FOREWORD

It is with immense pleasure and pride that I present to you Volume 6, Issue 2 of the Federal Polytechnic Ilaro - Journal of Pure and Applied Sciences (FEPI-JOPAS). As we continue to embark on a journey of academic excellence, innovation, and knowledge dissemination, this edition brings to the fore a selection of thought-provoking, interdisciplinary research that showcases the ingenuity and expertise of our scholarly community. This volume features five exceptional manuscripts, each contributing to the advancement of scientific understanding in diverse fields, ranging from applied technology and energy systems to plant-based biochemistry and predictive algorithms. These papers are the result of rigorous research and critical analysis, reflecting the depth of scholarship fostered at the Federal Polytechnic Ilaro and beyond.

The first manuscript, authored by Alabi and Ojo titled, *A Comparative Analysis of Extreme Gradient Boosting and Support Vector Regression for Modeling Benchmark Crude Oil Prices*, offers valuable insights into the application of machine learning algorithms in predicting one of the most volatile and essential global commodities—crude oil. By comparing the performance of two sophisticated modeling techniques, this study paves the way for more accurate forecasting in the energy sector. In second paper by Kumoye and Remi-Esan considered *Comparative Analysis of Phytochemical Constituents, Proximate Composition and Energy Value of Hibiscus sabdariffa L. Leaf and Petal Extracts*. The authors delve into the rich nutritional and medicinal potential of *Hibiscus sabdariffa*. This paper not only adds to the growing body of knowledge on natural products but also has significant implications for the health and wellness industry, as well as sustainable agricultural practices.

In the third paper by Ayodele and Sodeinde, the work titled *Student Academic Performance Prediction System Using Ensemble Algorithm*, explores the transformative potential of artificial intelligence in education. The authors have developed a system that predicts student performance based on multiple factors, providing a tool for improving academic outcomes and informing educational strategies. Paper four by Sanni and Onigbara is the Next on, *Development of a Crucible Furnace Fired with Liquefied Petroleum Gas - Butane Gas*, introduces an innovative approach to energy efficiency in industrial processes. This research presents a design for a crucible furnace that utilizes liquefied petroleum gas (LPG) as an alternative fuel, demonstrating how this can

contribute to a more sustainable and cost-effective solution for melting operations in the manufacturing sector. Finally, the article by Sanni and Onigbara titled *Design and Construction of Thermoelectric Generator Using Parabolic Trough Collector* addresses the urgent need for sustainable energy solutions. This paper outlines the design and construction of a thermoelectric generator powered by solar energy, contributing to the exploration of renewable energy technologies and their integration into real-world applications.

Each manuscript in this issue is a testament to the vibrant academic environment and wide readership scientific platform created by FEPI-JOPAS for prolific authors of internation repute. I commend the authors for their dedication to advancing knowledge in their respective fields. As we continue to foster research and innovation, it is our hope that these works will inspire further inquiry, discussion, and development, both within our institution and in the global academic community.

I extend my sincere gratitude to the reviewers, editorial team, and contributors who have made this publication possible. Without their tireless efforts, this journal would not be the prestigious platform it has become. As we look ahead to future editions, let us continue to uphold the values of academic rigor, curiosity, and the pursuit of knowledge. Thank you for your continued support of the Federal Polytechnic Ilaro - Journal of Pure and Applied Sciences. I am looking forward to receiving your manuscripts for subsequent publications. This journal is indexed in Arican Journal Oline (AJOL) and can be access online at https://www.ajol.info/index.php/fepijopas. You can also visit our website (https://fepi-jopas.federalpolyilaro.edu.ng/) for more information or contact us via e-mail at *fepi.jopas@federalpolyilaro.edu.ng*
Thank you and best regards.

Prof. Olayinka O. AJANI
Editor-in-Chief
Federal Polytechnic Ilaro - Journal of Pure and Applied Sciences (FEPI-JOPAS)

A COMPARATIVE ANALYSIS OF EXTREME GRADIENT BOOSTING AND SUPPORT VECTOR REGRESSION FOR MODELING BENCHMARK CRUDE OIL PRICES

ALABI, Nurudeen O. & OJO, Gabriel O.
Department of Mathematics & Statistics, Federal Polytechnic, Ilaro, Ogun State.
Corresponding author email: nurudeen.alabi@federalpolyilaro.edu.ng

ABSTRACT
Organization of Petroleum Exporting Countries (OPEC) and non-OPEC supply, oil balance, oil demand by Organization for Economic Cooperation and Development (OECD) and non-OECD members, money market managers' long positions, US consumer price index and spot prices of crude oils like New York Mercantile Exchange West Texas Intermediate (NYMEX WTI), Intercontinental Exchange (ICE) Brent, OPEC Reference Basket (ORB), and other crude oils are basic elements driving the patterns seen in the market pricing of crude oils. Data between 2008 and 2022 were obtained for this study from OPEC Monthly Oil Market Reports. This research evaluates the performance of two machine learning models, Support Vector Regression (SVR) and Extreme Gradient Boosting (XGBoost), in predicting crude oil prices for three major benchmarks: OPEC Reference Basket (ORB), NYMEX WTI, and ICE Brent. Using key performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2), the study highlights the strengths and weaknesses of each model in both stable and volatile market conditions. SVR shows strong predictive accuracy, particularly for ICE Brent, but struggles with price volatility in the ORB and NYMEX WTI datasets. XGBoost is more robust in handling volatility and non-linear relationships. The findings have important economic implications for market participants, suggesting that while SVR is suited for stable pricing environments, XGBoost is better equipped to handle the unpredictability of more volatile markets.

Keywords: Ensemble Algorithms, XGBoosting, ICE Brent, NYMEX WTI, ORB, OPEC, OECD Support Vector Regression

1.0 INTRODUCTION

Ensemble algorithms are sophisticated techniques that combine the predictions of multiple base models to improve overall forecasting accuracy and robustness. They capture complex nonlinear relationships, handle high-dimensional data, and effectively deal with noisy and dynamic patterns inherent in data. These algorithms leverage the diversity of multiple models to mitigate overfitting and improve generalisation performance. By evaluating their accuracy, robustness, and computational efficiency, we seek to gain insights into the most effective algorithm for this specific application, which can inform decision-making processes, risk management strategies, and investment decisions in the global crude oil market.

