Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.sciencedirect.com/journal/energy-and-ai

Using machine learning to expound energy poverty in the global south: Understanding and predicting access to cooking with clean energy

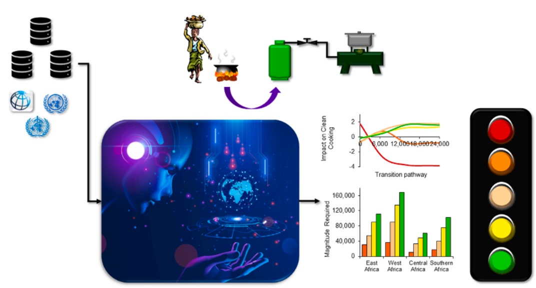
M.D. Mukelabai [a,*], K.G.U. Wijayantha [a,b], R.E. Blanchard [a]

[a] Centre for Renewable Energy Systems Technology (CREST), Wolfson School of Mechanical, Electrical, and Manufacturing Engineering, Loughborough University, Epinal Way, Loughborough, Leicestershire LE11 3TU, UK
[b] Centre for Renewable and Low-Carbon Energy, Cranfield University, College Road, Cranfield, Bedfordshire MK43 0AL, UK

HIGHLIGHTS

- Machine learning expounds energy poverty in the global south.
- Financial resilience necessary to modernize household energy systems in developing countries.
- First-of-a-kind data-driven analysis offers practical clean-cooking pathways.
- Electricity access is not linked to tackling energy poverty in developing countries.
- A rapid paradigm shift necessary to address energy poverty.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Energy modeling
Artificial intelligence
Explainable AI
Developing world
Hydrogen economy

ABSTRACT

Efforts towards achieving high access to cooking with clean energy have not been transformative due to a limited understanding of the clean-energy drivers and a lack of evidence-based clean-energy policy recommendations. This study addresses this gap by building a high-performing machine learning model to predict and understand the mechanisms driving energy poverty - specifically access to cooking with clean energy. In a first-of-a-kind, the estimated cost of US$14.5 trillion to enable universal access to cooking with clean energy encompasses all the intermediate inputs required to build self-sufficient ecosystems by creating value-addition sectors. Unlike previous studies, the data-driven clean-cooking transition pathways provide foundations for shaping policy and building energy models that can transform the complex energy and cooking landscape. Developing these pathways is necessary to increase people's financial resilience to tackle energy poverty. The findings also show the absence of a linear relationship between electricity access and clean cooking - evidencing the need for a rapid paradigm shift to address energy poverty. A new fundamental approach that focuses on improving and sustaining the financial capacity of households through a systems approach is required so that they can afford electricity or fuels for cooking.

1. Introduction

Achieving high access to clean cooking has continued to elude the global south despite worldwide attention and research activities. Even though the access rate to clean cooking increased by 12% between 2010 and 2020, about 2.4 billion people lacked access to efficient and non-

* Corresponding author.
E-mail address: m.d.mukelabai@lboro.ac.uk (M.D. Mukelabai).

https://doi.org/10.1016/j.egyai.2023.100290

Available online 22 July 2023

polluting cooking technologies [1]. The increment in access rate was only in India, Pakistan, Indonesia, China, and Brazil. The access rate to clean cooking was static for the other global south countries [1]. Previous studies show that 470 million more people will lack access to clean cooking in 2030 [2]. The global south will not attain universal access for 4.5 billion people by 2030 [3], including over 1 billion people in Sub-Saharan Africa by 2025 [4]. This slow-moving access rate to clean cooking causes deforestation, habitat loss, indoor pollution and inequality, sustains low economic activities, and causes about 3.2 million deaths per year, including over 237,000 deaths of children under the age of 5 in 2020 [1,3,5]. It is important to mention that clean cooking can mean many things, among which it can mean cooking food cleanly or cooking clean food. Clean cooking in this study implies cooking with clean energy.

Existing literature aimed at improving access to clean and affordable energy for everyone focused on increasing electricity access, techno-economic comparative analyses, energy efficiency measures, and behavior analysis [6,7,16,17,8–15]. Studies centred around the access rate to clean cooking focus on fuel stacking, population heterogeneity and affordability constraints to explore the access rate of clean cooking [2-4,18–23]. These studies use structured energy-econometric approaches that hypothesize perfect markets and predictable consumer behavior. The models assume that access to clean cooking depends on income and fuel choices with negligible interconnection to other socio-economic sectors. This fragility highlights why the efforts towards achieving high access to clean cooking have not been transformative. For example, the number of people with access to electricity in Sub-Saharan Africa is more than 2.5 times the number of people with access to clean cooking. In addition, the models lean on policy-based scenarios, such as climate-based mitigation policies [2,18,20] and low energy demand scenarios [24], thus exposed to high variability. There are few data-driven approaches in the literature. These studies focus on predicting Africa's electricity mix [25] and analysing the progress and failure of electric utilities to adapt to the energy transition [26].

Data-driven modeling approaches that can expound and predict the potential future of the clean-cooking landscape without generalizing regions are lacking in the literature. The existing studies are limited in their investigation by utilizing sales data to assess the adoption of liquefied petroleum gas cylinder (LPG) stoves for cooking [27] and surveys to model supply and demand-side factors of LPG stoves consumption [28]. Other studies employ statistical analysis techniques such as chi-square, linear, logistic and quantile regression models on preselected variables to investigate factors that affect access to clean cooking in rural India, Latin American and Caribbean countries [29–32]. There is an acute shortage of holistic studies that employ a systems approach to generate an accurate image of the cooking landscape through historical data-based learning to inform future decisions.

The lack of data-driven modeling approaches translates into a lack of evidence-based clean cooking policy recommendations. There is limited understanding of the drivers to improve access to clean cooking. Yet, it is crucial to understand the constraints and enablers of clean cooking. And an accurate representation of the cooking sector will enable decision-makers and investors to make informed decisions. Data-driven modeling approaches will add significant value to the literature by helping to propound the mechanisms affecting access to clean cooking. Thus, the key central question in this study is to identify, verify, and quantify the recommendations and steps necessary to transform the cooking landscape in the global south.

To bridge these research gaps and answer this central question, the objectives of this study were to (i) develop a data-driven approach to understand the drivers of access to clean cooking in Africa by using holistic country-level data, (ii) determine the variables with the most impact on access to clean cooking from a historical dataset and predict the potential future, (iii) establish and quantify the pathways required to attain universal access to clean cooking, and (iv) estimate the costs of achieving universal clean cooking for everyone in Africa.

To meet these objectives, Section 2 provides a detailed description of the data sources and the machine learning model. Section 3 investigates the variables with the most impact on access to clean cooking, while Section 4 details the predicted the potential future. Section 5 describes the clean cooking transition pathways before discussing the results and concluding in Section 6.

2. Materials and methods

2.1. Data sources

The data in this study was collected from world-leading databases. The country-level indicators were collected from the World Bank indicators [33]. The data on primary energy consumption was collected from Our World in Data [34] that collects its data from EIA [35] and BP [36]. The data on access rates of clean cooking was collected from the World Health Organization [37]. The data on gross value outputs, intermediate inputs, and gross value added was collected from United Nations [38]. These databases provide publicly accessible high-quality data, and country-level comparable statistics about development, combating climate change, and poverty reduction. Supplementary Table 1 gives a summary of the data.

The data points included the past three decades. Reducing the data points to the past two decades reduced the learning capability of the model. The final features in the dataset comprised 11,480 data points. These included 31 African, 1 South American (Mexico), and 6 Asian countries (India, Indonesia, Papua New Guinea, Myanmar, Bangladesh, and Philippines. Asian and South American countries were added in the data set to reinforce the learning and provide an unbiased predicting capability of the model. The countries were selected because they are transitioning to clean cooking and have all the country-level data.

2.2. Machine learning model

The machine learning model is based on a state-of-the-art open-source gradient boosting library that successfully handles categorical and numerical data and outperforms existing publicly available gradient-booting-based algorithms LightGBM, XGBoost, and H2O [39,40]. Gradient boosting on decision trees is a powerful machine learning technique that trains complex models to maximize the prediction accuracy, thus making it one of the most effective ways to build ensemble models [41,42]. The dataset containing the target and features is described as $D = \{(\mathrm{x_i}, \mathrm{y_i})\}_{i=1}^{\mathrm{n}}$, where $\mathrm{x_i} = (\mathrm{x_i^n}, \ldots \mathrm{x_i^n})$ is a vector of the features. $\mathrm{y_i} \in \mathbb{R}$ is the target (access to cooking with clean energy). The features (input) are the independent variables while the target (output) is the dependant variable.

Gradient boosting constructs an ensemble predictor by performing gradient descent in a functional space. Robust predictors are constructed through an iterative process that combines weaker models in a greedy fashion [43]. The aim of a learning task in gradient boosted decision trees is training a function $\mathrm{F}: \mathbb{R}^{\mathrm{m}} \rightarrow \mathbb{R}$ that minimizes a differentiable loss function $\mathscr{L}(\mathrm{F}) := \mathrm{L}(\mathrm{y_i}, \mathrm{F}(\mathrm{x}))$ [43]. The training involves building trees and computing the difference between the observed probability and the predicted probability (r), Eq. (1). This is the derivative of the loss function that gradient boost derives its name from. M denotes the number of trees and n denotes the samples in the dataset.

$$\mathrm{r_{im}} = -\left[\frac{\partial \mathrm{L}(\mathrm{y_i}, \mathrm{F}(\mathrm{x_i}))}{\partial \mathrm{F}(\mathrm{x_i})}\right]_{\mathrm{F(x)} = \mathrm{F}_{m-1}(\mathrm{x})} \quad \text{for } i = 1 \ldots \mathrm{n} \tag{1}$$

The model was built in python using Catboost (categorical boosting), an open source algorithm developed by Yandex and freely available on GitHub [44]. CatBoost is a modification of the standard algorithm based on gradient boosting by applying ordered boosting [40]. The algorithm prevents target leakage (a type of overfitting in gradient-boosted algorithms) and implements a new and efficient algorithm for processing categorical features. The algorithm handles categorical features during training unlike during pre-processing thus enabling the usage of the entire dataset for training [39]. These advantages were important in the analysis to obtain unbiased results by reducing the likelihood of over-fitting. The model attained excellent predictive performance. See Section 2.7 and Supplementary Table 2.

Obtaining reproducibility and high predictive performance of the machine learning model required substantial data preprocessing (feature engineering), determining important features, optimizing the hyperparameters, evaluating the model performance, validating the model, and interpretating the black-box model. These processes are outlined in the following subsections.

