Original Research Article

OBESITY RISK PREDICTION USING GREEN AI STRATEGY

Abstract

Aims/ objectives: With the consistent global increase in obesity population, researchers have proposed both machine learning and neural network techniques in predicting obesity beforehand from lifestyle and demographic factors. However, extant literatures in the pipeline subscribes to existing Red AI approach, that is, enhancing model's accuracy at whatever cost without recourse to the effect of such model on the environment in terms of carbon footprint and efficiency.

Study design: This paper set forward a green AI model selection strategy for obesity risk prediction which depends on multi-criteria determination method, factoring in computation time, in choosing the optimal model for final deployment.

Methodology: In this research, we applied two deep learning and three machine learning models. We used the algorithms of Convolution Neural Network (CNN), Artificial Neural Network (ANN), Random Forest (RF), Logistic Regression (LR) and Decision Trees (DT) learning algorithms. Our proposed method identifies the optimal model for final deployment based on equal compromise between model's performance and computation time. The estimated Carbon footprint and Energy consumption of training the models used in this study have been computed using green Algorithms.

Results: From our comparative analysis, our green AI selection strategy favored Random forest model, which scored 97.16% in accuracy, took 0.03420 seconds of computation time, 2.36mg of CO_2e carbon footprint and 2.62×10^{-3} WH energy consumption during model training and validation.

Conclusion: This paper's contributions are significant to the support of the ongoing call for Green AI, especially within the healthcare sector.

Keywords: Obesity; Machine Learning; Green AI; Deep Learning; Carbon Footprint; Energy Consumption

1 Introduction

Artificial Intelligence (AI) has found widespread application most especially in the healthcare sector, over the past few years. AI has been used to improve medical diagnosis (Kandar and Praba, 2020), speed up drug discovery (Alexander et al., 2023), manage healthcare data (Tatineni, 2022) and perform robotic surgery (Joseph et al., 2023). In the case of applying AI to medical diagnosis, various diseases have been identified and treated from its signs and symptoms. Diseases such as Cancer (Xiaoyin et al., 2023), Prostate disease (Boluwaji et al., 2023), diabetes (Fayroza and Abbas, 2023) and so on has been diagnosed or detected by using AI method. Obesity risk has also been predicted using AI approach for different categories of individuals like children (Pritom et al., 2023), adolescents (Orsolya et al., 2024) and adults (Maria et al., 2023). Obesity is a global health burden, defined as excess accumulation of fat in the body. The disease has increased three times between 1975 and 2013. In 2022, 1 in 8 people in the world were living with obesity, which is about 890 million adults above the age of 18 years were living with the disease and if the trend continues to rise, it is estimated that close to one third of global population of adults will be overweight and more than one billion will be obese by 2025(Organization, 2019). Obesity is a major cause of several chronic diseases such as cardiovascular disease, cancer, arthritis and diabetes, which are the leading causes of death around the globe (Hurby et al., 2016).

Researchers have applied several machine learning and deep learning algorithms for the prediction of obesity risk, employing diverse techniques. Most literatures on the subject, have used machine learning models on dataset containing the physical condition, eating habit and physical description of the participants as a criterion for their predictive analysis (Nirmala et al., 2022; Maria et al., 2023; Elias et al., 2021; Faria et al., 2021). Other prediction approaches involve the application of obese thermal images (Snekhalatha and Thanaraj, 2021), features of the built environment (Adyasha and Elaine, 2018) and genetic factor (Fatemeh et al., 2016). Both machine learning and deep learning algorithms employed in these studies have recorded significant results. Nirmala et al. trained five machine learning algorithms on datasets from two sources; real dataset gotten from students from various colleges in Tamil Nadu and dataset from University of California Irvine (UCI) machine learning repository. Their dataset comprise of physical description, eating habits and physical condition features mapped to the obese status of the participants. From their comparative analysis, Random forest classifier had an overall best accuracy of 99.% before tuning. After tuning, Logistic regression had the highest accuracy of 99.68% (Nirmala et al., 2022). Faria et al. (2021) developed an online application with a user friendly interface, which takes inputs from users, about certain physical factors revolving around their daily activities such as height, weight, food routine etc. Users get feedback from the online system on their mobile interface. To develop the system, nine different machine learning algorithms were trained on a dataset obtained from both online and offline sources from students, club members and campuses in Dhaka city. Logistic regression had the highest accuracy of 97.0% in their study. Snekhalatha and Thanaraj (2021) designed a customized deep learning model for grouping into obese and normal classes forearm, abdomen and shank thermal images of 100 participants. In their study, they designed four new customized Convolution Neural Network (CNN) models for recognizing Brown Adipose Tissue (BAT) activation in the thermal photographs. In their study, VGG-16 had the overall best accuracy of 79% on obesity detection among the pre-train neural network category while Custom-2 had the best accuracy of 92% on the task among the Customized Networks. Adyasha and Elaine (2018) applied 2014 yearly approximate of prevalent obesity derived from 500 cities project census. Six cities in the USA were selected for the study. They applied an 8 layered VGG-NN-F network to learn from an approximate of 1.2 million ImageNet pictures to identify

features relating to 1000 categories. They observed that when these environmental features are combined with other data, it can be used to evaluate the rate of obesity to assist programs targeted on reducing obesity risk. Their findings presented a solid relationship between a person's surrounding and obesity level. Fatemeh et al. (2016) observed the relationship between biological risk factors in childhood relating to their surrounding and medical factors, using the Gradient Boosting model on a dataset obtained from cardiovascular risk in young Finns study cohort. It was observed in their study that the action of gene in predicting obesity risk is more accurate in infants compared to older children.