The global crude oil market is highly robust, with relatively few producers. In 2021, the total world crude oil demand averaged 96.44 million barrels per day (mb/d), with OECD countries, non-OECD and OPEC-13 providing 29.56, 31.9, and 28.9 mb/d respectively. Crude oil prices are determined by the market forces of demand and supply, with higher demand in rapidly developing countries driving prices in the upward direction. Additionally, speculative activities of money managers through their total futures and net-long positions in crude oil futures contribute significantly to price volatility. The exchange value of the US dollar also plays an important role in the fluctuations of crude oil prices globally. Seven fundamental factors are responsible for crude oil price fluctuations. Financial markets, OPEC and non-OPEC crude oil supply, crude oil balance, oil demand by OECD and non-OECD members, and spot prices of crude oils such as NYMEX WTI, ICE Brent, EIA's Imported Refiner Acquisition Cost (IRAC), OPEC Reference Basket (ORB) and other crude oils determine the differences in other crude oils globally. ExxonMobil (2021) claims that worldwide crude oils can be divided into light, medium, and heavy categories based on density (API gravity). Sweet crude oils have a sulphur level between zero percent and 0.59 percent of the weight, while sour crude oils range from 0.62 percent to 3.85 percent by weight. The new OPEC Reference Basket contains thirteen crude oils, including Saharan Blend (Algeria), Iran Heavy (Iran), Zafiro, Basrah Medium (Iraq), Girrasol, Kuwait Export (Kuwait), Es Sider (Libya), Bonny Light (Nigeria), Djeno, Merey, Arab Light (Saudi Arabia), Murban (UAE), and Rabi Light. The American Petroleum Institute (API) gravity and Sulphur content of the WTI are respectively 39.6 degrees and 0.24 percent, while Brent has an API gravity of 38.3 degrees and contains 0.37 percent sulfur. The NYMEX futures price for crude oil is the value of a futures contract which depends purely on market forces on trading one thousand barrels of the light-sweet crude oil at a particular time. The IRAC represents the volume-weighted average cost of all heavy and light crude oils imported into the United States during a specific period. To ensure stable, open, and predictable oil markets, producers and consumers must act responsibly regarding supply and demand security. The behavior of crude oil. The behavior of crude oil market prices on a worldwide scale is influenced by several factors, as was previously mentioned. The directional relationships and interactions between these variables and spot prices are shown in Figure 1.

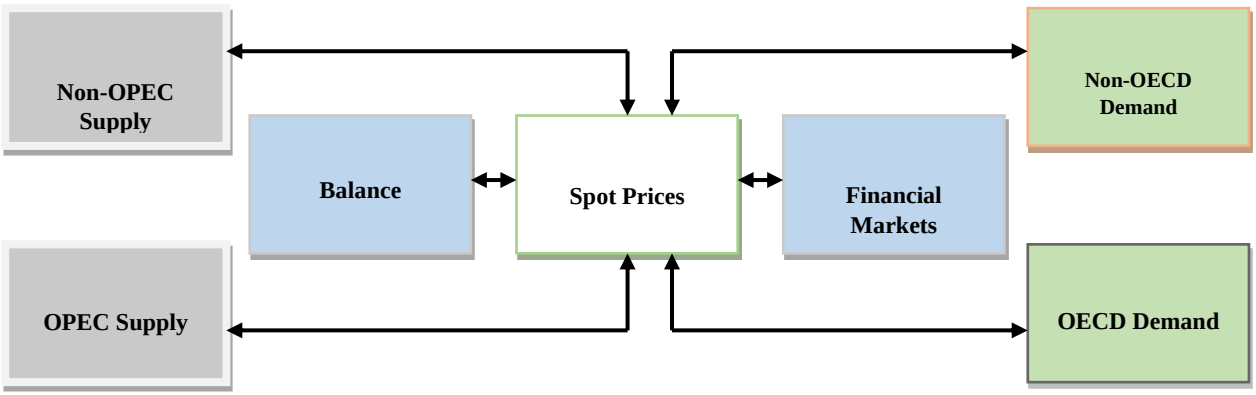


Figure 1: Determinants of Crude Oil Prices and their Interactions
Source: US Energy Information Administration (EIA, 2022)

These links serve as the foundation for this study, which aims to establish two objectives

- Compare the performances of three widely used ensemble machine learning algorithms such as extreme gradient boosting, and support vector regression on the crude oil datasets.
- To determine the most relevant determinants for modeling the benchmark crude oil prices in the global crude oil market.

Researchers from all around the world have conducted several studies on the modeling and forecasting of crude oil prices. For instance, Lu *et al*. (2021) created a framework for selecting and forecasting the key variables that affect the price of crude oil. The study used a Long-Short Term Memory Network, Spike-Slab LASSO, Bayesian model average (BMA), and an elastic net regularized generalized linear model (GLMNCI) (LSTM). Based on a random walk, six forecasting methods—Walvet neural networks (WNN), Elman neural networks (ENW), neutral ELM networks, autoregressive integrated moving average models (ARIMA), and generalized regression neural network model—were compared (GRNN). According to the analysis, the LSTM has the greatest precision. In a different work, Bai, Yuying, and Shonyang (2021) generalized integral-valued forecasts to include uncertainty and variability in the price of crude oil and presented a two-stage forecasting method based on interval-valued time series. This process outperformed some benchmark models when compared, it was discovered. Krzysztof and Liu conducted a study that is comparable in that it deals with the significant problem of uncertainty that is present in time series analysis in 2021. The study compared several time-varying VARs that included geopolitical risk as an endogenous component. They concluded that real prices for crude oil vary significantly over time.

Dondukova and Lin (2021) used the Euler-Mamyama scheme as an approximation of the Heston model to model the volatility of WTI and Brent. In particular, it was discovered that the stochastic volatility model outperformed all GARCH models when they were examined via the RMSE and MAE. Waqas et al. (2018) developed an ensemble empirical model decomposition (EEMD) To anticipate the price of crude oil. It has been demonstrated that this alternative method to traditional econometric methods aids in dealing with non-stationarity and non-linearity of time series, particularly crude oil prices. Wajdi and Dawud (2018) fitted a crude oil price regression on its key variables chosen using PCA. These are the geopolitical and fundamental factors. Analysis revealed that the most crucial factors affecting crude oil prices are fundamentals and the responsibilities of OPEC members. Wassin and Ibrahim (2018) analyse the linear and non-linear regression models to examine the relationships between crude oil prices and stocks. They used the informational value of oil demand and the link between crude oil and equities prices to fit the linear models. Similarly, a non-linear model was fitted using fuzzy neural networks and genetic algorithms. In terms of the accuracy of the statistical forecasts made outside of the sample, some of the fitted linear models performed the best.