2.3. Feature engineering

The numerical features were converted to categorical data. This conversion was done through normalization, multiplying by 120, and rounding the features to integers, Eq. (2). A CatBoost classifier recognises integers as categories while floats are recognised as continuous variables. By doing this, the numerical country-level indicators were converted to categorical indicators during training and converted back to numerical features during result analysis and presentation.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \times 120 \tag{2}$$

This simple yet powerful feature engineering technique increased the prediction power and reduced convergence time by a magnitude. The data was split into a training set comprising 8120 data points and a test set comprising 3360 data points. Each data point comprises a target y_i (access to clean cooking), and features x_i (mechanisms driving access to clean cooking), from $i = 1 - 8,120$ in the training set and $i = 1 - 3,360$ in the testing set.

2.4. Model interpretation

Interpreting black-box models is crucial in machine learning. SHAP (Shapley Additive Explanations) was applied to interpret the model. SHAP is based on game theory introduced by Lloyd Shapley [45] and is a unified approach that explains the output of any machine learning model [46,47]. This state-of-art framework introduced by Lundberg and Lee in 2017 [46] has been successfully applied in literature to explain the results of black box models [48–52]. The SHAP values explained the impact of features on predicting access to clean cooking. The importance of a feature is computed by evaluating the model's output with and without this feature. The SHAP algorithm derived for tree ensembles lowers the complexity of calculating the precise SHAP values from $O(TL2^2)$ to $O(TLD^2)$ where T is the number of trees, L is the maximum number of leaves in any tree, M is the number of features, and D is the maximum depth of any tree [53]. This exponential reduction in model complexity enables explainability and prediction from previously hard-to-deal with models with thousands of trees and features in a fraction of a second. The SHAP values explained the impact of features (mechanisms driving access to clean cooking) on predicting access to clean cooking. The importance of a feature was computed by evaluating the model's output with the feature, $f_x(S)$, and without this feature, $f_x(S \cup \{i\})$, Eq. (3). The difference gives the contribution of the feature to the subset. $\varnothing$ is the Shapley value, F is the number of input features and N is the set of all input features. S is the set of non-zero feature indices (the features observed and unknown). $f_x(S) = E[f(x)|x_s]$ is the model prediction for an input x, where $E[f(x)|x_s]$ is the expected value of the function conditioned on a subset S of the input features.

$$\varnothing_i(f, x) = \sum_{S \subseteq N\{i\}} \frac{|S|!(F - |S| - 1)!}{F!} [f_x(S \cup \{i\}) - f_x(S)] \tag{3}$$

The Shapley value is the only method that satisfies the desirable properties efficiency, symmetry, dummy, and additivity resulting in a fair distribution. The efficiency property assures that feature contributions sum up to the difference of prediction and the average. Symmetry assures that the contributions of two feature values must be the same if they contribute equally to all possible groups of features. Dummy assures that features that do not change the predicted value in any group of features have a Shapley value of 0. Additivity assures that the Shapley values for a feature can be computed individually for each tree and averaged.

To explain SHAP simply, shapley values can be thought of in terms of a game. If access to clean cooking (the target) is winning a game, and the mechanisms (the features) are the players. Then shapley values explain the average contribution of a player to winning a game. Each player contributes differently to winning a game and interacts differently with other players. Thus, SHAP compares the performance of the team with and without a specific player. This computation gives the marginal contribution (marginal value) of a player to a team. In the machine learning model, the features are treated like a player to compute their contribution to predicting access to clean cooking.

The global feature importance was computed by averaging the absolute Shapley values per feature across the data, Eq. (4).

$$\frac{1}{F} \sum_{i=1}^{F} |\varnothing_i|, \ \varnothing_i \in \mathbb{R} \tag{4}$$

2.5. Feature selection

Feature selection is another crucial process when building machine learning models. Selecting the right combination of features significantly improves the model performance by capturing critical spatial effects. SHAP was applied [46] analysis in the model to select the features that have a high marginal contribution to predicting access to clean cooking. After identifying the top 20 features, dropping low-ranking features during the analysis resulted in a different order of feature importance. For example, a feature ranked number 19 can rank as number 16 after dropping the 20th feature. Therefore, the process was iterated to capture the relative importance of features in the absence of a feature and to rigorously select features. All the features that had a low marginal contribution on predicting access to clean cooking were dropped to remain with 9 features out of 34 features. The important features are (in order of importance) energy consumption, households and NPISHs final consumption expenditure, female literacy, services value addition, electricity access, industry value addition, GDP per capita, agric-forestry-fishing value addition, and fertilizer consumption.

The features that have a low marginal contribution on predicting access to clean cooking included agriculture land, forest area, Gini coefficient, poverty head count ratio, corruption, urban population, rural population, total unemployment rate, female unemployment rate, male waged salaried workers, female waged salaried workers, female employment in agriculture, male employment in agriculture, adult literacy rate, cereal yield, crop production index, food production index, energy investment by the private sector, expense per GDP, energy imports, medium and high-technology exports, renewable electricity consumption, fossil fuel consumption, total population, and adjusted net national income per capita. Supplementary Table 1 gives a summary of the model data.

2.6. Hyperparameter optimization

Hyperparameter optimization is a critical and delicate aspect when building machine learning models to enhance model performance. Hyperparameters are parameters not determined by a model but control the learning process. Thus, the optimization process determined the optimal mix of hyperparameters that maximized the model performance.

Optuna was applied in this model, an open-source, state-of-the-art, and first-of-its-kind optimization algorithm. Optuna built the hyperparameter optimization by maximizing the objective function that took a set of hyperparameters as an input and returned its score to evaluate the performance of hyperparameters [54]. The MultiClass objective function was applied in the model, Eq. (5). w_i is the weight of the ith sample. t_i is the label value for the i th sample from the input data for training and a_i is the result of applying the model to the ith sample. C is the number of classes, n is the total number of samples.

$$\frac{\sum_{i=1}^{n} w_i \log\left(\frac{e^{a_{it_i}}}{\sum_{j=0}^{C-1} e^{a_{ij}}}\right)}{\sum_{i=1}^{n} w_i},\ t \in \{0, \cdots, C\} \tag{5}$$

The evaluation metric applied in the model is the area under the curve (AUC), Eq. (6). The AUC is the ability of a classifier to differentiate between classes. AUC(j|k) is the AUC with class j as the positive class and class k as the negative class. The objective function builds the searching space of neural network architecture without depending on variables specified externally. The algorithm significantly outperforms existing optimization frameworks through a versatile and efficient sampling and pruning algorithm. The algorithm facilitated the construction of the search space dynamically. See Supplementary Table 3 for all the optimized hyperparameters.

$$\frac{1}{C(C-1)} \sum_{j=1}^{C} \sum_{k>1}^{C} (AUC(j|k) + AUC(k|j)),\ AUC(j|k) \neq AUC(k|j) \tag{6}$$

2.7. Model performance and validation

After hyperparameter optimization and training, the sci-kit-learn evaluation metrics were implemented [55] to assess the model performance. The ability of the model to correctly predict classes was determined by computing the model accuracy, Eq. (7).

$$\frac{\sum_{i=1}^{n} w_i \left[\text{argmax}_{j=0,\ldots,C-1}(a_{ij}) == t_i\right]}{\sum_{i=1}^{n} w_i},\ t \in \{0, \ldots, C-1\} \tag{7}$$

The fraction of positive predictions was determined by computing the precision of the model, Eq. (8). The precision of the model is computed independently for each class numbered from 0 to C − 1. TP denotes positive and FP denotes false positive.

$$\frac{TP}{TP + FP} \tag{8}$$

The fraction of positive classes correctly predicted as positive was determined by computing recall, Eq. (9). Recall is computed independently for each class numbered from 0 to C − 1. FN denotes false negative.

$$\frac{TP}{TP + FN} \tag{9}$$

The weighted harmonic mean of the precision and recall F1 score was determined by computing the F1 score (F-measure), Eq. (10). F1 score is computed independently for each class numbered from 0 to C − 1.

$$2 \times \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

The fraction of wrongly predicted classes was determined by computing the zero one loss, Eq. (11).

$$1 - \frac{\sum_{i=1}^{n} w_i \left[\text{argmax}_{j=0,\ldots,C-1}(a_{ij}) == t_i\right]}{\sum_{i=1}^{n} w_i},\ t \in \{0, \ldots, C-1\} \tag{11}$$

Like the zero one loss, the hamming loss is the fraction of the imperfectly predicted classes to the total classes, but it is computed as the hamming distance between two sets of samples and penalises the individual classes, Eq. (12).

$$\frac{\sum_{i=1}^{n} w_i \left[\text{argmax}_{j=0,\ldots,C-1}(a_{ij}) = t_i\right]}{\sum_{i=1}^{n} w_i} \tag{12}$$

The trained model was validated with the testing (validation) dataset to ensure the reliability and reproducibility of the model. The empirically acceptable fraction of validation datasets is 20 - 30% to ensure robustness of validation results. Thus, 29.2% of the dataset was used for validating the model (see Section 2.3). The stratified k-fold cross-validation iteration technique in sci-kit-learn was implemented to validate the model [55]. The stratified k-fold cross-validation is an extension of the conventional k-fold cross-validation. Using the stratified 5-fold cross-validation technique guaranteed that the size of the target features is equal within the training and validation data sets. This approach reduced overfitting, thus creating a robust and high-performing model.

The ability of the model to correctly predict classes (accuracy) on the trained and test datasets is 93% and 96%, respectively. The fraction of the imperfectly predicted classes (hamming loss) on the trained and test dataset is 6.65% and 2.58%, respectively. The fraction of positive predictions (precision) on the trained and test dataset is 94% and 95%, respectively. The fraction of positive classes correctly predicted as positive (recall) on the trained and test datasets is 93% and 96%, respectively. See the performance metrics in Supplementary Table 2. The performance of the model was compared with XGBoost [56] (a gradient-boosted algorithm) to ascertain the robustness of the results. The Catboost classifier outperformed the gradient-boosting classifier by preventing overfitting through regularization. Even though both models perform well on the trained and test data sets, the XGBoost gradient-boosting classifier achieved very high metrics - denoting low bias and high variance in the model. Machine learning models with low bias and high variance are ineffective when dealing with real-world data because of their inability to resist learning the noise stemming from determinate-sized data sets [57].

