidden Layers, the variations in the number of Units in the hidden layers are

1,1; 5,5; 10,10; 15,15; 20,20; 25,25; 30,30; 35,35; 40,40; 45,45; 50,50; 55,55; 60,60; 70,70; 75.75. The experimental results are presented in Table 1 and 2.

Table 1: Multi-layer Perceptron with 1 Hidden layers Architecture

Number of Units	Accuracy	Precision	Recall	F1 Score
1	**0.5161**	**0.1032**	**0.2000**	**0.1362**
10	0.7771	0.9079	0.5834	0.6639
20	0.9498	0.8926	0.9185	0.9044
30	0.9508	0.9190	0.9203	0.9182
40	0.9608	0.9419	0.9251	0.9324
50	0.9558	0.8974	0.9499	0.9196
60	**0.9689**	0.9476	**0.9511**	**0.9490**
70	0.9558	0.9317	0.9238	0.9272
80	0.9629	0.9387	0.9314	0.9344
90	0.9669	**0.9614**	0.9266	0.9421
100	0.9608	0.9469	0.9295	0.9365
110	0.9598	0.9387	0.9137	0.9249
120	0.9639	0.9574	0.9288	0.9402
130	0.9418	0.9355	0.8596	0.8846
140	0.9578	0.9388	0.9236	0.9284
150	0.9357	0.9224	0.8894	0.8970

Table 2: Multi-layer Perceptron with 2 Hidden Layers Architecture

Number of Units	Accuracy	Precision	Recall	F1 Score
1, 1	**0.5241**	**0.2441**	**0.2195**	**0.1742**
5, 5	0.7631	0.6450	0.5010	0.5396
10, 10	0.7751	0.7738	0.5778	0.6223
15, 15	0.8183	0.8512	0.6962	0.7523
20, 20	0.9217	0.8692	0.8878	0.8750
25, 25	0.9438	0.9248	0.8943	0.9088
30, 30	0.9388	0.8499	0.9072	0.8679
35, 35	0.9528	0.9282	0.9195	0.9231
40, 40	0.9498	0.9025	0.9017	0.9017
45, 45	0.9388	0.9207	0.8746	0.8955
50, 50	0.9618	0.9534	0.9251	0.9378
55, 55	0.9639	**0.9742**	0.9054	0.9350
60, 60	0.9659	0.9679	0.9220	0.9416
65, 65	0.9588	0.9371	0.9253	0.9305
70, 70	**0.9709**	0.9521	**0.9356**	**0.9430**
75, 75	0.9227	0.9132	0.8666	0.8759

Based on the experimental results the models attained reasonable scores; in terms of accuracy, the highest is 97% and the lowest score is 52%.