By combining prior knowledge from the current and anticipated structure of the oil markets with the Bayesian technique, Chul-Yoon and Sung-Yoon (2016) predicted long-term crude oil prices. The model's stated factors for determining crude oil prices include factors including global oil demand and supply, economic conditions, upstream costs, and geographic occurrences. The OLS and neural network were contrasted using this model. The fitted model was found to perform better on the forecasting performance test alone than these two models. Merk (2016) used the daily return of crude oil prices to examine the relationship between global financial crises and volatility. They concluded that crude oil prices are extremely volatile and substantially respond to shocks from the global financial crisis after fitting volatility models such as the APGARCH and FIAPGARCH models. To estimate the price of crude oil, Ani and Rubaidah (2014) devised a hybrid model that combines wavelet and multiple linear regression. PCA and the Mallat wavelet transform are used in this work. To choose the best model for the multiple linear regression on the WTI, they used Particle Swarm Optimisation (PSO). The WMLR outperformed the ARIMA, MLR, and GARCH models when they were put side by side.

2.0 MATERIALS AND METHODS

2.1 Materials

The data used in this study includes 176 monthly observations between 2008 and 2022 of the prices of a barrel of ORB (orb_t), WTI (wti_t), and Brent ($brent_t$) in US dollars, as well as data on OPEC supply ($opec\_supply_t$), non-OPEC supply ($nonopec\_supply_t$), money market managers' net long positions ($mo\_net\_long_t$), OECD demand ($oecddemand_t$), non-OECD demand ($nonoecddemand_t$), the oil balance ($balance_t$), and US Consumer Price Index (us_cpi_t). The source of data used is the OPEC Monthly Oil Market Reports.

2.2 Methods

2.2.1 Extreme Boosting Machine Learning Algorithm

Extreme Boosting is a popular ensemble machine learning algorithm used in various domains, including time series forecasting. It encompasses implementations like XGBoost, LightGBM, and CatBoost, each with unique features and optimisations. The algorithm follows a boosting concept and ensemble methodology. These algorithms allow base learners, typically decision trees, to be sequentially trained to minimize the overall prediction error by emphasizing difficult instances. The algorithm iteratively adds decision trees to the ensemble, updating the predictions and minimising the loss function. Parameter tuning and model optimisation are crucial for effective application. Vital considerations include the learning rate, tree depth and complexity, and regularisation parameters. The formula for updating the ensemble model involves combining the previous predictions with the contribution of the newly added tree. Each iteration minimises a specific loss function concerning the ensemble's errors. The specific implementation and formulas may vary depending on the Extreme Boosting algorithm used. Additional optimisations and enhancements are introduced in algorithms like *XGBoost*, *LightGBM*, and *CatBoost* to improve training speed, accuracy, and handling of different data types. In Extreme Boosting, the algorithm aims to minimise a specific loss function during the training process. The choice of loss function depends on the task at hand, such as mean squared error (MSE) or mean absolute error (MAE) for regression problems. During each iteration of the boosting process, a new decision tree is added to the ensemble to reduce the overall loss. The tree is constructed to minimise the chosen loss function concerning the residuals or errors made by the ensemble model on

the previous iterations. By iteratively updating the ensemble and adding trees that target the remaining errors, the algorithm gradually improves the overall model performance and reduces the prediction errors. This process continues until a predefined stopping criterion is met, such as reaching a maximum number of iterations or achieving satisfactory performance. The iterative nature of Extreme Boosting, combined with the minimisation of the loss function, allows the algorithm to effectively learn complex patterns and make accurate predictions.

Mathematical Aspect of Extreme Gradient Boosting Ensemble

Given a dataset with *n* observations and *m* features

$$D = \left\{ \left(x_i, y_i \right) \right\} \left(\left| D \right| = n, x_i \in R^m, y_i \in R \right) \quad (1)$$

Where we predicted the value of the ensemble tree model using *K* additive functions expressed as

$$\hat{y}_i = \sum_{k=1}^{K} f_k \left(x_i \right), f_k \in F \quad (2)$$

Where *F* is the regression tree space computed as

$$F = \left\{ f\left(x \right) = \omega_x \left(x \right) \right\} \left(q : R^m \to T, \omega_q \in R^T \right) \quad (3)$$

q represents the structure of each tree, *T* is the number of nodes (leaves) in the tree and f_k is a function that corresponds to an independent tree structure *q* and leaf weights ω. Given a training dataset with input features *x* and target variable *y*, and an ensemble model represented by *f* (*x*), the goal is to minimise the loss function

$$L^{(t)} = \sum_{i=1}^{n} l\left(\left(y_i, \hat{y}_j^{(t-1)} \right) + f_t \left(x_i \right) \right) + \Omega\left(f_k \right) \quad (4)$$

where *l* is a differentiable convex objective function to calculate the error between predicted and measured values; y_i and y_i are regulated and predicted values, respectively; *t* shows the repetitions to minimise the errors; and **Ω** is the complexity penalised with the regression tree functions:

$$\Omega\left(f_k \right) = \gamma T + \frac{1}{2} \lambda \left\| \omega \right\|^2, \quad (5)$$

ω is the vector of the score for the blades, and *c* is the minimal loss required for the further isolation of a blade node. λ is the regularisation function. In addition, *c* and λ are parameters that can control the complexity of the tree, and the regularisation term helps to avoid overfitting by smoothening the final learned weights. Taylor expansion is applied to the objective function to simplify it further as

$$F = \sum_{i=1}^{m} \left[f_t \left(x_i \right) g_i + \frac{1}{2} \left(f_t \left(x_i \right) \right)^2 h_i \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \quad (6)$$

where g_i and h_i are the first and second derivatives obtained on the loss function, respectively. Specifically,

$$g_i = \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)} \right) \quad (7)$$

$$g_i = \sum_{i=1}^{n} l\left(y_i, y_i^{(t-1)} \right) + \sum_{k=1}^{t-1} \Omega\left(f_k \right) \quad (8)$$

$$h_i = \partial^2_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)} \right) \quad (9)$$

The optimal objective function in equation (6) as a function of the 1st and 2nd derivatives is

$$L^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T + C \quad (10)$$

$$G_j = \sum_{i \in I_j} g_i$$

and (11)

$$H_j = \sum_{i \in I_j} h_i$$

The *XGBoosting* process involves a series of steps where a new decision tree is added to the group to reduce errors. This new decision tree, denoted as *g*, is trained to make predictions that are closer to the actual target variable, *y*, compared to the current group of models. After training, this new tree becomes part of the ensemble. This cycle repeats, with each new tree trained to improve predictions based on the current ensemble. The ultimate goal of Extreme Gradient Boosting is to create an ensemble model, *f*(*x*), that minimizes the total errors across all these steps. By minimising errors at each step and updating the ensemble accordingly, *XGBoost* gradually enhances the model's predictive accuracy, leading to a more precise final model. In this study, the researchers compared *XGBoost* with other ensemble methods like bagging, random forest, and support vector regression, as proposed in the existing literature.