However, most of these researchers were only concerned on improving accuracy of their model, a Red AI approach of model selection (i.e. choosing model purely based on their accuracy) without factoring in the efficiency and environmental friendliness of the models. Also, little attention was paid to model computation time and the carbon footprint of training the models. Similar trend is noticeable outside health care theme, developers have trained a lot of heavy weight models consuming humongous power and greatly increasing carbon emission. Researchers from Open AI trained GPT-3 with 175 billion parameters for 355 hours and it emitted almost 500 million gram of carbon dioxide equivalent (gCO2e), same as the emission of five cars in their life time (David et al., 2021). Other similar model such as BLOOM with 176 billion parameters for 1200 hours (Teven et al., 2023), Gopher with 280 billion parameters for 920 hours (Jordan et al., 2022), Palm with 540 billion parameters for 1200 hours (Rohan et al., 2023) and GPT-4 with 1,800 billion parameters for 2280 hours (Bates et al., 2022). The computation time of training these models grows linearly with their carbon footprint (Lannelongue et al., 2021).

Besides the computation time, the processing cost of such models are enormous, model like Alpha GO requires 1920 CPUs and 280 GPUs, the cost of reproducing the same experiment was about $35,000,000 (Silver et al., 2017). Schwartz et al. (2020) referred to AI research studies of this nature as Red AI, where machine learning or deep learning models are pushing the state-of-the-art and results at whatever cost. They gave the name Green AI to AI studies that reduces cost of computation and minimizes the resources spent (Schwartz et al., 2020). Even though, Green and expense sensitive AI study is somewhat modern, in health informatics, a handful of researchers have followed the practice of efficient AI technology. In these researches, certain features of the model are optimized during training and prediction by the researchers to improve efficiency thereby freezing out extraneous ones. Some of the promising approaches to making Green in AI include; Algorithm optimization, hardware optimization, data center optimization, pragmatic scaling factor and the use of energy consumption tools like CarbonTracker, CodeCarbon, Green Algorithms and PowerTop (Veronica et al., 2024). Ezenkwu et al. (2023) in detecting Mpox disease has applied the Green AI strategy for selecting optimal model. They developed a model inefficiency equation which incorporates model performance and computation time for choosing a model. The outcome of their strategy for model selection is the same as expert's decision on optimal model in a situation where decision should be made between computation time and model error. Similarly, the cost of feature annotation are incorporated in the process of feature selection by other researchers (Wei et al., 2019; Das et al., 2021). Outside Healthcare, Lannelongue et al. (2021) developed a simple technique that uses known factors such as hardware, runtime, tool requirements and data center location to estimate the carbon footprint of the model. Our study supports the current shift from Red to Green AI technology through the strategic model selection by applying obesity risk prediction for analysis.

The main purpose of this paper is the novelty of applying Green AI strategy in obesity risk prediction. This paper gives insight that compels the upshot or a more environmentally friendly and efficient AI technology within healthcare and beyond. Five different models were selected for the experiment based on popular opinion from literature review. The algorithms of CNN, ANN, LR, RF and DT were applied, showing clearly that the strategy nominated can determine the best model depending on the researchers preferred variable that bears the required outcome. The remainder of this paper is as follows: Section 2.0 presents relevant background information on Green AI strategy. The research methodology has been provided in Section 3.0, while results and discussions are

illustrated in Section 4.0. Section 5.0 concludes the paper.

2 Green AI Strategy

By definition, Green AI is an AI study that seeks to reduce the cost of computation; resources spent and encourage techniques that promote a balance between performance and efficiency Schwartz et al. (2020). Conversely, Red AI is a research in AI with the ultimate goal of improving model accuracy (i.e. performance) and thereby consuming a lot of resources at whatever cos26 Schwartz et al. (2020). To illustrate this, in Natural Language Processing (NLP) Brown et al. (2020), Chat GPT-3 is among the best models, but has 175 billion variables and 96 layers. Consequently, in Computer Vision (CV) Xie et al. (2020), EfficientNet-L2 is one of the best models but with 480 million parameters and learned from 130 million images. The craving for large parameters and training data necessitated the use of GPUs and TPUs for deep learning task to speed up the training process. The great concern here becomes, hardware cost, power consumption and carbon footprint accruing to the intensive energy used by the hardware Anthony et al. (2020), Erion et al. (2021). Schwartz et al. (2020) developed an equation for Red AI incorporating variables responsible for the cost of a Red AI Result(R), that cost is directly proportional to the amount spent in processing an Example (E), the quantity of Dataset(D) trained and how many Hyperparameters(H) used, as shown in Equation (1).

$$\text{cost}(R) \propto E \cdot D \cdot H \tag{1}$$

In this study, we applied the model inefficiency equation (Fatemeh et al., 2016) for the selection of our optimal model as illustrated in Equation (2). Equation (2) depends is a statistical product model. It factors in each models prediction time and error (100% - accuracy) for the selection of final deployment model depending on how it performed and how efficient it was on the task.