2.8. Clean-cooking pathways and future access rates of clean cooking

Furthermore, SHAP dependence plot analysis was applied to establish evidenced-based clean-cooking scenarios. The feature values were plotted on the x-axis and the corresponding Shapley value on the y-axis, Eq. (13). ∅ is the marginal contribution of a feature *i* and *x* is a feature from *i* to *F* (the number of input features). The effects of the features by varying the values were also explored. This analysis was achieved by creating varied synthetic data sets. This analysis revealed the model behavior and enabled the postulation of data-driven clean-cooking pathways.

$$\{(x_i, \varnothing_i)\}_{i=1}^{F} \tag{13}$$

The future access rates of clean cooking were determined by comparatively evaluating two assumptions. The assumptions are (1) the past three decades influence the next three decades, and (2) the past two decades influence the next three decades. The future country-level indicators were estimated through linear extrapolation based on these assumptions and ensured the robustness of the results by comparing the results internally and externally. For example, the percent difference between the South African GDP forecast is only 1.64% between the first assumption and the forecast by OECD [58]. The percentage difference between the second assumption and the OECD estimate is 15.8%. The second assumption was applied in the model because it captures the combined effects of the UN millennium development goals and the UN 2030 Sustainable Development Goals.

2.9. Energy balance and cost analysis

The data on primary energy consumption was extracted from Our World in Data - which gets most of its data from BP. Our World in Data and BP use the substitution method to calculate the share of energy consumption by source [59]. This approach accounts for energy losses during conversion processes. For example, about two-thirds of fossil fuel energy is lost through thermal losses, while renewable energy has negligible transformation losses. Thus, the total primary energy is not the final energy consumption because it includes fossil fuels. Primary energy also includes renewable and nuclear energy. This complex fossil fuel energy loss was compensated in the analysis by multiplying primary energy consumption with a correction factor of 0.5 derived from Eq. (14). This correction factor accounts for the primary energy consumed (PEC) and the average share of fossil fuels in the primary energy (SFF) from 1990 to 2020, and a standard thermal efficiency factor of 0.38 (η_t). The average share of fossil fuels from the EIA [35] is 82%, see Supplementary Fig. 11. The correction factor allows the quantity of renewable electricity generated to equal the final energy consumed.

$$(\mathrm{PEC} - \%\ \mathrm{SFF} * \mathrm{PEC}) + \%\mathrm{SFF} * \mathrm{PEC} * \eta_t \tag{14}$$

By applying this conversion in the energy balance analysis, the postulated total energy required equals the renewable electricity generation and the planned electricity generation from existing studies. The sectoral enery consumption was estimated by considering the postulated contribution of each sector to the total required energy. The residential energy consumption was estimated by accounting for the households without access to clean cooking and the growing population, Eq. (15).

$$\mathrm{RREC} = \left[\frac{\mathrm{CP} - \mathrm{PWACC}}{\mathrm{H}} + \frac{\mathrm{FP} - \mathrm{CP}}{\mathrm{H}}\right] * \sum(\varphi_{\mathrm{C}} + \varphi_{\mathrm{HW}} + \varphi_{\mathrm{L}} + \varphi_{\mathrm{RC}} + \varphi_{\mathrm{O}}) \tag{15}$$

Where RREC is the required residential energy consumption and CP is the current population. PWACC is the population without access to clean cooking. The PWACC also denotes the population in energy poverty. H is the average number of households [60] and FP is the forecasted population. C, HW, L, RC, and O are the average cooking, hot water, lighting, refrigeration and cooling, and other household energy needs respectively. The required renewable energy capacity accounts for the predicted installed generation capacity in 2030 and required energy consumption for all the sectors, Eq. (16). RIGC is the required installed generation capacity, ER is the energy required, and PGC is the predicted generation capacity. CP is the capacity factor of the respective generation technology, GT.

$$\mathrm{RIGC} = \frac{\mathrm{ER} - \mathrm{PGC}}{\mathrm{CP}_{\mathrm{GT}}} \tag{16}$$

Extending the SHAP analysis described in Section 2.8 that informed the clean-cooking transition pathways enabled us to estimate the cost required to transition to clean-cooking ecosystems. These estimates encompass all the intermediate inputs required to build self-sufficient clean-cooking communities by creating value-addition sectors - thereby creating jobs and eliminating poverty. The gross value added to gross value output ratios of the agric-forestry-fishing, industry (including construction), and services value addition from the UN [38] were applied onto the transition pathways to estimate the costs required to transition to clean-cooking ecosystems. These ratios were applied in the cost analysis because they are steady from 1990 to 2020 - implying the ratios will remain unchanged during the next decades. The different regional material efficiencies were captured by capturing regional effects. See the ratios in Supplementary Figs. 7–10. Finally, energy efficiency improvements were included in the energy balance analysis to account for technological learning and energy efficiency measures. GO and GVA in Eq. (17) are the gross output and gross value added, respectively. The industry value addition includes value added in manufacturing, construction, mining, electricity, water, and gas. The agric-forestry-fishing value addition comprises forestry, hunting and fishing, including crop and livestock production. The services value addition constitutes wholesale and retail, transport, government, financial, professional, and personal services such as education, health care, and real estate. Fig. 1 shows a flowchart of the machine learning model and analysis.

$$\mathrm{Totalcosts} = \sum_{\mathrm{industry}} (\mathrm{GO} - \mathrm{GVA}) + \sum_{\mathrm{services}} (\mathrm{GO} - \mathrm{GVA}) + \sum_{\mathrm{agric-forst-fishing}} (\mathrm{GO} - \mathrm{GVA}) \tag{17}$$

3. Impact of country-level features on access to clean cooking

The findings indicate that primary energy consumption, households and NPISH (non-profit institutions serving households) final expenditure, and female literacy have more impact on predicting access to clean cooking (Fig. 2). GDP per capita, agric-forestry-fishing value addition, and fertilizer consumption have the lowest impact. On the other hand, the results reveal that female literacy has a lower impact on countries that have transitioned to class 4 (80 - 100%) - denoting that it is not a crucial driver to sustain access to clean cooking in countries with more than 80% access to clean cooking. Apart from the agric-forestry-fishing value addition with lower impact, all the features have more impact for countries to transition from class 0 (0 - 19%) to class 1 (20 - 39%). Beside energy consumption and household expenditure, the features share comparable importance to transition to class 2 (40 - 59%) from class 1. Interestingly, services value addition and agric-forestry-fishing value addition have more impact on class 4. This finding underscores that income-generating activities are crucial to sustaining high access rates of clean cooking in a country.

Contrasting feature importance were noticed when the analysis was applied to the African regions to understand how the features drive access to clean cooking regionally (Fig. 3). The features have more impact in North Africa than in the rest. This region has successfully transitioned to class 4 of access to clean cooking apart from Libya. For example, the overall impact of primary energy consumption in North Africa is four times more than in other regions (Fig. 3a). In addition to services and agric-forestry-fishing value addition, the findings reveal that industry value addition, fertilizer consumption, and household expenditure have integral roles in sustaining access to clean cooking above 80% - further highlighting the importance of income-generating activities.

Even though North Africa has high levels of access to clean cooking, the importance of electricity access is lower in classes 1, 2 and 4 (Fig. 3b). This finding indicates the significance of using alternative clean cooking technologies for societies to transition to class 4 and evidences the role that gases such as green hydrogen can play in improving the levels of clean cooking while solving the hydrogen chicken-egg [61] problem. For example, over 75% of the households in Egypt use liquefied petroleum gas cylinder stoves for cooking because they do not require grid connections and are cheaper than using electricity for cooking [62]. Income-generating activities such as industrialization have more impact in West Africa - indicating the importance of industrial activities in the region in improving access to clean cooking (Fig. 3f). Fertilizer consumption, agric-forestry-fishing value addition, and services value addition have more impact in Southern Africa, whereas Central Africa shows the lowest impact.

The results indicated the existence of various relationships between the features and their effect on predicting access to clean cooking (Fig. 4 and Supplementary Fig. 1 to Fig. 4.). The impact of primary energy consumption on predicting class 4 depicted the existence of a logarithmic progression. This finding denotes that continued growth in energy consumption for a country that has transitioned to class 4 (80 - 100%) has a static effect on access to clean cooking. Electricity access depicts the existence of an exponential relationship with access to clean cooking - which reveals that high electricity rates are required to transition to class 4 access to clean cooking.

On the other hand, household expenditure indicates a linear relationship. Female literacy shows that high rates have a negligible impact

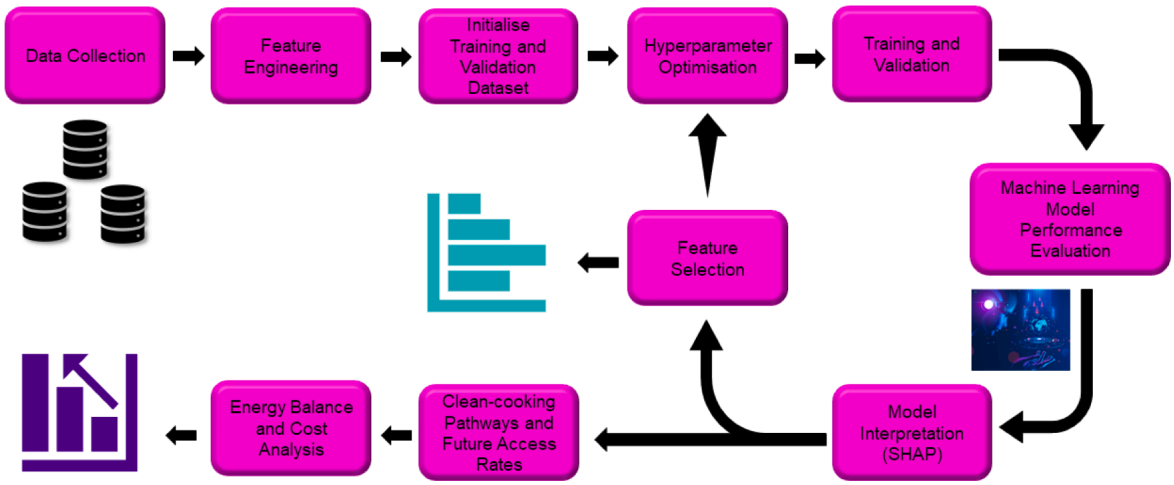


Fig. 1. Flowchart for the machine learning model and analysis.