2.2.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a non-linear regression method that can capture non-linear relationships between the input features and the target variable. It is based on the concept of support vector machines (SVMs), which were originally developed for classification tasks. SVR uses a kernel function to map the input features into a higher-dimensional space where the relationship between the features and the target variable may become linear. In this transformed space, SVR attempts to find a linear regression model that fits the data while maximising the margin (the distance between the regression line and the closest data points, which are the support vectors). The choice of kernel function, such as the radial basis function (RBF) kernel, allows SVR to model complex, non-linear relationships. It has the flexibility to capture both linear and non-linear relationships in the data, depending on the choice of kernel and model parameters. This makes SVR a powerful tool for regression tasks with complex data patterns. Support Vector Regression (SVR), support vectors, and weights play important roles in understanding the model. The support vectors are data points from the training dataset that have the most influence on the SVR model. These are the data points that are closest to the SVR's decision boundary (the hyperplane) and are used to define the margin. They are the points for which the model's prediction is either equal to the target value or falls within a certain distance (the margin) from the target value. They play a crucial role in determining the SVR model's performance and are the ones responsible for shaping the model's regression line. Hence, Support vectors are the critical data points that have the most influence on the model's prediction. In SVR, weights represent the importance or contribution of each support vector to the model's prediction. Each support vector is associated with a weight that signifies its significance in shaping the regression model. Weights are essentially the coefficients of the support vectors in the SVR equation. Support vectors with larger weights have a more substantial impact on the model's decision boundary and, consequently, on its predictions. Vector weights provide

information about the importance of different data points in the SVR model, helping to understand which data points are the most influential in making predictions. Support vectors are the data points closest to the decision boundary, and they have a direct impact on the SVR model's predictions. The weights associated with these support vectors indicate their relative importance in shaping the model. By examining support vectors and their weights, we gain insights into which data points are driving the SVR model's predictions and the significance of each of these points in the regression analysis. By analysing the weights, allows us to pinpoint the most influential data points and focus on their characteristics, which are valuable for understanding the key factors driving the SVR model's performance and enhancing its predictive capabilities. High-level SVR algorithm involves data pre-processing by standardizing or normalising the input features to ensure they are on a similar scale, choosing a kernel function (linear or radial basis function) to transform the input features and formulation of the SVR optimisation problem. The objective is to find a hyperplane that has the maximum margin while minimising the error (the difference between the predicted and actual values). The introduction of a soft margin allows some instances to be within the margin or even on the wrong side of the hyperplane. This is done to handle cases where a strict margin might not be achievable due to noise or outliers. We apply cross-validation to control the complexity of the model and prevent overfitting. We define a loss function that penalises errors in prediction. Common loss functions include epsilon-insensitive loss or mean squared error. Solve the dual problem using optimisation techniques. The solution provides the support vectors and their corresponding weights. Use the obtained support vectors and weights to make predictions on new, unseen data. The basic idea is to find a hyperplane that not only fits the data well but also has a maximum margin. SVR is especially useful when dealing with non-linear relationships between input features and the target variable, thanks to the kernel trick that allows for implicit mapping to higher-dimensional feature spaces.

To determine the support vectors and their associated weights, we formulated a dual SVR problem which was solved using the Lagrangian dual optimisation. Generally, the dual problem is a maximisation problem. The Lagrangian function was formulated by combining the objective function of the primal problem with the constraints each multiplied by the Lagrange multiplier α_i. The Lagrangian for linear and radial basis function kernels are given as follows.

Radial Basis Function Kernel (RBF)

The primal problem for SVR with an RBF kernel $K(X_i, X_j)$ is given as

$$Minimize\ \frac{1}{2}\|\boldsymbol{w}\|^2 + C\ \sum_1^n(\xi_i + \xi_i^*) \qquad (12)$$

Subject to constraints:

$$\begin{aligned} y_i - f(x_i) &\leq \varepsilon + \xi_i \\ f(x_i) - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0,\ \forall\, i = 1, 2, 3, \ldots\ldots\ldots, n \end{aligned} \qquad (13)$$

ε is the acceptable error, ξ_i, ξ_i^* are the slacks and $f(x_i)$ are the predicted output using the RBF kernel. C in the minimisation problem is the regularisation parameter. The Lagrangian for the RBF kernel is

$$L(\mathbf{w}, b, \alpha, \alpha^*, \xi_i, \xi_i^*) = \frac{1}{2}\|w\|^2 + C\sum_1^n(\xi_i + \xi_i^*) + \sum_1^n \alpha_i^*\left(\xi_i^* - \varepsilon - y_i - f(x_i)\right) + \sum_1^n \alpha_i \left(\xi_i - \varepsilon + y_i - f(x_i)\right) \qquad (14)$$

$f(x_i)$ is the output of the RBF expressed as:

$$f(x_i) = \sum_{j=1}^n \alpha_j(\mathbf{x_j})K(\mathbf{x_j}, \mathbf{x_i}) + b \qquad (15)$$

where $K(\mathbf{x_j}, \mathbf{x_i})$ is the RBF kernel. The dual problem involves maximising the Lagrangian for the Lagrange multipliers α, α^* subject to the constraints $\alpha, \alpha^* \geq 0$ for all i. The RBF kernel function and its parameters play a crucial role in the dual optimisation problem.

Linear Kernel

Just like the RBF kernel, the primal for the dual problem in equation (1) for the linear kernel is subject to the following constraints

$$\begin{aligned} y_i - (\mathbf{w}\cdot\mathbf{x_i} + b) &\leq \varepsilon + \xi_i \\ (\mathbf{w}\cdot\mathbf{x_i} + b) - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0,\ \forall\, i = 1, 2, 3, \ldots\ldots\ldots, n \end{aligned} \qquad (16)$$

The Lagrangian for the linear kernel is

$$L(\mathbf{w}, b, \alpha, \alpha^*, \xi_i, \xi_i^*) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_1^n(\xi_i + \xi_i^*) + \sum_1^n \alpha_i^*(\xi_i^* - \varepsilon - y_i + \mathbf{w}\cdot\mathbf{x_i} + b) + \sum_1^n \alpha_i \left(\xi_i - \varepsilon + y_i - \mathbf{w}\cdot\mathbf{x_i} - b\right) \qquad (17)$$

The estimation process is one of finding a hyperplane with a maximum margin while minimising the error which involves solving the constrained optimisation primal problem. The slack variables allow for the introduction of soft margins which allows some instances to be within the margin or on the wrong side of the hyperplane. We employed the epsilon insensitive loss function and the mean square error to penalise errors in prediction. The solution of the Lagrangian dual optimisation provides the support vectors and their associated weights which were used to define the function $f(x)$ for predicting the target variable.