$$\eta^{\langle m \rangle} = \left(\frac{e^{\langle m \rangle}}{\|\{e^{\langle m \rangle}\}_{m=1}^{M}\|_{\infty}} \right) \alpha + \left(\frac{\tau^{\langle m \rangle}}{\|\{e^{\langle m \rangle}\}_{m=1}^{M}\|_{\infty}} \right) (1 - \alpha) \tag{2}$$

From Equation (2), m represents model 'm' inefficiency $m_1, \ldots, M$; $\tau^{(m)}$ and $e^{(m)}$ represent the computation time and error of model 'm' respectively, $\|.\|_{\infty}$ is the infinity norm, and α stands for an inclusive number from 0 to 1. α describes how much attention is paid to error relative to computational time. If α is 1, then the models are selected entirely based on error, and if α is 0, the computational time becomes the only factor for model selection. An α of 0.5 provides equal attention to the two parameters. Equation (3) is the optimized model equation which reduces the equation of inefficiency as shown below:

$$m^{*} = \arg\min_{m} \left(\eta^{\langle m \rangle} \right) \tag{3}$$

Loic *et al.* (2021) developed equations for computing Energy Consumption (EC) and Carbon Footprint (CF), shown in Equations (4) and (5) respectively, using known operational factors as follows:

$$EC = t \times (n_c \times P_c \times u_c + n_m \times P_m) \times PUE \times 0.001 \tag{4}$$

Where *EC* is the energy consumption (in kWh), t represents the running time in hours, n_c is the number of CPU cores, u_c is the usage factor, P_c is the power draw per core, n_m is the memory usage in gigabytes, P_m is the memory power draw, and *PUE* is the Power Usage Effectiveness of the data center.

$$CF = EC \times CI \tag{5}$$

Where *CF* is the carbon footprint, *EC* is the energy consumption, and *CI* is the carbon intensity (i.e., the carbon footprint per kilowatt-hour of energy).

3 Methodology

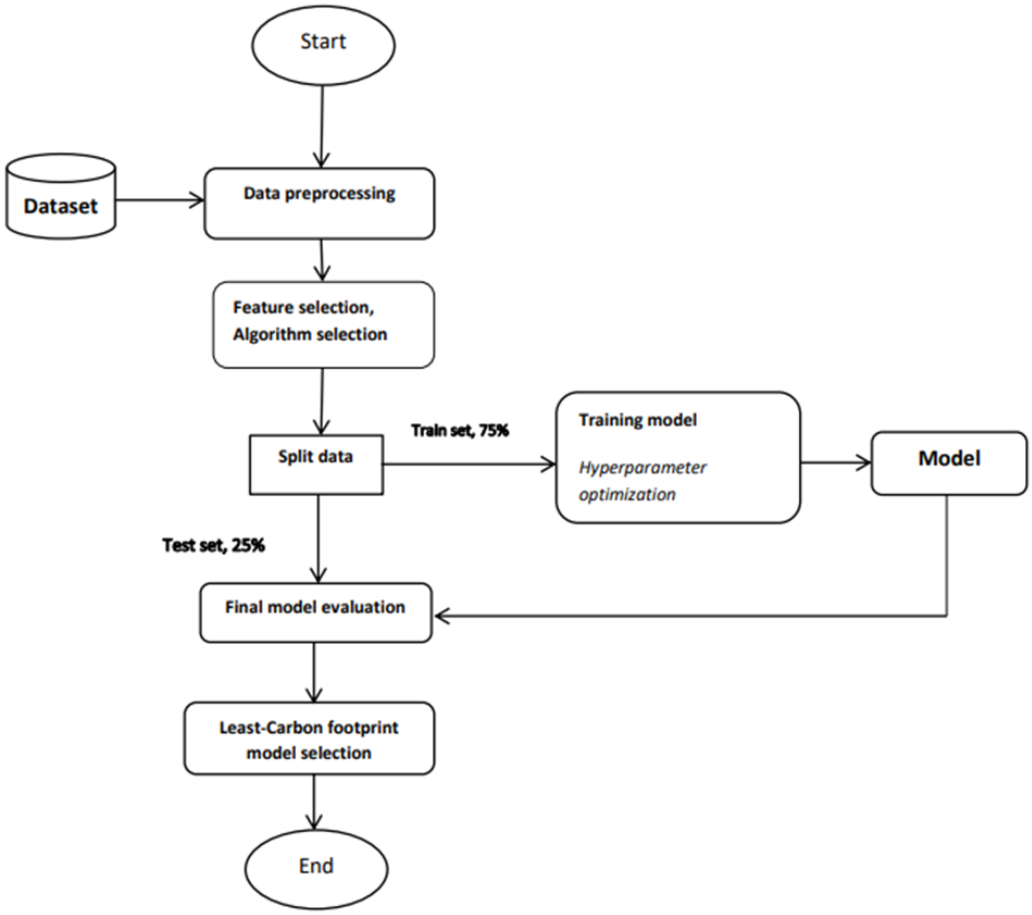

Figure 1: The System Model

3.1 System Model

Figure 1 illustrates the system Architecture for model development. Before training the selected models for the study on the dataset; we first carried out the data preprocessing step. The dataset preprocessing method involve: label encoding, feature scaling, data normalization and segmentation. Features represent inputs to be used for the prediction, while label represents that data used for the prediction. In most cases, dataset may contain string column that disobeys the principle of tidy data, label encoding translates labels into numeric format so they may be readable by the machine. Based on alphabetical order, each label is issued a unique integer. Feature scaling normalizes the range of the independent variables or features of data. This is a data preprocessing step that shifts and rescales values so they range between 0 and 1. StandardScaler imported from Scikit-learn library was used to scale the dataset. For normalization process, Min-Max Scaler was used to ensure that all the features are on the same scales, preventing bias in the model towards features with larger

scale thereby improving the models' ability to generalize and make accurate prediction. Finally, the dataset was segmented into two, 75% for training and 25% for testing and evaluation. This division was to ensure that the model is trained on a sufficiently large portion of the data to learn patterns and relationships while the test set serves as an independent validation to assess how well the model generalizes to unseen data.