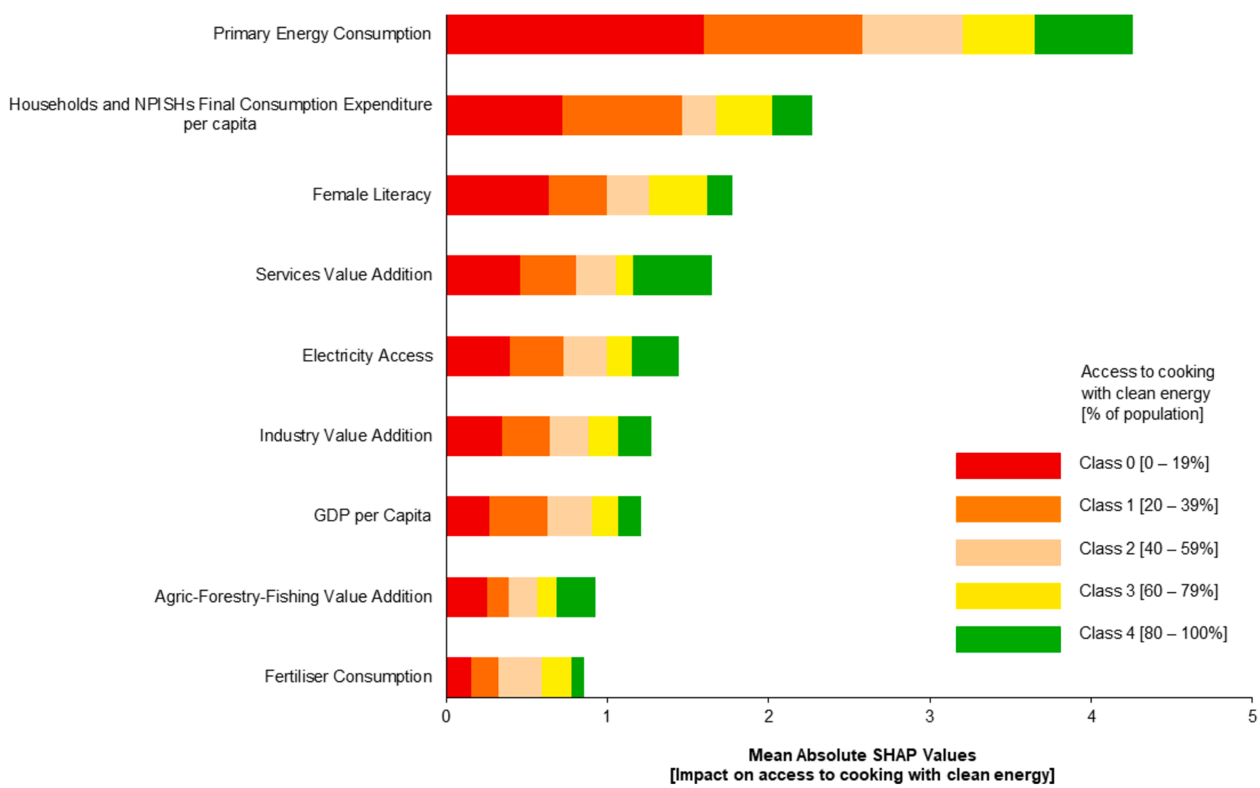


Fig. 2. The impact of country-level features on predicting the different classes of access to clean cooking. (Class 0: 0 - 19% population with access to clean cooking, class 1: 20 - 39%, class 2: 40 - 59%, class 3: 60 - 79%, class 4: 80 - 100%). For example, Algeria is a class 4, while Kenya is a class 0 (See Fig. 4). Higher values of SHAP values in a class indicate a higher average impact. SHAP (Shapley Additive Explanations) values explain the marginal contribution of features on predicting access to clean cooking. The marginal contribution of a feature, in mean absolute SHAP values, is computed by evaluating the model's output with and without this feature. See Methods 2.4. Primary energy consumption is the most important feature with an overall impact in all the classes higher by 47% to the second most important feature, households and NPISHs (non-profit institutions serving households) final expenditure per capita.

on sustaining access to clean cooking in countries with over 80% of access to clean cooking. Agric-forestry-fishing value, industry, and services value addition, as well as fertilizer consumption and GDP, appear to plateau, showing that these features have boundaries beyond which more growth has less impact on access to clean cooking.

4. Predicted access to clean cooking

There are currently about 940 million people in Africa without access to clean cooking [63,64] (Fig. 5a). The model predicts that the number of people with access to clean cooking in Africa will increase by 69 million in 2030 and 121 million in 2050 compared to 2020 (without the countries with missing data, Fig. 5a, b, and c). The average rate of accessing clean cooking in 2030 and 2050 is thus about 8% and 11%, respectively. The estimated access rates to clean cooking are in line with existing studies [3,65].

The findings show that these access rates are too low to outpace population growth - implying that over 840 million people in Africa will have no access to clean cooking in 2030. And over 1.1 billion people in Africa will have no access to clean cooking in 2050. Instead of reducing

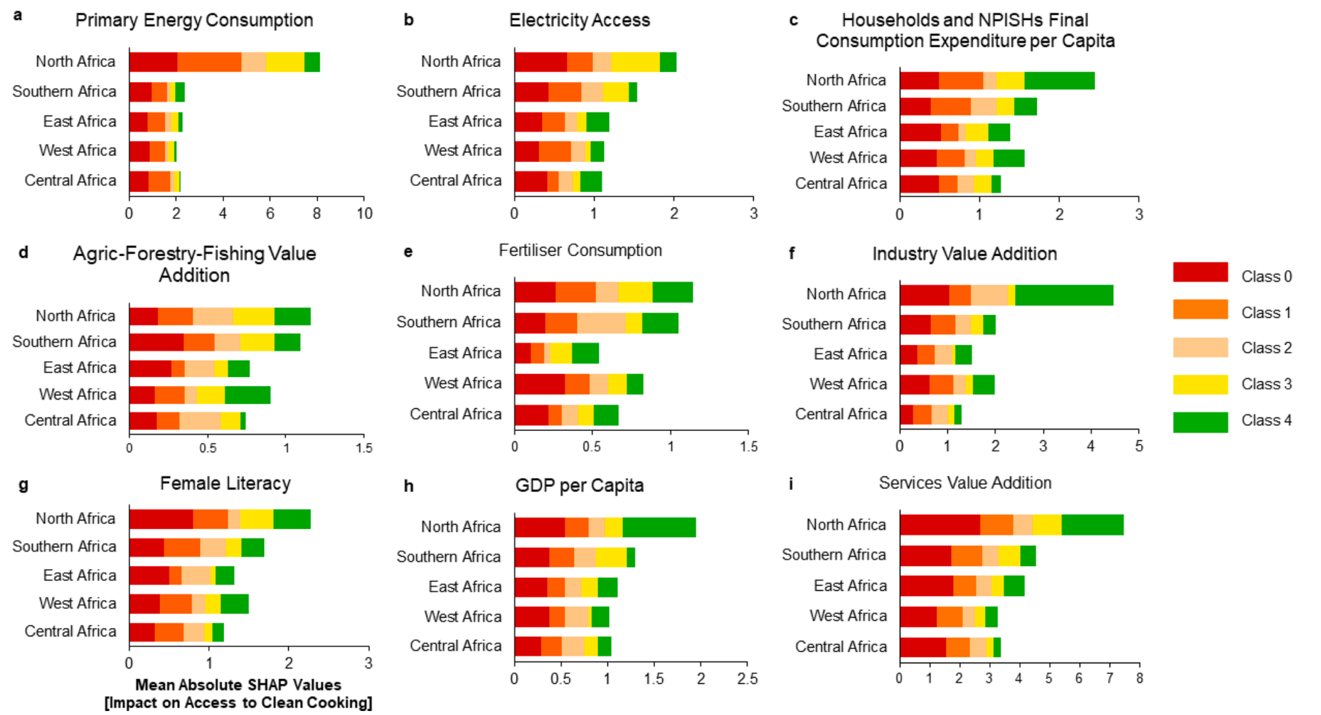


Fig. 3. The impact of country-level features on predicting regional access to clean cooking. Higher values of features in a class indicate a higher average impact. See Methods 2.4. (Class 0: 0 - 19% access to clean cooking, class 1: 20 - 39%, class 2: 40 - 59%, class 3: 60 - 79%, class 4: 80 - 100%). (NPISHs - non-profit institutions serving households).