3.0 RESULTS AND DISCUSSION

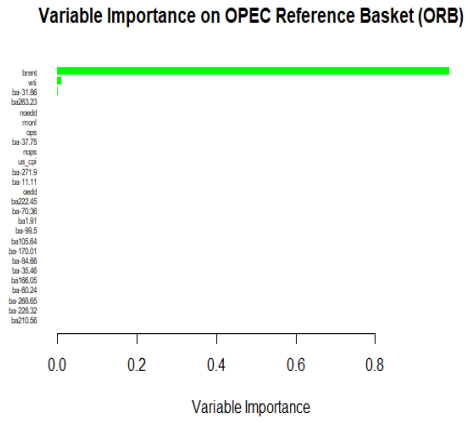
We present the analysis and results derived from the application of the two discussed machine learning models to forecast benchmark crude oil prices, including ORB, Brent, and WTI. The performance of the Support Vector Regression (SVR) and XGBoost was evaluated using key metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 for both training and testing datasets. By comparing these metrics, we aim to establish the models' effectiveness in predicting crude oil prices across different datasets, providing insights into their accuracy and generalisation capabilities. The following are the detailed findings and visual comparisons, leading to a thorough assessment of each model's predictive performance.

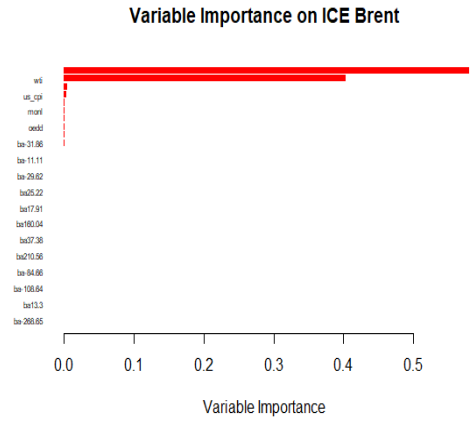
Table 1: Summary of Training on the XGBoost models with hyperparameters

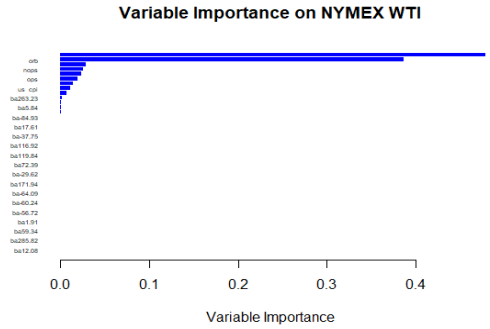
model	nround	max_depth	eta	gamma	colsample_bytree	min_child_weight	subsample
xgboost_wti	100	3	0.1	0.1	0.6	1	0.8
xgboost_brent	100	3	0.1	0.2	0.6	1	0.8
xgboost_orb	100	3	0.1	0.1	1	1	1

Three different models on ORB, Brent and WTI were built using stringent parameters. The training process for XGBoost models on the three oil benchmarks is outlined in Table 1. For each model, the tuning parameters *nrounds* and *max_depth* were held constant at 100 and 3, respectively. The models were optimised using Root Mean Square Error (RMSE) to identify the best-performing parameters. Common final values across the models include a learning rate (eta) of 0.1, a *min_child weight* of 1, and subsampling (subsample) values set at 0.8 for WTI and Brent and 1 for ORB. The specific differences in the models come from gamma, which regulates the minimum loss reduction for splitting nodes. It was set at 0.2 for WTI, 0.1 for Brent, and ORB. Additionally, *colsample_ bytree* (the proportion of features used for each tree) was 0.6 for both WTI and Brent but set to 1 (all features) for ORB. Each model was trained with 182 features, using 100 iterations (boosting rounds), and employed the reg: squared error objective to minimise squared error during regression. The table presents hyperparameters for three XGBoost models (*xgboost_vti*, *xgboost_brent*, and *xgboost_orb*). All models undergo 100 boosting rounds with a maximum tree depth of 3, ensuring the trees are shallow to control overfitting. The learning rate (eta) is set to 0.1 for all, balancing model performance and convergence speed. The minimum loss reduction to make a split (gamma) is set at 0.1 for *xgboost_vti* and *xgboost_orb*, allowing moderate splits, while *xgboost_brent* has a stricter requirement with a gamma of 0.2, limiting tree complexity unless significant gains are achieved. For feature sampling (*colsample_bytree*), *xgboost_vti* and xgboost_brent use 60% of the features for each tree, while xgboost_orb uses all features (100%). The minimum sum of instance weights needed to create a new leaf node (*min_child_weight*) is set to 1 across all models, ensuring trees do not split unless meaningful. Finally, *xgboost_vti* and *xgboost_brent* use 80% of the training data for each tree (subsample = 0.8), while *xgboost_orb* uses all the data (subsample = 1), potentially increasing the risk of overfitting for the latter.

Figure 2: Variable Importance on ORB, ICE Brent and NYMEX WTI







The three variable importance plots illustrate how different variables contribute to the prediction models for oil benchmarks: OPEC Reference Basket (ORB), ICE Brent, and NYMEX WTI. In the ORB model, Brent price overwhelmingly dominates the predictions with an importance close to 0.8, while the other variables contribute very little. This suggests that the model heavily relies on an ICE Brent to make accurate predictions for the ORB. For the ICE Brent model, a similar trend is observed, where one variable WTI shows the highest importance at around 0.5. However, the influence of other variables is minimal, indicating that the Brent model also focuses on NYMEX WTI as the key predictor for most of its forecasting accuracy. In contrast, the NYMEX WTI model exhibits a broader distribution of variable importance. While one variable (ORB) still stands out with an importance around 0.4, other variables play a more significant role compared to the other two models. This suggests that the WTI model captures a wider range of factors in making predictions, making it more dependent on multiple variables.