3.2 System Setup

All experiments were conducted using Ubuntu 23.04 Operating System 64 bit with Intel Core i5, 2.30GHz frequncy and 8Gigabyte RAM. We used the Scikit-learn, Numpy, Pandas, Matplot and Seaborn libraries on a Jupyter Notebook for the experiments. For the deep learning training, Keras was used, a free downloadable source library having a high level interface with Tensorflow for the different deep learning architectures. Three machine learning models were selected for the study, Logistic regression, Decision tree and Random forest models, which were selected based on popular suggestions from other authors. Logistic Regression model was hyper parameter tuned as follows: Max-iteration =1000 and 'lib-linear' solver. Decision tree model was tuned using Criterion='entropy' while Random forest model tuned by n-estimator=200 and criterion ='entropy'. We also experimented on two deep learning models selected based on further studies from previous authors on the same task. The deep learning models, Artificial Neural Network (ANN) and Convolution Neural Network (CNN) where trained on TensorFlow framework. Both Neural Network models underwent similar tuning; Adam Optimizer, batch size of 32, learning rate of 0.1, activation function of Relu, 100 Epochs and Categorical-Crossentropy loss function as shown in table 1. The obesity risk prediction classes were categorized into seven different classes: Insufficient weight, Normal weight, Overweight level 1, Overweight level 2, Obesity type 1, Obesity type 2 and Obesity type 3. Beside the normal class category, those in other categories were at a risk of suffering from obesity co-morbidities.

Table 1: Deep learning model configuration

Parameter	**Values**
Optimizer	Adam
Batch size	32
Learning rate	0.1
Metrics tracked	Accuracy, precision, recall, F1-score, and computation time
Activation Function	ReLU
Epochs	100
Loss function	Categorical cross-entropy

4 Results and Discussion

4.1 Dataset

The dataset used for this study was gotten from UCI online machine learning repository on obesity. The dataset had 16 attributes which describes the eating habits, the physical conditions and physical descriptions of 2111 participants from Peru, Columbia and Mexico. Attributes associated with eating habits were: Frequency of consuming vegetables (FCVC), Number of main meals (NCP), consuming

food between meals (CAEC), water intake (CH20), and alcohol intake (CALC). Those features connected to physical conditions were: Calories consumption monitoring (SCC), regular exercise (FAF), Technological device usage (TUE), Means of mobility (MTRANS), while attributes related to physical description of participants were: Gender, age, height and weight. All these attributes were mapped to the obesity status of the participants, comprising of seven different classes which are: Insufficient weight, Normal weight, Overweight level 1, Overweight level 2, Obesity type 1, Obesity type 2 and Obesity type 3. The different classes were unevenly distributed with Obesity level 3 having the largest percentage of about 16.62% and Insufficient weight level with the least percentage of 12.85%. Dataset was saved in Comma Separated variable format (.CSV) and has two set of data types, object and float data types.

4.2 Evaluation Metrics

Some the performance metrics deployed in the study for evaluating the models were: Accuracy, Precision, Recall, F1-score and Computation time.

4.2.1 Accuracy

Accuracy in classification problem, represent the ratio of the number of accurate predictions of the model to all types of prediction that we're conducted. Challenge however associated with accuracy is it's assumption of equal cost for various error kinds. For example when a model has an accuracy of 90% it can be said to be excellent, very good, good, mediocre, poor or terrible based on the task performed. Accuracy is calculated has shown in Equation (1).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

4.2.2 Precision

Precision in classification problem represents the ratio of the total number of correctly classified positives to the total number of positives predicted by the model. Model with high precision signifies that its positive label is truly positive (i.e. low false positives). Precision was calculated as shown in Equation (2).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

4.2.3 Recall

Recall is calculated by dividing positively predicted observations by the total positively predicted observations. It signifies the capability of the model to forecast all the positive instances. Recall was calculated as shown in Equation (3).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

4.2.4 F1 Score

F1-score represents the average values of recall and precision. It takes into account FP and FN. In a situation where there is an uneven distribution of class, F1 score is more important than accuracy. Accuracy is excellent where FP and FN have similar cost. Where the cost of FP and FN is different, considering precision and recall is best. F1 score was calculated as shown in Equation (4).

$$\text{F1-score} = \frac{2 * RECALL * PRECISION}{RECALL + PRECISION} \tag{4}$$

Where:

- **True Positives (TP):** Represent cases where both the observed and predicted class are True. For example, when an obese person is classified by the model as obese.
- **True Negatives (TN):** Represent cases where both the observed and predicted class are False. For example, when a non-obese person is classified as non-obese.
- **False Positives (FP):** Represent cases where the observed class is False and the predicted class is True. For example, when a non-obese person is classified as obese.
- **False Negatives (FN):** Represent cases where the observed class is True and the predicted class is False. For example, when an obese person is classified as non-obese.