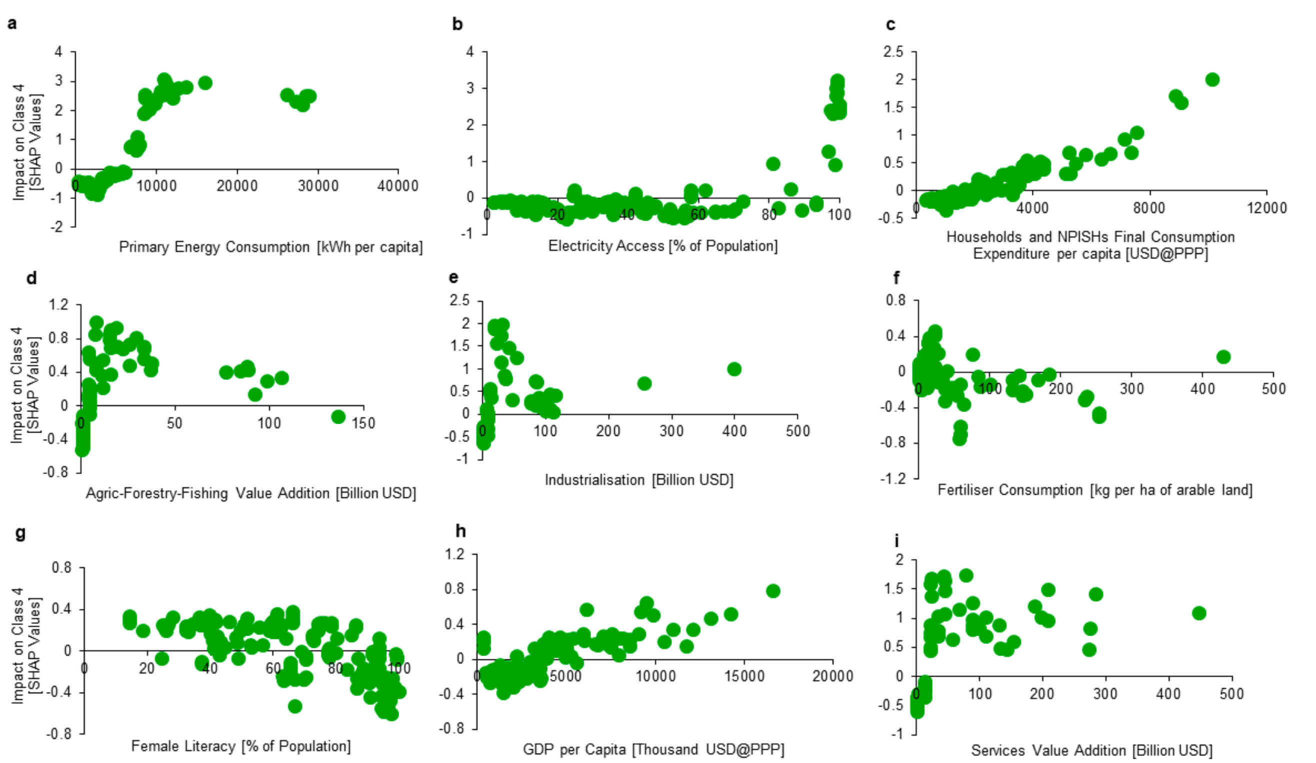


Fig. 4. The Impact of country-level features on predicting class 4 (80 - 100% access to clean cooking). a-i, The plots show the type of relationship between the features and class 4. See Methods 2.4. For example, lower values of primary energy consumption and electricity access reduce the probability of transitioning to class 4 of access to clean cooking. (a) The impact of primary energy consumption, (b) electricity access, (c) households and NPISHs (non-profit institutions serving households) per capita converted by power purchasing parity conversion factor (PPP), (d) agric-forestry-fishing value addition (includes forestry, hunting and fishing, as well as cultivation of crops and livestock production), (e) industry value addition (it comprises value added in manufacturing, construction, mining, electricity, water, and gas), (f) fertilizer consumption (nitrogenous, potash, and phosphate fertilizers), (g) female literacy, (h) GDP per capita converted by power purchasing parity, and (i) services value addition (includes wholesale and retail, transport, government, financial, professional, and personal services such as education, health care, and real estate services).

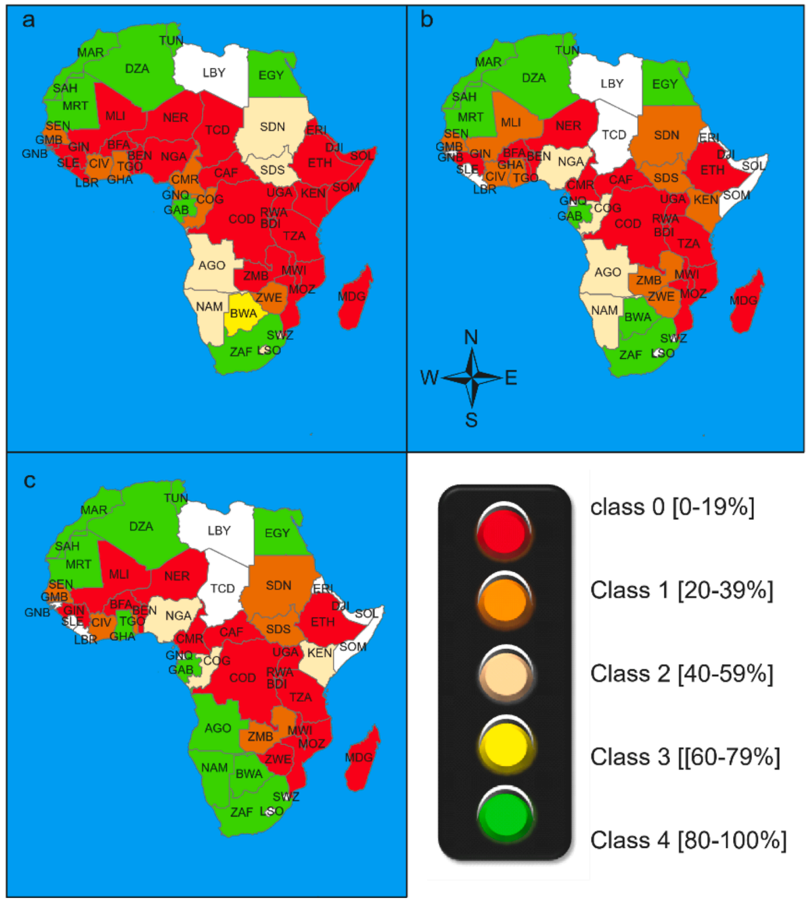


Fig. 5. The predicted access rate of clean cooking - A clean-cooking traffic light system. (a) Access rate to clean cooking in 2020, (b) access to clean cooking in 2030, (c) access to clean cooking in 2050. Libya, Chad, Somalia, Sierra Leone, Guinea Bissau, Liberia, Equatorial Guinea, Somali Land, Eritrea, Djibouti, Lesotho, and Swaziland were dropped from the analysis due to missing data. Supplementary Table 4 gives the country names and abbreviations.

from 940 million, an additional 117 million people in Africa may not have access to clean cooking in 2050. This stark image of the clean-cooking landscape indicates the need to postulate data-driven clean-cooking pathways with an achievable potential.

5. Clean cooking transition pathways

Evidenced-based clean-cooking scenarios were established by varying the feature values (Fig. 5). The results revealed that primary energy consumption impacts transitioning to class 1 (20 - 39%) between 3000 and 9,000 kWh per capita per year (Fig. 5a). The maximum effect is around 6,000 kWh per capita per year. For countries to transition to classes 2 (40 - 59%), 3 (60 - 79%), and 4 (80 - 100%), energy consumption is impactful above 2,000 kWh per capita per year. The high primary energy consumption required to transition to class 1 signifies the importance of energy in a country to spur economic activities. The maximum impact in classes 2, 3, and 4 is around 17,000 kWh per capita per year before plateauing. The required final energy consumption per capita per year is about two times lower than the primary energy consumption after applying a correction factor of 0.5, see methods 2.9.

The electricity access rate impacts transitioning to class 1 after about 40% of electricity access (Fig. 5b). Meanwhile, electricity access impacts transitioning to class 2 up to about 45% of electricity access, whereas it impacts transitioning to class 3 after about 50% of electricity access. Access to electricity impacts transitioning to class 4 after about 85% of electricity access. These findings critically highlight the absence of a linear relationship between electricity access and access to clean cooking in the global south - implying that high rates of electricity access do not translate into high rates of access to clean cooking due to high electricity tariffs and load-shedding [66]. These significant findings underscore the need for a rapid paradigm shift in the developing world energy sciences. The evidence necessitates a fundamental change in approach to tackling energy poverty. Increasing the financial resilience of households through a systems approach should be the focal point because a house with financial capacity can afford electricity or fuels for cooking.

The household expenditure impacts transitioning to class 1 between 1500 and 7000 US$ per capita per year, with the maximum impact of around 3000 US$ per capita per year (Fig. 5c). Transitioning to over 40% of access to clean cooking requires substantial household expenditure between 4000 and 5000 US$ per capita per year. However, the impact stalls at around 7000 US$ per capita per year in class 3 and shows an exponential effect in class 4. The household expenditure of 7000 US$ per capita-year translates to a daily expenditure of US$19.18 per day for countries to transition to clean-cooking ecosystems.

The income-generating activities in classes 2, 3, and 4, namely, agric-forestry-fishing, industry, and services value addition, including GDP, show plateauing characteristics after about US$200 billion per year and 17,000 US$ per capita per year, respectively. This plateauing indicates boundaries beyond which more growth has less impact on sustaining clean cooking in an ecosystem. The female literacy rate impacts transitioning to classes 1, 2, and 3 after about 70%, 85%, and 80% of the female literacy rate, respectively. On the other hand, it impacts transitioning to class 4 up to about 80% of the female literacy rate. Fertilizer consumption impacts transitioning to class 4 between 300 and 400 kg per hectare per year (Fig. 6).

Striking differences between the regions and countries were observed when the pathways were applied to determine the magnitude

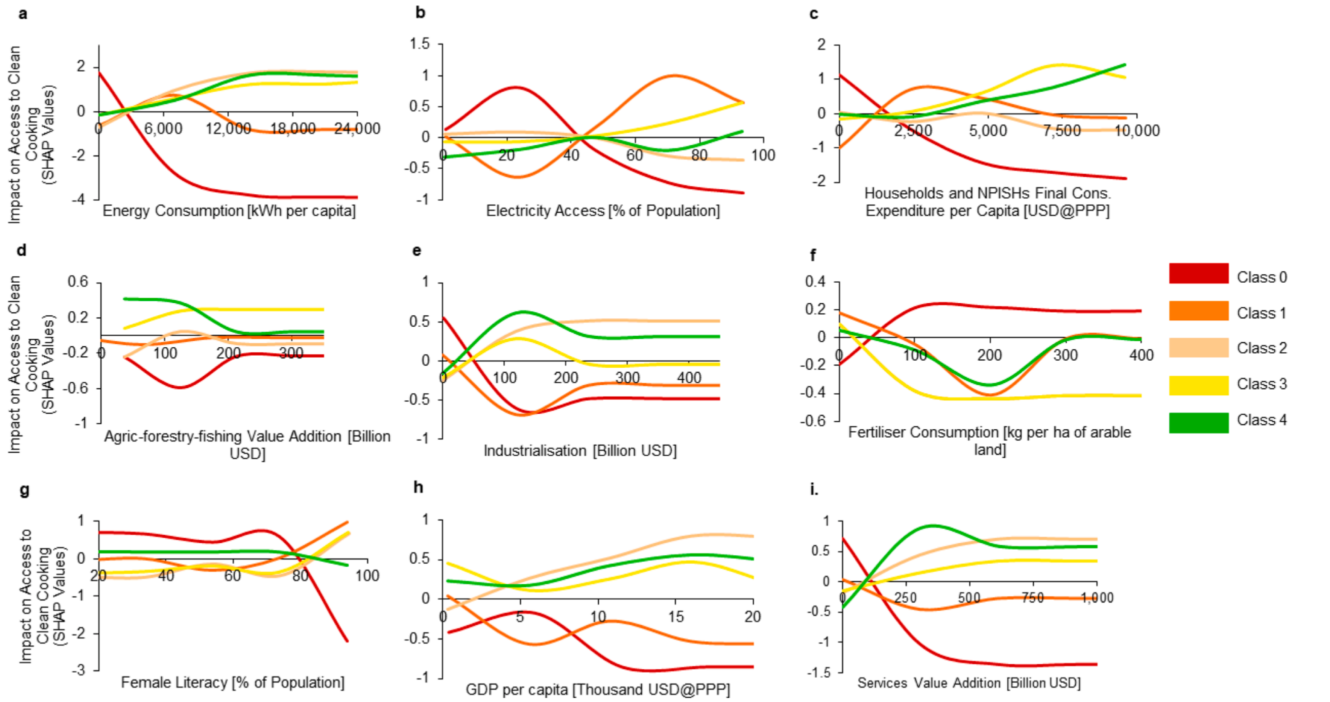


Fig. 6. Evidenced-based clean-cooking scenarios. See Methods 2.4 and 2.8. (Class 0: 0 - 19% access to clean cooking, class 1: 20 - 39%, class 2: 40 - 59%, class 3: 60 - 79%, class 4: 80 - 100%). (PPP - purchasing power parity conversion factor. NPISHs - non-profit institutions serving households).

of increment required for countries to transition to classes 1, 2, 3, and 4 based on their current states (Fig. 7 and Supplementary Fig. 5). The primary energy consumption in West Africa should increase by 168 MWh per capita per year to transition to class 4. The energy consumption in Southern and East Africa should increase by 103 and 111 MWh per capita per year, respectively. Central Africa should increase its consumption by 61 MWh per capita per year. The required final energy consumption per capita per year is two times lower than the primary energy consumption after applying a correction factor of 0.5, see methods 2.9.