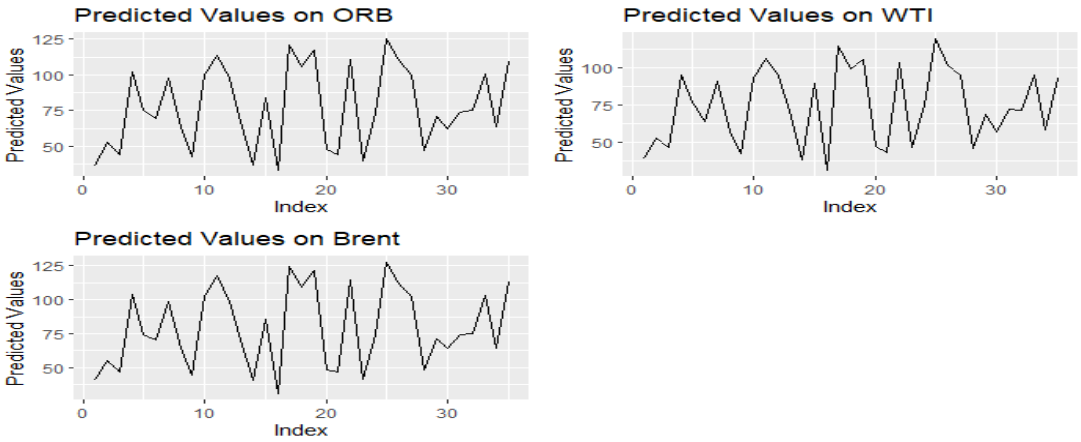


Figure 3: Plot on Predicted Values on ORB, WTI and Brent

The predicted values generated by the XGBoost models for the OPEC Reference Basket (ORB), NYMEX WTI, and ICE Brent benchmarks are presented in Figure 3. Across all three benchmarks, the predictions show a high degree of variability, with values fluctuating between 50 and 125. This suggests that the models capture the dynamic changes in the oil markets, reflecting how volatile and unpredictable the price movements can be. For the ORB and Brent benchmarks, the predictions display more extreme fluctuations, reaching up to 125, with noticeable peaks and troughs throughout the index range. This highlights that the models are sensitive to shifts in the market, perhaps responding to key influencing factors like supply and demand changes, geopolitical events, or production adjustments. The WTI model shows similar oscillations but within a slightly narrower range, peaking around 100. While the predicted values still show sharp changes, the range indicates that the WTI model may be capturing slightly less variability compared to the ORB and Brent models. Overall, all three models provide predictions that align with the complex and volatile nature of the oil price benchmarks they represent.

We carried out training on three SVR models based on a linear kernel and epsilon regression. Epsilon regression allows the model to ignore small errors within a specified margin (epsilon). The main parameters listed include cost (C), which controls the trade-off between maximising the margin and minimising error, and epsilon, which defines how much deviation from the actual values is tolerated. All models use a linear kernel, meaning the relationship between the input features and the target variable is assumed to be linear. The WTI model has a cost of 1, which strikes a balance between penalising errors and maximising the margin. The epsilon value is set at 0.1, allowing some tolerance for prediction errors. With 76 support vectors, this model has fewer data points influencing the decision boundary, suggesting a simpler model with a relatively straightforward decision margin. The higher cost in this model likely leads to better performance but with a risk of overfitting. The ORB and Brent models both have a lower cost (C = 0.1), meaning they tolerate more misclassifications in favour of a wider margin. They also have a smaller epsilon (0.01), making them more sensitive to small prediction errors. These models have more support vectors (118 for ORB and 116 for Brent), indicating a more complex decision boundary. The higher number of support vectors suggests that more data points are influencing the model, possibly due to the lower cost and the focus on minimising even small errors. We evaluated the performance of the Support Vector Regression (SVR) model in predicting crude oil prices for three major benchmarks: OPEC Reference Basket (ORB), NYMEX West Texas Intermediate (WTI), and ICE Brent. Each benchmark represents a critical component of the global oil market, and accurate price predictions are essential for risk management, market analysis, and strategic decision-making. By analysing major performance metrics such as MSE, RMSE, and R^2, we assess the predictive accuracy of the SVR model across both training and testing datasets. This analysis aims to highlight the model's strengths and potential areas for improvement in forecasting oil prices for these three critical benchmarks.

Table 2: Summary of performance metrics on SVR models using the training and test datasets

	Training			Testing		
Metric	ORB	Brent	WTI	ORB	Brent	WTI
MSE	2.71	1.126	5.593	1.502	1.501	42.72
RMSE	1.646	1.061	2.365	1.225	1.225	6.536
R^2	0.995	0.999	0.989	0.996	0.998	0.944

Table 2 presents performance metrics for the SVR models. The table is divided into training and testing sets for each dataset. While the

model performs well on the ORB and Brent datasets, it struggles to generalise to the WTI dataset. This implies that SVR is not well suited to the WTI dataset. This is also shown in Figure 4.

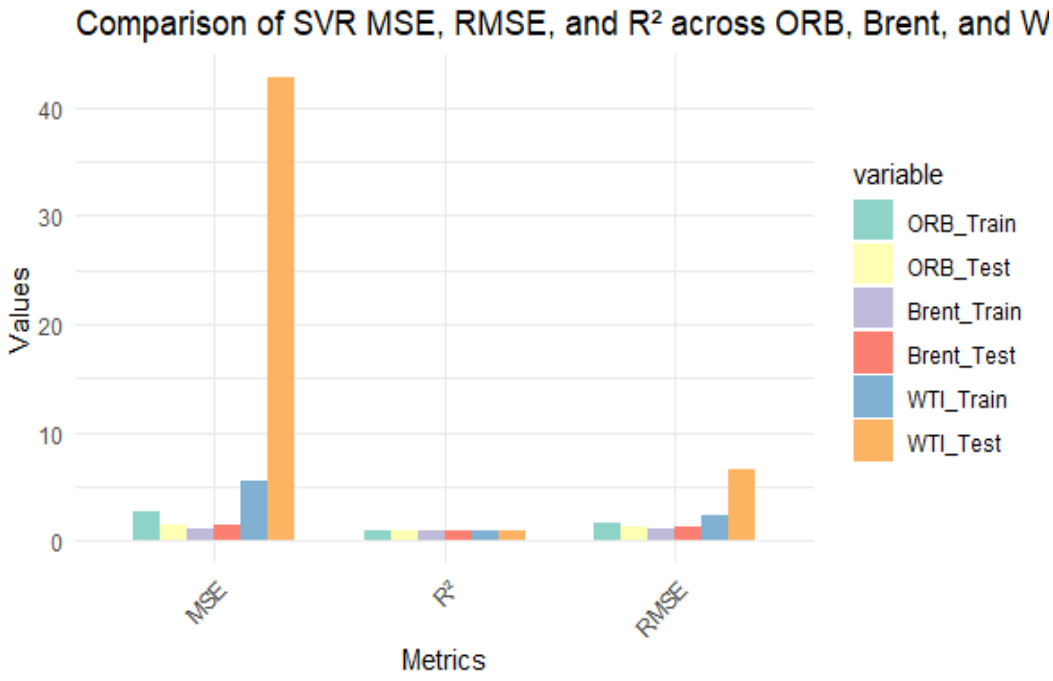


Figure 4: Bar Chart on Comparison of MSE, RMSE, and R^2 across ORB, Brent and WTI SVR models