4.2.5 Computation Time

Model computation time signifies a quantity of time needed to get prediction from a new input for a model with a batch size of one. In other words, it is the time the algorithm spends in making prediction relative to what it is predicting. Computation time is measured in seconds, the SI unit of time. By acknowledging the implications of model computation time, researchers and practitioners can prioritize efficiency, scalability and sustainability in AI and machine learning applications. Slow computation time hinders real time processing, affecting application such as medical diagnosis. Computation time impacts the number of inputs processed per unit time, influencing the overall system performance. Increased computation time is linked with increased carbon footprint of the model and limits model scalability making it challenging to handle large datasets or user bases.

4.3 Results

After training the selected model for the study on the selected dataset, it was necessary to validate their performance using the validation set by conducting a comparative analysis with performance metrics such as accuracy, recall, precision, F1-score and computation time. The ANN model exhibited a loss of 1.9157 initially which steadily decreased to 0.0489 while the accuracy rose from 17.14% to 99.6% on the training set. Also, the CNN model initially exhibited a loss of 1.7040 which progressively decreased to 0.2457 while the accuracy increased from 37.44% to 90.92% on the training set as shown in figure 2. Table 2 illustrates the performance of the models in terms of accuracy, precision, recall and F1-score. From the result, all the models scored above 90% in all the performance metrics tracked. Nevertheless, Random forest model had the overall best performance in all the metrics tracked while Logistic Regression performance was the worst on the task. Models with high accuracy enable informed decisions and dependency on technology. However, accuracy of a model could be influenced by data quality control, model evaluation and testing. To improve accuracy of the model some researchers used ensemble methods, transfer learning techniques, regularization method and data augmentation. Increasing model complexity often improves accuracy but requires more computation resources and time. Large training dataset can improve accuracy but require more computation time. Choice of algorithm affects both accuracy and computation time. Conversely, the implication of rapid model computation time can be significant affecting various aspects of AI applications. Rapid computation time enables real time processing, influences the overall system performance and encourages model scalability. Apart from performance implication, fast computation time requires few computation resources, decreasing energy consumption and cost. Health care diagnosis demands high accuracy but computation time is crucial. Whereas existing researchers in the pipeline follow the strategy of Red AI favoring models with the best performance only, we adopted the approach of Green AI to select our best model for final deployment. Two factors were

considered for the selection of our final model. These factors are performance and computation time. Figure 3 illustrates each models computation time, energy consumption and carbon footprint of the models. The result demonstrates that CNN model have a higher computation time than ANN model as well as all the ML models. The computation time required to train both neural network models in this study was more than 12 times the computation time needed to train the three ML models. This is because Neural Network models have a larger number of parameters, layers and connections. They employ nonlinear activation functions which require more computation than linear transformations used in traditional machine learning. They often use iterative optimization algorithms (for example stochastic gradient descent, Adam) that require multiple passes through the data. In order to prevent over fitting in Neural Networks, techniques like regularization, dropout and early stopping, add computation time overhead. Neural network often requires multiple epochs, increasing overall training time. Nevertheless, Random forest model had the least computation time, about 27 times less than the computation time of the CNN model. Figure 3 also showed that the ML models recorded less carbon footprints and energy consumption as compared to their neural network model counterparts. This agrees with the linearity of relationship between model computation time, carbon footprint and energy consumption of the models.