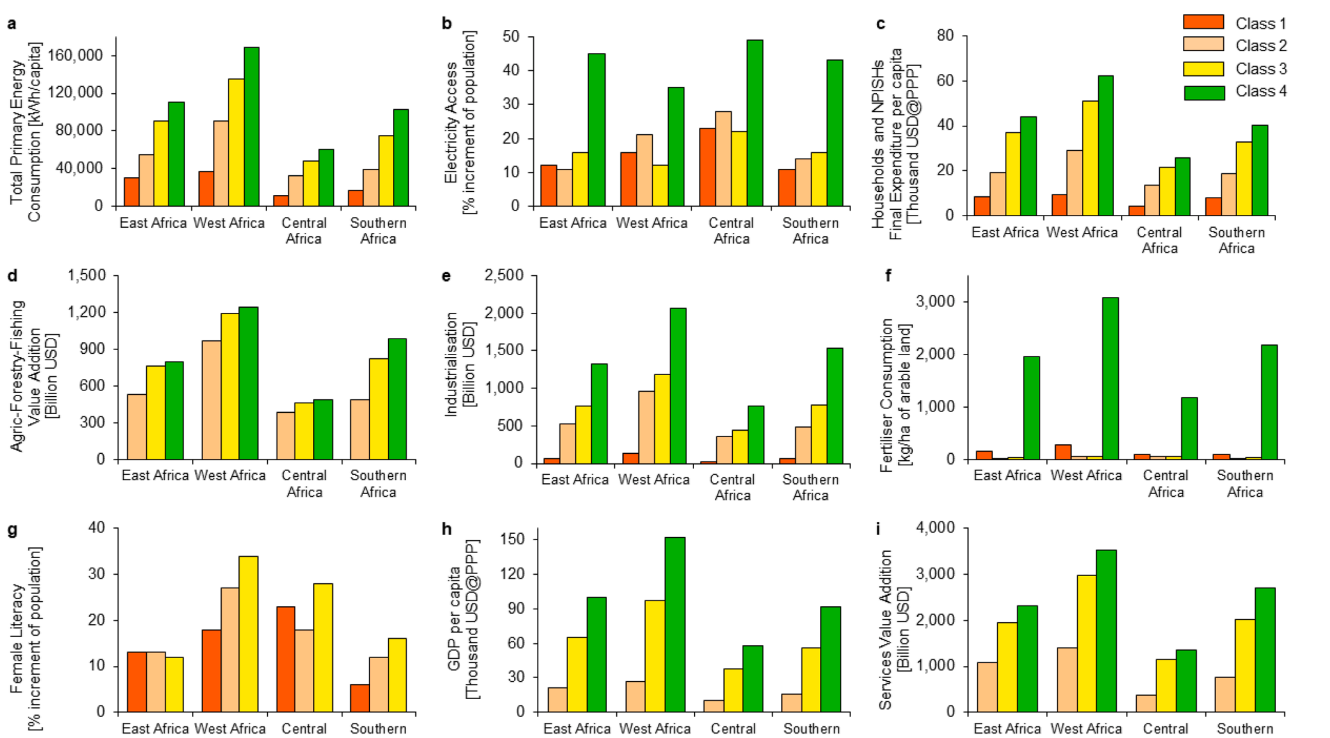


Fig. 7. Data-driven postulated clean-cooking transition pathways. (class 1: 20 - 39% access to clean cooking, class 2: 40 - 59%, class 3: 60 - 79%, class 4: 80 - 100%). Class 0 is omitted because it is the lowest class to transition from (PPP - purchasing power parity conversion factor. NPISHs - non-profit institutions serving households).

The magnitudes of increment required for all the features apart from electricity access are higher in West Africa. Southern and East Africa have comparable required magnitudes, while Central Africa has the lowest. Female literacy has a negligible impact on transitioning to class 4. Thus, its value is 0 in this class. The increment in the value of industrialization and fertilizer consumption required to transition to class 1 totals US$282 billion per year and 621 kg per hectare per year of fertilizer consumption, respectively. The estimated cost of achieving universal access to clean cooking (Supplementary Fig. 6) is higher in West Africa at US$5.2 trillion per year, followed by Southern Africa at US$4.7 trillion per year, East Africa at US$2.9 trillion per year, and Central Africa at US1.7 trillion per year.

An energy balance analysis was conducted to understand and visualize the magnitude of energy flows required per sector (Fig. 8). The results reveal that the required increment in residential energy consumption is 29% of the total energy needed for countries to transition to clean cooking. This finding signifies the importance of energy consumption in spurring wealth-generating activities that will sustain a clean-cooking ecosystem. The rejected energy is a third of the required final energy consumption - underlying the importance of energy efficiency measures at the end user to optimise generation capacity and raw materials when developing the energy economy.

Industrial activities consume more energy, followed by commercial and public buildings, residential buildings, agriculture and transportation. Producing fuels from renewable electricity to meet non-electrical energy consumption and energy storage results in slightly higher installed capacities to meet the power-to-fuel conversion losses. For example, producing hydrogen using conventional electrolysers requires increasing the generation capacity by about a third whereas using E-TAC electrolysers (Electrochemical-thermally Activated Chemical) with a system efficiency of 95% [67] requires an increment in the generation by only five hundredths. Moreover, the total final energy required in Sub-Saharan Africa is a staggering 183 times lower than the combined wind and solar PV potential. The combined renewable energy potential (wind, solar PV, and concentrated solar power) is 763,823 TWh per year in Africa [68]. Even so, the planned electricity generation in Sub-Saharan Africa should increase three times to meet the required final energy consumption.

6. Discussion and conclusions

The drivers necessary to transition from energy poverty and transform the cooking landscape in the global south have been determined in this study. The country-level features with the most impact are (in order of importance) energy consumption, households and NPISHs final consumption expenditure, female literacy, services value addition, electricity access, industry value addition, GDP per capita, agric-forestry-fishing value addition, and fertilizer consumption.

The results reveal that income-generation activities (services, industry, and agric-forestry-fishing value addition, fertilizer consumption) play integral roles in transitioning and sustaining access to clean cooking above 80%. Hydrogen could play a crucial role in these sectors by stimulating manufacturing supply chains on fertilizer production and farming, steel and cement production, and decarbonizing mining and transport sectors. For example, meeting the required final energy consumption for the industrial and transport sectors through a hydrogen-based economy [69] could increase the gross value added by US$92 billion per year and create about 3 million jobs. Meeting the required energy for all sectors through solar photovoltaics [70] could create about 14 million manufacturing, construction, installation, and operation and maintenance jobs. The findings also show that Sub-Saharan African countries should consume between 300 and 400 kg per hectare per year of fertilizer - underscoring the integral role renewable ammonia will play in facilitating ecosystems with self-sufficient clean cooking. This consumption is 28 times more than the current fertilizer consumption.

The impact of electricity access on improving access to clean cooking highlights the absence of a linear relationship - implying that high rates

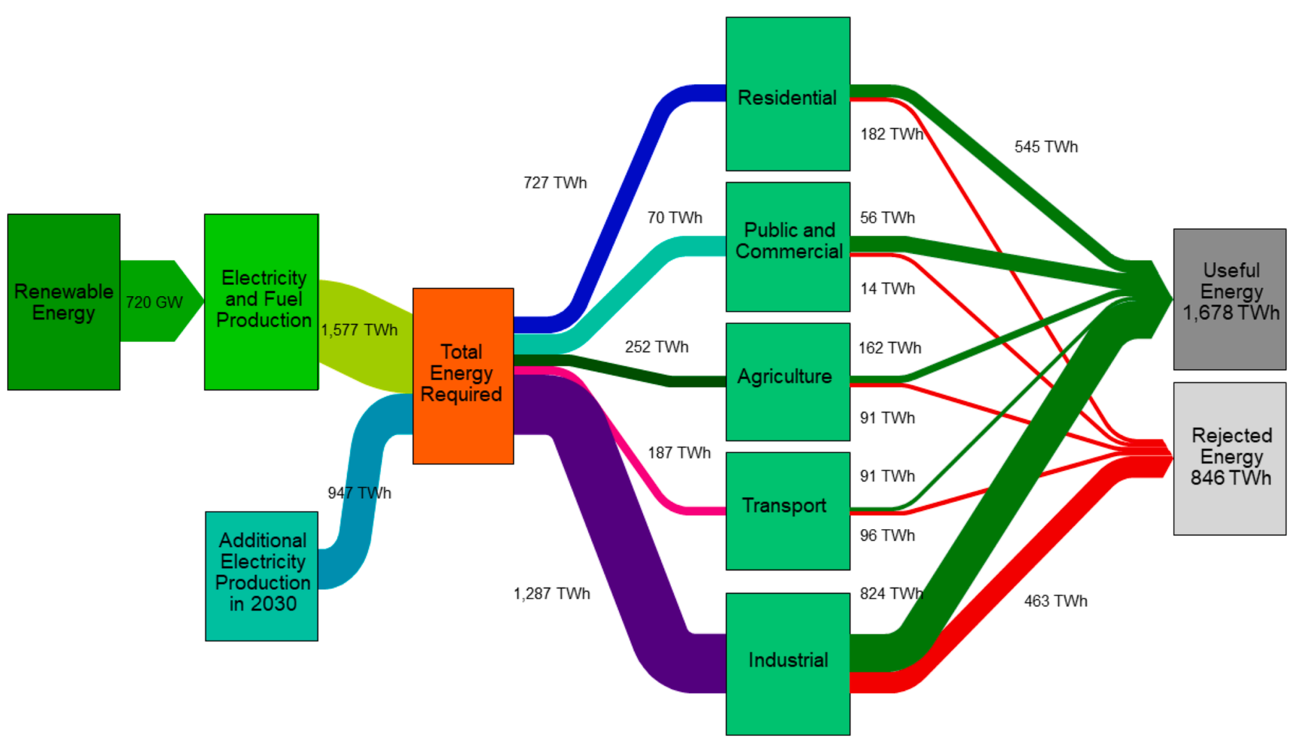


Fig. 8. Final energy consumption flows per sector required to transition to over 80% of access to clean cooking in Africa. The renewable energy capacity input is solar PV (see methods 2.9). A mix of renewable energy technologies will result in a lower required installed capacity due to a higher cumulative capacity factor. The additional (planned) generation capacity (947 TWh) in Sub-Saharan Africa in 2030 is extracted from Alova et al. (2021) and comprises fossil fuel, hydro, non-hydro renewable electricity, and nuclear electricity.