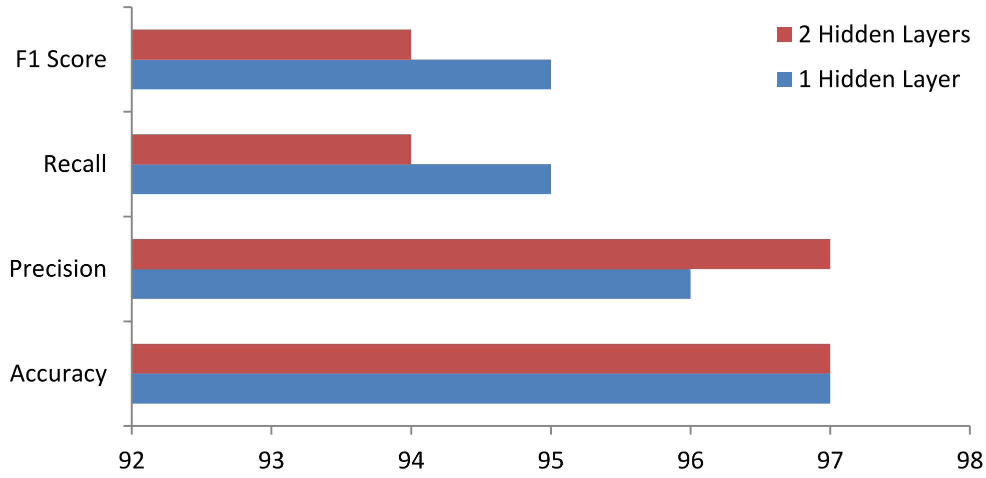

Figure 4: Comparison of Highest Classification Accuracy

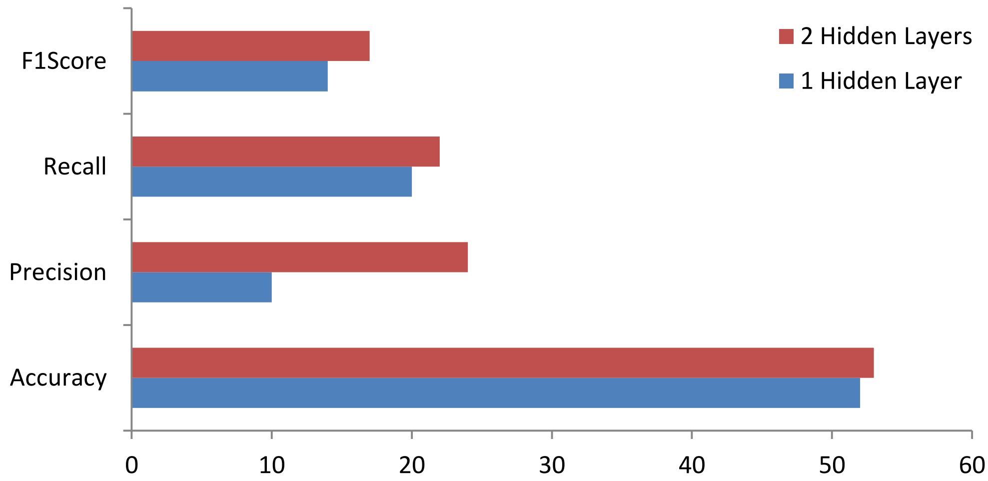

Figure 5: Comparison of Lowest Classification Accuracy

The highest score for precision, recall and F1 Score were obtained as 97%, 95% and 95 respectively; and the lowest scores for precision, recall and F1 Score are 10%, 20% and 14% respectively. Figure 4 shows the comparison of the Highest Classification accuracies obtained from the experiment architectures while Figure 5 shows the lowest attended classification accuracies from the experiment architecture.

3.2 Discussion

From the 30 experiments conducted with variations in the number of Hidden layers and number of Units in the Hidden layers, the highest classification result was obtained using 2 Hidden Layers architecture with [70,70] units in the Hidden layers; the architecture attends accuracy of 97.09%, precision of 95.21%, recall of 93.56% and F1 Score of 94.30%.

In the previous studies of [9], for the SVM-based data-driven lighting energy consumption prediction model for office buildings in Philadelphia, the SVM model attend coefficient of variation (CV) of 6.83%. The Household Electricity Consumption Prediction model in the previous studies of [11], the SVM regression model achieved a mean absolute percentage error of 9.42%.

The hybrid machine learning model that combines multi-layered perceptron, support vector regression, and CatBoost to forecast energy consumption achieved a performance of 15.72754 MAE, 472.9644 MSE, 21.74774 RMSE and 0.037885 RMSLE [12].

4. Conclusion

Based on all the experiment conducted in this study, in general, the study shows the significance of using 2-hidden layers architecture and also shows the influence of using variations in the number of units in the hidden layers to find the optimal number of units for the classification of households' cooking energy sources. The best accuracy using 1–hidden layer was obtained using (60) units in the hidden layer; and the best accuracy for the 2-hidden layer architecture was obtained using (70, 70) units in the hidden layers. This study shows that given some specific household's characteristics, the household's Major Cooking Fuel Source in Nigeria can be determined with an accuracy of 97%, precision of 95%, recall and F1 Score of 94% each.

Recommendation

In future, this study can be improved by exploring the use of different numbers of hidden layers in the MLP architecture with variations in the number of units, as well as using other machine learning algorithms.

References

[1]. Mishra, V., Retherford, R. D. and Smith, K. R., (2005). Cooking Smoke and Tobacco Smoke as Risk Factors for Stillbirth. *International Journal of Environmental Health Research,* 15(6), 397-410.