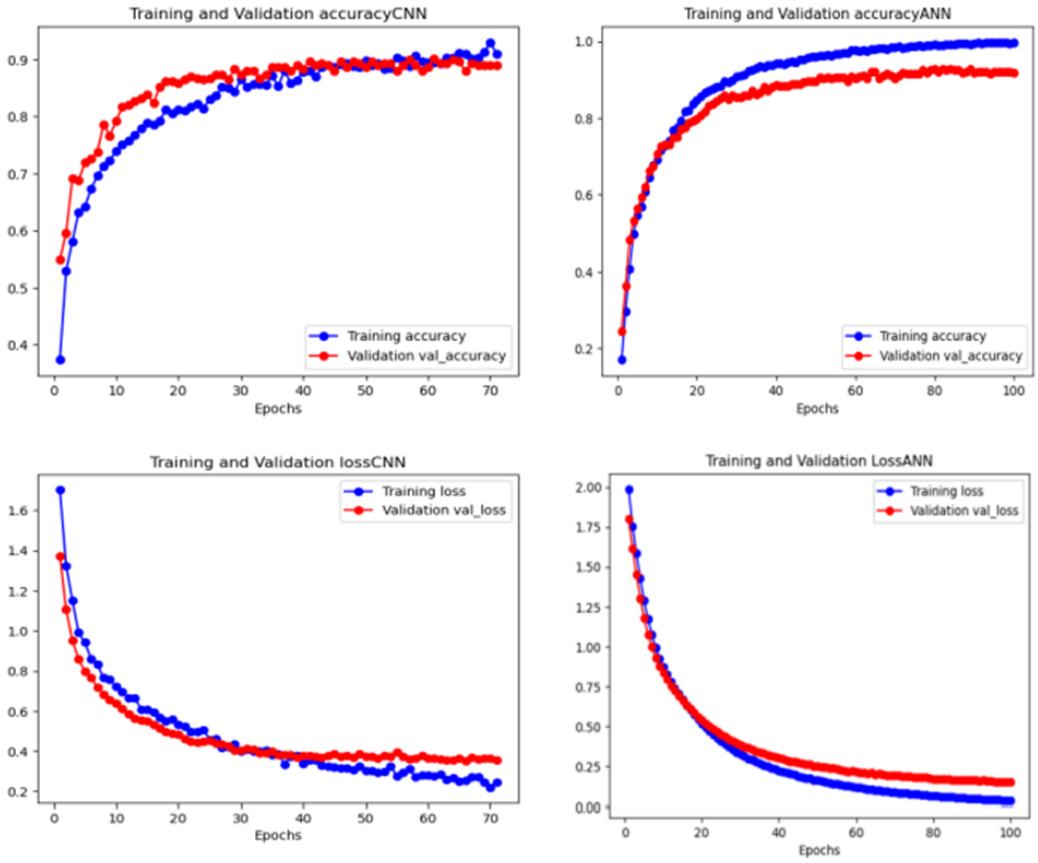

Figure 2: Training and validation loss for ANN and CNN models respectively

Figure 4 shows a graphical representation of model error, computation time, and inefficiency

plotted on the same axes to demonstrate the choice of M^* at $\alpha = 1$. When $\alpha = 1$, the choice of model depends entirely on error, similar to the Red AI approach. The Random Forest model had the optimal value M^* (i.e., the least inefficiency), followed closely by the Decision Tree model (denoted as m_1). The ANN model is ranked third (m_2), while the CNN model is fourth (m_3). Logistic Regression is the worst-performing model based on this strategy.

Figure 4, where $\alpha = 0.5$, model selection is based equally on performance and computation time. Here, both performance and computation time are weighted equally—this is the strategy adopted for this study. Although the Random Forest model remains the optimal model (M^*), closely followed by the Decision Tree model (m_1), Logistic Regression is now ranked third (m_2), displacing ANN to fourth place (m_3). CNN is regarded as the most inefficient model based on our Green AI selection approach.

In Table 3, we compare our work with other reviewed literature. Most of the authors fail to consider their models' training computation time, energy consumption, and carbon footprint. Some of the benchmarked studies even reported lower accuracy than the one selected in our study. Notably, our best-performing model—Random Forest—also aligns with the best-performing model in most of the other published work.

Table 2: Performance comparison of classification models

Model	Accuracy	Precision	Recall	F1-Score
ANN	0.9280	0.9287	0.9250	0.9259
CNN	0.9072	0.9058	0.9030	0.9021
LR	0.9034	0.9009	0.9021	0.9006
DT	0.9621	0.9635	0.9614	0.9619
RF	**0.9716**	**0.9713**	**0.9713**	**0.9712**