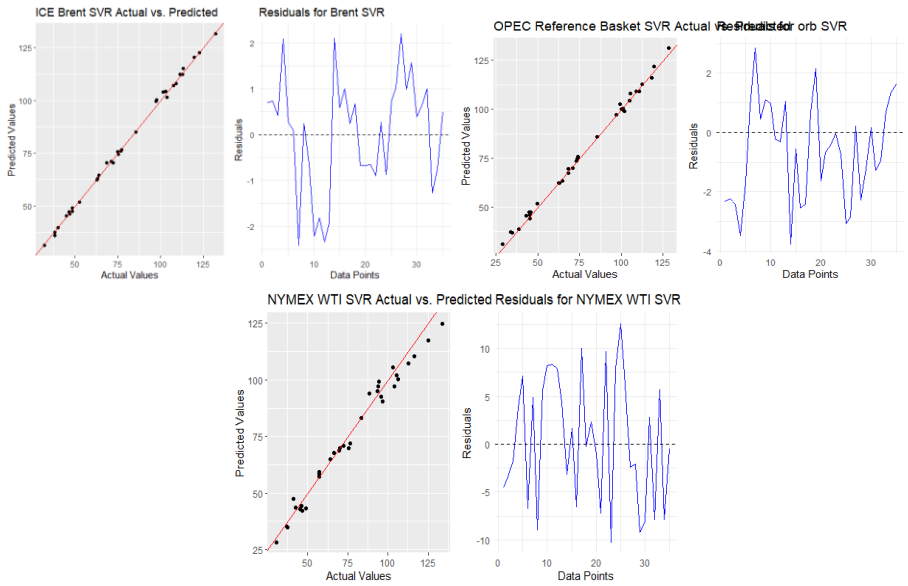


Figure 5: Plots on Predicted Values of SVR on ORB, NYMEX WTI and ICE Brent

For the OPEC Reference Basket (ORB), the actual versus predicted plot shows a good alignment between the SVR model's predictions and actual prices, with data points closely following the diagonal line (Figure 5). This suggests that the SVR model effectively captures the overall trends in ORB pricing. However, the residual plot reveals significant volatility, with wide fluctuations above and below the zero line. This indicates that while the model generally tracks price trends, it struggles to maintain consistent accuracy, especially during periods of price shifts, leading to larger prediction errors. For ICE Brent, the actual versus predicted plot similarly shows a strong alignment between predicted and actual values, indicating good model performance in forecasting Brent crude prices. The residual plot shows less volatility than the ORB, suggesting that the SVR model is more accurate for Brent pricing. The smaller fluctuations in residuals imply that the model is better at capturing Brent price movements, though some minor prediction errors remain.

In the WTI plots, the actual versus predicted values reveal that the SVR model performs quite well for NYMEX WTI prices. Most data points lie close to the diagonal line, indicating that the predicted values are fairly aligned with the actual values. This suggests that the SVR model is capable of capturing the general price trends in the WTI market. However, there are a few deviations, especially at the extreme values, implying that the model may have difficulty accurately predicting outliers or more volatile price changes. The residual plot shows the difference between the actual and predicted values. Residuals fluctuate significantly, indicating some instability in the model's predictions. The presence of positive and negative residuals means that the model occasionally underestimates or overestimates prices, with larger fluctuations seen in some data points. This suggests that while the model works reasonably well, it struggles to consistently predict prices with precision, particularly during periods of high volatility.

The SVR model performs well in predicting both ORB and Brent crude prices, with accurate alignment between actual and predicted values in both cases. However, the ORB residuals exhibit greater volatility, suggesting more significant prediction errors for this benchmark compared to Brent. For Brent, the smaller residual fluctuations imply more stable and reliable forecasts, which benefits market participants involved in Brent-linked contracts. Economically, better prediction accuracy for Brent means more precise hedging and risk management strategies, while the volatility in ORB predictions could increase uncertainty for stakeholders relying on ORB forecasts, leading to potential financial miscalculations.

The performance of XGBoost and SVR models using three metrics MSE, R^2, and RMSE are compared in Table 3 and Figure 4. Generally, the XGBoost consistently underperforms SVR across all datasets and metrics except the NYMEX WTI dataset, demonstrating lower prediction accuracy. Specifically, SVR achieves lower MSE and RMSE values, indicating better prediction accuracy. SVR's dominance is less pronounced in the WTI dataset. The superior performance of SVR over XGBoost in predicting crude oil prices has significant economic implications.

Table 3: Summary of Performance Metrics on XGBoost models using the training and test datasets

Metrics	XGBoost_ORB	XGBoost_Brent	XGBoost_WTI	SVR_ORB	SVR_Brent	SVR_WTI
MSE	3.066	2.901	2.896	1.502	1.501	42.71
RMSE	1.751	1.703	1.701	1.225	1.225	6.536
R^2	0.994	0.995	0.996	0.996	0.998	0.944

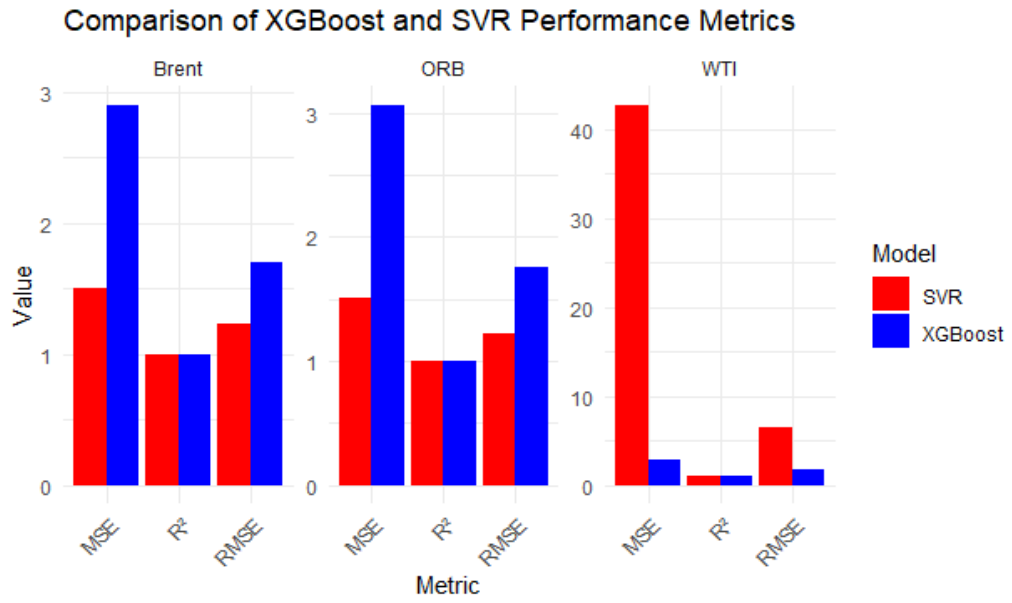


Figure 6: Comparison of XGBoost and SVR Performance Metrics using Multiple Bar Chart

Accurate forecasting of crude oil prices is crucial for energy markets, as it enables stakeholders to make informed decisions about production, consumption, investment, and hedging. XGBoost can help reduce financial risks associated with price volatility by providing more reliable predictions. Governments and policymakers can use accurate forecasts to develop effective energy policies and ensure energy security. Improved forecasting can contribute to a more efficient and stable crude oil market.