[2]. Smith, K. R. and Haigler, E., (2008). Co-Benefits of Climate Mitigation and Health Protection in Energy Systems: Scoping Methods. *Annual Review of Public Health,* 29(1), 11-25.

[3]. Kapsalyamova, Z., Mishra, R., Kerimray, A., Karymshakov, K. and Azhgaliyeva, D., (2021). Why is Energy Access not Enough for Choosing Clean Cooking Fuels? Sustainable Development Goals and Beyond. *Asian Development Bank Institute,* 1234, 1-30.

[4]. Maina, Y. B., Kyari, B. G. and Maina, M. B., (2019). Socio-Economic Determinants of Household Dirty Energy Use in Nigeria. *Journal of Agricultural Economics, Environment and Social Sciences, 5*(1 & 2), 64 - 71.

[5]. Ogwumike, F. O., Ozughalu, U. and Abiona A., (2014). Household Energy Use and Determinants: Evidence from Nigeria. *International Journal of Energy Economics and Policy*, 4(2), 248-262.

[6]. Maina, Y. B., Kyari, B. G. and Maina, M. B., (2019). Households' Clean Energy Demand in Nigeria and Its Implication on the Environment. *Journal of Agricultural Economics, Environment and Social Sciences*, 5(1&2), 35-45.

[7]. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K. (2019). Tackling Climate Change with Machine Learning, Cornell University, Ithaca, New York, 1-111. *arXiv:1906.05433v2 [cs.CY] 5.*

[8]. Nzabona, A., Tuyiragize, R., Asiimwe, J. B., Kakuba, C. and Kisaakye, P., (2021). Urban Household Energy Use: Analyzing Correlates of Charcoal and Firewood Consumption in Kampala City, Uganda. *Journal of Environmental and Public Health,* Vol. 2021 1-8. https://doi.org/10.1155/2021/5904201

[9]. Amasyali, K. and El-Gohary, N., (2016). Building lighting energy consumption prediction for supporting energy data analytics. *International Conference on Sustainable Design, Engineering and Construction*, 145, 511-517.

[10]. Madhusudanan, C. R., (2019). *A Machine Learning Framework for Energy Consumption Prediction.* All Theses, Tiger Prints. Clemson University, Clemson, South Calorina. 3184. https://tigerprints.clemson.edu/all_theses/3184.

[11]. Shen, M., Sun, H. and Lu, Y., (2017). Household Electricity Consumption Prediction Under Multiple Behavioural Intervention Strategies Using Support Vector Regression. *9th International Conference on Applied Energy, ICAE2017.* Cardiff, UK. 142, 2734-2739. DOI: 10.16/j.egypro.2017.12.218

[12]. Khan, P. W., Byun, Y., Lee, S., Kang, D. and Kang, J., (2020). Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Non-renewable Power Sources. *Energies,* 13(18) 4870, 1-16.

[13]. Koksal, M., Ugursal, V. and Fung, A., (2002). Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. *Applied Energy*, 71(2), 87-110. *10.1016/S0306-2619(01)00049-6.*

[14]. Ula, M., Muhathir, M. and Sahputra, I., (2022). Optimization of Multilayer Perceptron Hyperparameter in Classifying Pneumonia Disease Through X-Ray Images with Speeded-Up Robust Features Extraction Method. *International Journal of Advanced Computer Science and Applications,* 13(10), 203-210.

[15]. Friends of the Earth Africa. (2016). *Dirty Energy in Africa.* Friends of the Earth International, Amsterdam. www.foei.org.

[16]. Ayodele, T. O., (2010). *Types of Machine Learning Algorithms.* New Advances in Machine Learning,Yagang Zhang (Ed.), IntechOpen. DOI: 10.5772/9385.

[17]. Hastie, T., Tibshirani, R. and Friedman, J., (2008). *The Elements of Statistical Learning, Data Mining, Inference and Prediction.* Stanford, California: Springer Series in Statistics.

[18]. Bishop, M. C., (1995). *Neural Networks for Pattern Recognition.* Oxford: Oxford University Press.

[19]. National Renewable Energy Laboratory (NREL), (2001). *Renewable Energy: An Overview.* U.S. Department of Energy (DOE), Golden, Colorado. 1-8. https://www.nrel.gov/docs/fy01osti/27955.pdf