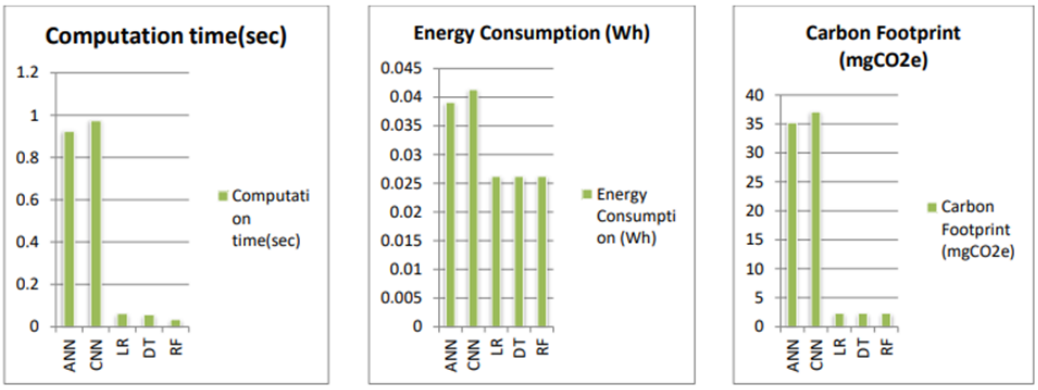

Figure 3: Computation time, energy consumption and carbon footprint of models

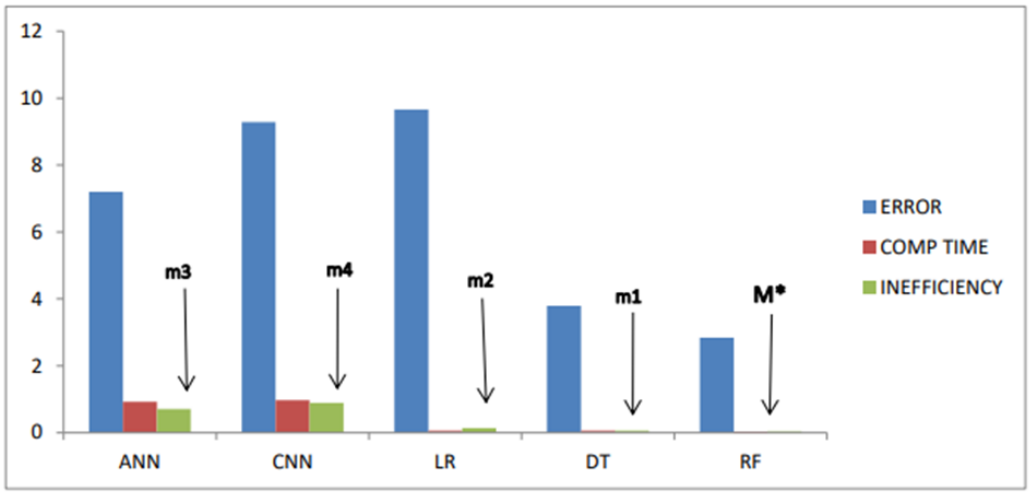

Figure 4: Model selection using the model inefficiency equation for $\alpha = 1$

5 Conclusion

Extant researchers in obesity risk prediction primarily followed the Red AI approach, focusing on improving the percentage of accuracy of learning algorithms. While this approach has been significantly successful within health informatics, it contradicts the commitments of many Government and organizations towards decarburizing energy systems within digital transformations. This paper presents an approach based on Green AI in the selection of an efficient model for obesity risk prediction from Physical conditions, physical description and eating habits of the participants. Although the proposed model can be generalized to unlimited variables, we considered only performance and computation time as the decision variables for selecting our optimal model. Our future research will address the issue of optimizing the hyper parameters of the selected models.

References

Adyasha, M. and Elaine, O. N. (2018). Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. *Health Informatics*, 1.

Alexander, B. G., Alfonso, C., and Aleiandro, S. G. (2023). The role of ai in drug discovery: challenges, opportunities and strategies. *Pharmaceuticals*, 16:891.

Anthony, L. F. W., Kanding, B., and Selvan, R. (2020). Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. *arXiv preprint arXiv:2007.03051*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2022). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

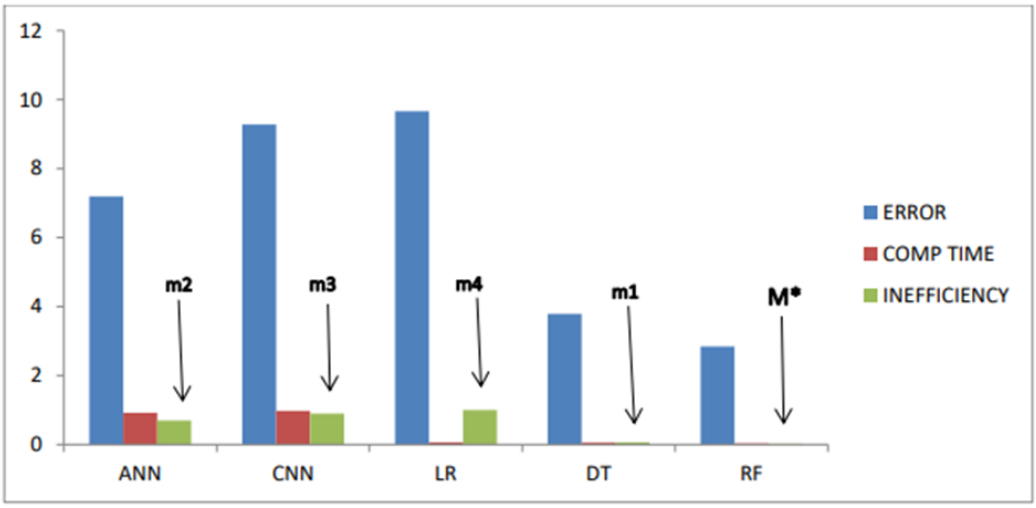

Figure 5: Model selection using the model inefficiency equation for α=0.5

Boluwaji, A. A., Kehinde, A. O., and Benjamin, S. A. (2023). Application of svm algorithm for early differential diagnosis of prostate cancer. *Data science and management*, 6:1–12.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Das, S., Iyer, R., and Natarajan, S. (2021). A clustering based selection framework for cost-aware and time test feature elicitation. In *8th ACM IKDDCODS and 26th COMAD*, pages 20–28.

David, P., Joseph, G., Quoc, L., Chen, L., Lluis-Miquel, M., Daniel, R., David, S., Maud, D., and Texierand, J. (2021). Carbon emissions and large neural network training. *Computers and Society*, 3.

Elias, R., Elen, R., Luis, N., Aneirson, S., and Fernando, M. (2021). Machine learning techniques to predict overweight or obesity. In *International conference on informatics and Data-driven medicine*, volume 4, page 3038.

Erion, G., Janizek, J. D., Hudelson, C., Utarnachitt, R. B., McCoy, A. M., Sayre, M. R., White, N. J., and Lee, S.-I. (2021). Coai: Cost-aware artificial intelligence for health care. *medRxiv*.

Ezenkwu, C. P., Stephen, B. U., Affiah, I., and Daniel, B. (2023). A green ai model selection strategy for computer-aided mpox detection. In *2023 IEEE Africon*.

Faria, F., Kazi, S. A. R., Ismail, J., and Tarek, H. (2021). A machine learning approach for obesity risk prediction. *Current research in behavioral sciences*, 2:1–9.

Fatemeh, S., Johanna, M., Niina, P., Markus, J., Nina, H., Terho, L., Jorma, V., Tanika, K., Changwei, L., Lydia, B., Laura, L. E., and Olli, T. R. (2016). Prediction of adulthood obesity using genetic and childhood clinical risk factors in the cardiovascular risk in young finns study. *American Heart Association Journals*, 10:1–9.