of electricity access do not translate into high rates of access to clean cooking due to high electricity tariffs and load-shedding. This finding necessitates the need for a new fundamental approach that focuses on improving the financial capacity of households through a systems approach so that they can afford electricity or fuels for cooking. It also indicates the significance of using decentralized clean cooking technologies for countries to transition to clean cooking and evidences the role that gases such as green hydrogen could play in improving the access rates of clean cooking as hydrogen becomes cost competitive with liquid petroleum gas. On the other hand, household expenditure indicates a linear relationship, while female literacy shows that over 70% of female literacy rates are required to transition to over 20% of access to clean cooking. The ideal household expenditure to transition to over 80% of access to clean cooking and clean-cooking ecosystems is around US$19.18 per day.

The results show that the country-level indicators have boundaries beyond which more growth has less impact on access to clean cooking. The results revealed that the maximum effect of primary energy consumption in classes 2, 3, and 4 is around 17,000 kWh per capita per year before plateauing. This primary energy consumption translates into a final energy consumption of 8,500 kWh per capita per year after applying a correction factor of 0.5. The income-generating activities in classes 2, 3, and 4, namely, agric-forestry-fishing, industry, and services value addition, including GDP, show plateauing characteristics after about US$200 billion per year and 17,000 US$ per capita per year, respectively. This plateauing indicates boundaries beyond which more growth has less impact on sustaining clean cooking in an ecosystem.

The increment in magnitude required for countries to transition to clean cooking has been determined. For example, the primary energy consumption in sub-Saharan Africa should increase by about 443 MWh per capita per year - equalling the total energy consumption of the UK, Australia, USA, Canada, Germany, Japan, China, France, and Norway (444 MWh per capita per year in 2021 [71]). However, the final energy consumption should increase by about 222 MWh per capita per year - two times lower than the required primary energy consumption. The required increment in residential energy consumption is only 29% of the total final energy needed for African countries to transition to clean cooking. This finding signifies the importance of energy consumption in spurring wealth-generating activities that will sustain a clean-cooking ecosystem. Furthermore, this important finding completely changes the narrative of an energy transition in energy-poor countries. The findings show that sub-Saharan African countries should increase their planned electricity generation to three times the planned generation in 2030. Thus, for developing countries, it is not a matter of transitioning, but a matter of increasing energy capacity. The total renewable potential in Africa (solar PV, wind and concentrated solar power) is 300 times higher than the final energy consumption required to transition to clean cooking. This finding shows that Africa can transition to self-sufficient clean cooking ecosystems through renewable electricity production and energy storage. Beside, the world has sufficient minerals to power the world with renewable energy [72], and Africa has the largest mineral reserves critical in transitioning to a renewable energy economy [73–75].

Finally, the estimated total cost of enabling universal access to cooking with clean energy in Africa is US$14.5 trillion. This cost equals yearly expenditures of US$2.1 trillion by 2030 or US$0.54 trillion by 2050. Other previous studies underestimated the cost of enabling universal access to clean cooking in the global south at US$4.5 - 156 billion per year by 2030 by focusing on expenditure towards stoves and fuels, funding subsidies, and installing modern energy cooking infrastructure [2,3,76]. The estimate in this study encompasses all the intermediate inputs required to build self-sufficient communities by creating value-addition sectors, thus, creating jobs and eliminating poverty. The annual cash flow (US$0.25 trillion) leaving Africa through unjust debt payments, multinationals exploiting tax loopholes, and corruption [77–79] can meet this estimated cost halfway. The annual cash flow from 148 developing countries can meet this estimated cost in just about five years.

Code availability

The python code for the open-source gradient-boosted tree-based machine-learning model in this study is available from the corresponding author upon reasonable request. The Catboost algorithm is freely available on GitHub (https://github.com/catboost/catboost).

CRediT authorship contribution statement

M.D. Mukelabai: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – review & editing, Writing – original draft. **K.G.U. Wijayantha:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. **R.E. Blanchard:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The country-level indicators were collected from the World Bank indicators (https://data.worldbank.org/indicator). The data on primary energy consumption was collected from our World in Data (https://ourworldindata.org/). The data on access rates of clean cooking was collected from the World Health Organization (https://www.who.int/data/gho/data/themes/air-pollution/household-air-pollution). The data on gross value outputs, intermediate inputs, and gross value added was collected from the United Nations (https://data.un.org/Default.aspx). All the data supporting the findings in this study are available within the paper and supplementary materials.

Acknowledgments

M.D.M., K.G.U.W. and R.E.B. are grateful for the financial support offered to M.D.M. for his doctoral research by the UK Engineering and Physical Sciences Research Council and Loughborough University through the EPSRC Sustainable Hydrogen Centre for Doctoral Training funded by the UK Research and Innovation (UKRI) [grant number EP/S023909/1].

Graphical abstract:

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.egyai.2023.100290.

References

[1] UN. The sustainable development goals report. New York, NY, USA: 2022.

[2] Pachauri S, Poblete-Cazenave M, Aktas A, Gidden MJ. Access to clean cooking services in energy and emission scenarios after COVID-19. Nat Energy 2021;6: 1067–76. https://doi.org/10.1038/s41560-021-00911-9.

[3] The state of access to modern energy cooking services (English). Washington DC: The World Bank Group; 2020.

[4] Stoner O, Lewis J, Martínez IL, Gumy S, Economou T, Adair-Rohani H. Household cooking fuel estimates at global and country level for 1990 to 2030. Nat Commun 2021;12:1–8. https://doi.org/10.1038/s41467-021-26036-x.

[5] WHO. Household air pollution - Fact Sheets 2022. 10.1093/med/9780199929573.003.0023.

[6] Gould CF, Pillarisetti A, Thompson LM, Saluja S, Nandan V, Urpelainen J. Using high-frequency household surveys to describe energy use in rural North India during the COVID-19 pandemic. Nat Energy 2023;8:169–78. https://doi.org/10.1038/s41560-022-01187-3.

[7] Khavari B, Ramirez C, Jeuland M, Fuso Nerini F. A geospatial approach to understanding clean cooking challenges in sub-Saharan Africa. Nat Sustain 2023;6:447–57. https://doi.org/10.1038/s41893-022-01039-8.

[8] Perros T, Allison AL, Tomei J, Parikh P. Behavioral factors that drive stacking with traditional cooking fuels using the COM-B model. Nat Energy 2022;7:886–98. https://doi.org/10.1038/s41560-022-01074-x.

[9] Poblete-Cazenave M, Pachauri S. A structural model of cooking fuel choices in developing countries. Energy Econ 2018;75:449–63. https://doi.org/10.1016/j.eneco.2018.09.003.

[10] Malakar Y, Day R. Differences in firewood users' and LPG users' perceived relationships between cooking fuels and women's multidimensional well-being in rural India. Nat Energy 2020;5:1022–31. https://doi.org/10.1038/s41560-020-00722-4.

[11] Moksnes N, Korkovelos A, Mentis D, Howells M. Electrification pathways for Kenya-linking spatial electrification analysis and medium to long term energy planning. Environ Res Lett 2017;12:095008. https://doi.org/10.1088/1748-9326/aa7e18.

[12] Dagnachew AG, Lucas PL, Hof AF, Gernaat DEHJ, de Boer HS, van Vuuren DP. The role of decentralized systems in providing universal electricity access in Sub-Saharan Africa - A model-based approach. Energy 2017;139:184–95. https://doi.org/10.1016/j.energy.2017.07.144.

[13] Dagnachew AG, Lucas PL, Hof AF, van Vuuren DP. Trade-offs and synergies between universal electricity access and climate change mitigation in Sub-Saharan Africa. Energy Policy 2018;114:355–66. https://doi.org/10.1016/j.enpol.2017.12.023.

[14] Panos E, Densing M, Volkart K. Access to electricity in the World Energy Council's global energy scenarios: an outlook for developing regions until 2030. Energy Strateg Rev 2016;9:28–49. https://doi.org/10.1016/j.esr.2015.11.003.

[15] Ahlborg H, Boräng F, Jagers SC, Söderholm P. Provision of electricity to African households: the importance of democracy and institutional quality. Energy Policy 2015;87:125–35. https://doi.org/10.1016/j.enpol.2015.09.002.

[16] Wickramasinghe A. Energy access and transition to cleaner cooking fuels and technologies in Sri Lanka: issues and policy limitations. Energy Policy 2011;39:7567–74. https://doi.org/10.1016/j.enpol.2011.07.032.

[17] Murshed M. Pathways to clean cooking fuel transition in low and middle income Sub-Saharan African countries: the relevance of improving energy use efficiency. Sustain Prod Consum 2021;30:396–412. https://doi.org/10.1016/j.spc.2021.12.016.

[18] Hollands AF, Daly H. Modelling the integrated achievement of clean cooking access and climate mitigation goals: an energy systems optimization approach. Renew Sustain Energy Rev 2023;173:113054. https://doi.org/10.1016/j.rser.2022.113054.

[19] Fuso Nerini F, Dargaville R, Howells M, Bazilian M. Estimating the cost of energy access: the case of the village of Suro Craic in Timor Leste. Energy 2015;79:385–97. https://doi.org/10.1016/j.energy.2014.11.025.

[20] Dioha MO, Kumar A. Exploring sustainable energy transitions in sub-Saharan Africa residential sector: the case of Nigeria. Renew Sustain Energy Rev 2020;117:109510. https://doi.org/10.1016/j.rser.2019.109510.

[21] Poblete-Cazenave M, Pachauri S, Byers E, Mastrucci A, van Ruijven B. Global scenarios of household access to modern energy services under climate mitigation policy. Nat Energy 2021;6:824–33. https://doi.org/10.1038/s41560-021-00871-0.

[22] Pachauri S, Rao ND, Cameron C. Outlook for modern cooking energy access in Central America. PLoS ONE 2018;13. https://doi.org/10.1371/journal.pone.0197974.

[23] Dagnachew AG, Hof AF, Lucas PL, van Vuuren DP. Scenario analysis for promoting clean cooking in Sub-Saharan Africa: costs and benefits. Energy 2020;192:116641. https://doi.org/10.1016/j.energy.2019.116641.

[24] Grubler A, Wilson C, Bento N, Boza-Kiss B, Krey V, McCollum DL, et al. A low energy demand scenario for meeting the 1.5 °c target and sustainable development goals without negative emission technologies. Nat Energy 2018;3:515–27. https://doi.org/10.1038/s41560-018-0172-6.

[25] Alova G, Trotter PA, Money A. A machine-learning approach to predicting Africa's electricity mix based on planned power plants and their chances of success. Nat Energy 2021;6:158–66. https://doi.org/10.1038/s41560-020-00755-9.

[26] Alova G. A global analysis of the progress and failure of electric utilities to adapt their portfolios of power-generation assets to the energy transition. Nat Energy 2020;5:920–7. https://doi.org/10.1038/s41560-020-00686-5.

[27] Kar A, Pachauri S, Bailis R, Zerriffi H. Using sales data to assess cooking gas adoption and the impact of India's Ujjwala programme in rural Karnataka. Nat Energy 2019;4:806–14. https://doi.org/10.1038/s41560-019-0429-8.

[28] Shupler M, Mangeni J, Tawiah T, Sang E, Baame M, Anderson de Cuevas R, et al. Modelling of supply and demand-side determinants of liquefied petroleum gas consumption in peri-urban Cameroon, Ghana and Kenya. Nat Energy 2021;6:1198–210. https://doi.org/10.1038/s41560-021-00933-3.

[29] Murshed M. The relevance of reducing income inequality for eliminating urban-rural divide in clean cooking fuel accessibility: evidence from Latin America and the Caribbean. Energy 2023;278:127718. https://doi.org/10.1016/j.energy.2023.127718.

[30] Ali J, Khan W. Factors affecting access to clean cooking fuel among rural households in India during COVID-19 pandemic. Energy Sustain Dev 2022;67:102–11. https://doi.org/10.1016/j.esd.2022.01.006.

[31] Byaro M, Dimoso P, Rwezaula A. Are clean energy technologies a panacea for environmental sustainability in sub-Saharan African countries? Environ Sci Pollut Res 2023. https://doi.org/10.1007/s11356-023-28438-4.

[32] Goswami A, Bandyopadhyay KR, Singh P, Gurtu A. Rural Energy Transition for Cooking in India—Revisiting the Drivers. Sustainability 2023;15:7635. https://doi.org/10.3390/su15097635.

[33] The world bank. Indicators 2022. https://data.worldbank.org/indicator (accessed February 1, 2022).

[34] Our World in Data. Research and data to make progress against the world'ss largest problems 2023:1–26. https://ourworldindata.org/ (accessed February 27, 2023).

[35] US energy information administration. International Energy Data 2023. https://www.eia.gov/international/data/world (accessed July 9, 2023).

[36] BP. Statistical review of world energy 2023. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html (accessed July 9, 2023).

[37] WHO. Household air pollution 2023. https://www.who.int/data/gho/data/themes/air-pollution/household-air-pollution (accessed February 27, 2023).

[38] UN. UN data: a world of information 2023:2023. https://data.un.org/Default.aspx (accessed February 27, 2023).

[39] Dorogush A.V., Ershov V., Gulin A. CatBoost: gradient boosting with categorical features support. ArXiv Prepr 2018:1–7.

[40] Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: unbiased boosting with categorical features. ArXiv Prepr 2017:1–23.

[41] Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal 2002;38:367–78. https://doi.org/10.1016/S0167-9473(01)00065-2.

[42] Yandex. CatBoost is a high-performance open source library for gradient boosting on decision trees 2023:1–12. https://catboost.ai/news/catboost-enables-fast-gradient-boosting-on-decision-trees-using-gpus (accessed February 27, 2023).

[43] Friedman J.H. Greedy function approximation: a gradient boosting machine 1999.

[44] Yandex. CatBoost 2023. https://github.com/catboost/catboost (accessed July 17, 2022).

[45] Shapley LS. A value for n-Person games. Contributions to the theory of games (AM-28). Princeton University Press; 1953. p. 307–18. https://doi.org/10.1515/9781400881970-018. vol. II, vol. 23.

[46] Lundberg SM, Lee S. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems; 2017. p. 1–10.

[47] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2:56–67. https://doi.org/10.1038/s42256-019-0138-9.

[48] Ding R, Yin W, Cheng G, Chen Y, Wang J, Wang R, et al. Boosting the optimization of membrane electrode assembly in proton exchange membrane fuel cells guided by explainable artificial intelligence. Energy AI 2021;5:100098. https://doi.org/10.1016/j.egyai.2021.100098.

[49] Meduri S, Nandanavanam J. Prediction of hydrogen uptake of metal organic frameworks using explainable machine learning. Energy AI 2023;12:100230. https://doi.org/10.1016/j.egyai.2023.100230.

[50] Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, et al. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: review, challenges and opportunities. Energy AI 2022;9:100169. https://doi.org/10.1016/j.egyai.2022.100169.

[51] Trebbien J, Rydin Gorjão L, Praktiknjo A, Schäfer B, Witthaut D. Understanding electricity prices beyond the merit order principle using explainable AI. Energy AI 2023;13:100250. https://doi.org/10.1016/j.egyai.2023.100250.

[52] Radaideh MI, Rigopoulos S, Goussis DA. Characteristic time scale as optimal input in Machine Learning algorithms: homogeneous autoignition. Energy AI 2023;100273. https://doi.org/10.1016/j.egyai.2023.100273.

[53] Lundberg S.M., Erion G.G., Lee S. Consistent individualized feature attribution for tree ensembles. ArXiv Prepr 2019.

[54] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. ACM; 2019. p. 2623–31. https://doi.org/10.1145/3292500.3330701.

[55] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn : machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[56] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery & data mining. ACM; 2016. p. 785–94. https://doi.org/10.1145/2939672.2939785. vol. 13-17- Augu.

[57] Mehta P, Bukov M, Wang CH, Day AGR, Richardson C, Fisher CK, et al. A high-bias, low-variance introduction to Machine Learning for physicists. Phys Rep 2019;810:1–124. https://doi.org/10.1016/j.physrep.2019.03.001.

[58] OECD. OECD data: live dataset 2023:2–3. https://stats.oecd.org/Index.aspx?DataSetCode=DP_LIVE (accessed February 27, 2023).

[59] Ritchie H. Primary energy production is not final energy use: what are the different ways of measuring energy? 2021:1–8. https://ourworldindata.org/energy-substitution-method (accessed February 27, 2023).

[60] United Nations. Household size and composition around the world. vol. 2. New York, USA: 2017.

[61] Mukelabai MD, Wijayantha KGU, Blanchard RE. Hydrogen technology adoption analysis in Africa using a Doughnut-PESTLE hydrogen model (DPHM). Int J Hydrog Energy 2022;47:31521–40. https://doi.org/10.1016/j.ijhydene.2022.07.076.

[62] The World Bank. Egypt to continue expanding household connections to natural gas grid 2014:1–6. https://www.worldbank.org/en/news/feature/2014/07/24/egyptian-households-to-abandon-lpg-cylinders (accessed February 22, 2022).

[63] Sustainable Energy for All. SEforALL analysis of SDG7 progress - 2022 2022. https://www.seforall.org/data-stories/seforall-analysis-of-sdg7-progress#:~:text

=Access to clean cooking&text=Improvements have been made in,million to ∼940 million people. (accessed February 23, 2023).
[64] IEA. Africa energy outlook 2019. Paris: 2019.
[65] IEA. SDG7: data and projections. Paris: 2022.
[66] Mukelabai MD, Wijayantha KGU, Blanchard RE. Hydrogen for cooking: a review of cooking technologies, renewable hydrogen systems and techno-economics. Sustainability 2022;14:16964. https://doi.org/10.3390/su142416964.
[67] Dotan H, Landman A, Sheehan SW, Malviya KD, Shter GE, Grave DA, et al. Decoupled hydrogen and oxygen evolution by a two-step electrochemical-chemical cycle for efficient overall water splitting. Nat Energy 2019;4:786–95. https://doi.org/10.1038/s41560-019-0462-7.
[68] Mukelabai MD, Wijayantha UKG, Blanchard RE. Renewable hydrogen economy outlook in Africa. Renew Sustain Energy Rev 2022;167:112705. https://doi.org/10.1016/j.rser.2022.112705.
[69] The Scottish Government. The scottish hydrogen assessment. Edinburgh: 2020.
[70] IRENA. Renewable energy and jobs. Abu Dhabi: 2013.
[71] Our world in data. Energy use per person 2023. https://ourworldindata.org/grapher/per-capita-energy-use (accessed May 4, 2023).
[72] Wang S, Hausfather Z, Davis S, Lloyd J, Olson EB, Liebermann L, et al. Future demand for electricity generation materials under different climate mitigation scenarios. Joule 2023:309–32. https://doi.org/10.1016/j.joule.2023.01.001.
[73] IEA. Africa energy outlook 2022. Paris: 2022. 10.1787/g2120ab250-en.
[74] IEA. The role of critical minerals in clean energy transitions. Paris: 2021.
[75] United Nations. Our work in Africa 2023:2–3. https://www.unep.org/regions/africa/our-work-africa (accessed February 27, 2023).
[76] Sustainable Energy for All. Energizing finance: understanding the landscape. Vienna: 2020.
[77] Global Financial Integrity. Illicit financial flows to and from developing countries: 2005-2014. Washington DC: 2019.
[78] Jonathon P. A decade to confront the climate emergency. UK: Simon & Schuster; 2020.
[79] Curtis M., Jones T. Honest accounts 2017: how the world profits from Africa's wealth. London: 2017.