Table 3: Comparison of our work with existing studies on obesity prediction.

Author	Model Used	Dataset Used	Best Model	CT	EC	CF
Nirmala et al. (2022)	LR, AdaBoost, RF, DT, SVM	UCI obesity dataset: physical description, eating habits, and physical condition	RF, 99.36% accuracy	Not stated	Not stated	Not stated
Faria et al. (2021)	KNN, LR, SVM, Naïve Bayes, CART, RF, MLP, AdaBoost, Gradient Boosting	Questionnaire-based data (online/offline) covering dietary and physical condition	LR, 97.07% accuracy	Not stated	Not stated	Not stated
Elias et al. (2021)	DT, SVM, KNN, GNB, MLP, RF, Gradient Boosting, XGBoost	16-question survey on dietary habits and physical condition	RF, 77.69% accuracy	Not stated	Not stated	Not stated
Kapil et al. (2018)	Ensemble of LR, RF, Partial Least Squares	Data on physical description of participants	89.68% accuracy	Not stated	Not stated	Not stated
SSnekhalatha and Thanaraj (2021)	VGG-16, VGG-19, ResNet-50, DenseNet-121	647 thermal images	VGG-16, 92% accuracy	Not stated	Not stated	Not stated
Our Study	CNN, ANN, RF, LR, DT	UCI obesity dataset: physical description, eating habits, and physical condition	RF, 97.16% accuracy	0.03432 sec	2.36 $mgCO_2e$	2.62e-3

Fayroza, A. K. and Abbas, M. A. (2023). Diagnosis of diabetes using machine learning algorithm. *Materials today*, 80:3200–3203.

Hurby, A., Manson, J. E., Qi, L., Malik, V. S., Rimm, E. B., Sun, Q., Willet, W. C., and Hu, F. B. (2016). Determinants and consequences of obesity. *Public health*, 106:1656–1662.

Jordan, H., Sebastian, B., Arthur, M., Elena, B., Trevor, C., Eliza, R., Diego, L. C., and Lisa Anne, H. (2022). Training compute optimal large language models. *Computation and Language*.

Joseph, E. D., Mohammed, H., Jado, K., Micheal, Z., and Hateem, M. (2023). Initial experience with a novel robotic surgical system in abdominal surgery. *Journal of robotic surgery*, 17:841–846.

Kandar, J. H. and Praba, J. J. (2020). Artificial intelligence in medical diagnosis: methods, algorithms and applications. *Machine learning with health perspective*, 13:27–37.

Lannelongue, L., Grealey, J., and Inouye, M. (2021). Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science*, 8:56–60.

Maria, A. S., Sunder, R., and Satheesh, R. (2023). Obesity risk prediction using machine learning approach. In *2023 International*, pages 1–7.

Nirmala, K. D., Karthi, S., Krishna, M. N., Jayanthi, P., and Kiranbharath, K. (2022). Machine learning based adult obesity prediction. In *International conference on computer communication and informatics*, pages 25–27.

Organization, W. H. (2019). Fact sheet: obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (Retrieved on 18th May 2024).

Orsolya, K., Fiona, C. B., Robert, P., Erin, E. D., and Kelley, P. (2024). Using explainable machine learning and fit bit data to investigate predictors of adolescent obesity. *Scientific reports*, 14:12563.

Pritom, K. M., Kamrul, H. F., Bryan, A. N., and Lisaann, S. G. (2023). Childhood obesity based on single and multiple well-child visit data using machine learning classifiers. *Sensors*, 23:759.

Rohan, A., Andrew, M. D., Orhan, F., Melvin, J., Dmitry, L., Alexandre, P., Siamak, S., Emanuel, T., Paige, B., et al. (2023). Palm 2 technical report. Technical report, Computation and Language.

Schwartz, R., Dodge, J., Noah, A. S., and Oren, E. (2020). Green ai. *Communications of the ACM*, 63:54–63.

Silver, D., Schrittwieser, J., and Simonyan, K. (2017). Mastering the game of go without human knowledge. *Nature*.

Snekhalatha, U. and Thanaraj, S. K. (2021). Computer aided diagnosis of obesity based on thermal imaging using various convolutional neural networks. *Biomedical signal processing and control*, 63:102233.

Tatineni, S. (2022). Integrating ai, blockchain and cloud technologies for data management in healthcare. *Journal of Computer Engineering and Technology*, 5.

Teven, L. S., Angela, F., Christopher, A., Ellie, P., and Suzana, I. (2023). Bloom: A 176b-parameter open-access multilingual language model. Technical report, INRIA. HAL.SCIENCEWeb Portal.

Veronica, B., Laura, M., Brais, C., and Amparo, A. (2024). A review of green artificial intelligence: towards a more sustainable future. *Neurocomputing*, 599:128096.

Wei, Q., Chen, Y., Salimi, M., Denny, J. C., Mei, Q., Lasko, T. A., Chen, Q., Wu, S., Franklin, A., Cohen, T., et al. (2019). Cost-aware active learning for named entity recognition in clinical text. *Journal of the American Medical Informatics Association*, 26(11):1314–1322.

Xiaoyin, J., Zuojin, H., Shuihua, W., and Yudong, Z. (2023). Deep learning for medical image-based cancer diagnosis. *Cancers*, 15:3608.

Xie, Q., Luong, M. T., Hovy, E., and Le, Q. V. (2020). Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10687–10698.