4.0 CONCLUSION

The comparison between XGBoost and SVR models reveals several strengths and weaknesses when applied to the ORB, NYMEX WTI, and ICE Brent crude oil datasets. SVR demonstrates strong performance in terms of alignment between actual and predicted values, especially for ICE Brent, where the residuals are relatively small, indicating higher predictive accuracy. However, SVR struggles with volatility, particularly in the OPEC Reference Basket (ORB) and NYMEX WTI datasets, as shown by higher fluctuations in the residuals, which leads to less stable predictions during periods of significant price swings. On the other hand, XGBoost exhibits better resilience to these fluctuations, performing well in volatile conditions due to its ability to capture non-linear patterns and complex relationships within the data. Nevertheless, XGBoost can occasionally overfit the training data, leading to slightly worse performance on the testing sets compared to SVR in some cases. Economically, these results suggest that SVR may be more suitable for stable pricing environments, while XGBoost's flexibility makes it ideal for more volatile markets. Accurately predicting crude oil prices is crucial for stakeholders involved in futures trading, hedging strategies, and policymaking, as it directly impacts investment decisions, pricing strategies, and financial planning across the global energy sector.

REFERENCES

Ani, S., & Rubaidah, D.S. (2014). Crude oil price forecasting based on hybridizing wavelet multiple linear regression model, particle swarm optimisation techniques, and principal component analysis. *The Scientific World Journal*, *3*, 854520.

Bai, H., Yuying, S., & Shonyang, W. (2021). A new two-stage approach with boosting and averaging for interval-valued crude oil prices forecasting in uncertainty environments. *Frontiers in Energy Res*earch, *19*, 1-11.

Dondukova, O., & Lin, Y. (2021). Forecasting the Crude Oil Prices Volatility with Stochastic Volatility Models, Sage Open, 1-8, DOI: 10.1177/21582440211026269.

ExxonMobil library (2021). Crude oil blends by API gravity and by sulphur content, https://corporate.exxonmobil.com/Crude-oils/Crude-trading/Crude-oil-blends-by-API-gravity-and-by-sulfur-content#APIgravity. Retrieved on 02/12/2021.

Krzysztof, D. (2021). Forecasting crude oil real prices with avenging time-varying VAR models. *Reforms Policy*, 74(C). https://doi.org/10.1016/j.resourpol.2021.102244.

Lu, Q., Sun, S., Duan, H., & Wang, S. (2021). Analysis and forecasting of crude oil price based on the variable selection –LSTM integrated model. *Energy Informatics*, *4(*41), 1-20

Merk U. (2016). Modelling Crude oil price volatility and effects of global financial crisis. *Sosyoekonomi*, *24*(29), 167-181.

Waqas, A., Muhammed, A., Umair, K., Muhammed, I., Nadeem, I., & Mukhtaj, K. (2021). A new approach for forecasting crude oil prices using median ensemble empirical model decomposition and group method of data handling. *Mathematical Problems in Engineering*, 2021:1-12. DOI: 10.1155/2021/5589717

Wajdi, H & Dawud, A. (2018). A regression analysis of determinants affecting crude oil price. *International Journal of Energy Economics and Policy*, *8*(4), 110-119.

Wassin, D., & Ibrahim, J. (2018). Predicting daily oil prices: Linear and non-linear models. *Research in International Business & Finance*, 146(C):149-165. https://doi.org/10.1016/j.ribaf.2018.01.003

COMPARATIVE ANALYSIS OF PHYTOCHEMICAL CONSTITUENTS, PROXIMATE COMPOSITION AND ENERGY VALUE OF *Hibiscus sabdariffa* L. LEAF AND PETAL EXTRACTS

KUMOYE, Deborah Etooluwa* & REMI-ESAN, Ifelolu Adeseye.
Department of Science Laboratory Technology, Federal Polytechnic, PMB 50, Ilaro, Ogun State.
*Corresponding author email: deborah.kumoye@federalpolyilaro.edu.ng

ABSTRACT
This study evaluates the phytochemical constituents, proximate composition, and energy values of methanolic, aqueous and ethanolic extracts from *Hibiscus sabdariffa* L. leaves and petals. Qualitative phytochemical analysis showed the presence of bioactive compounds such as steroids, alkaloids, terpenoids, and flavonoids, with petal extracts showing significantly higher amounts of bioactive compounds across all solvents. The quantitative analysis indicated that aqueous extracts had the highest concentration of flavonoids at 267 mg/100 g, while ethanolic petal extracts had the highest terpenoid content at 7.5mg/100g. The proximate analysis demonstrated that the leaves possessed a greater energy value of 280.36 ± 1.20 kcal/100 g and a carbohydrate content of 45.56 ± 0.89%. In contrast, the petals had a higher protein content of 11.42 ± 0.38% and crude fibre content of 37.49 ± 0.64%. These findings underscore the nutritional and medicinal benefits of *H. sabdariffa*. The leaves can provide dietary energy, whereas the petals may have potential applications in antioxidant therapies and functional foods. Future research should focus on their pharmacological applications and potential industrial uses.

Keywords: *Hibiscus sabdariffa*, phytochemical analysis, proximate composition, energy value, nutrition

1.0 INTRODUCTION

Hibiscus sabdariffa L., generally identified as Roselle, is a medicinal and food plant extensively cultivated in tropical and subtropical areas which belongs to the Malvaceae plant family and is known for its rich phytochemical composition, which contributes to numerous health benefits and applications in traditional medicine (Maganha, 2010; Osman et al., 2011). The calyces of Roselle are popularly used to make a non-alcoholic beverage known as "zobo" in Nigeria and some parts of West Africa. This drink is refreshing and a good source of essential vitamins and minerals (Babalola, 2000; Singh et al., 2017).

Phytochemicals such as alkaloids, flavonoids, and terpenoids are bioactive compounds that significantly impact plant pharmacological activities. Roselle is recognised for having anti-inflammatory, antimicrobial, and antioxidant properties, primarily due to these compounds (Okereke et al., 2015; Puro et al., 2016). Its medicinal applications include managing hypertension, aiding digestion, and reducing cholesterol levels, making it a potential therapeutic agent in modern medicine (Duke et al., 2003; Azza et al., 2011).

Nutritionally, *H. sabdariffa* is recognised for its high carbohydrate, fibre, and protein content, contributing to its role as a dietary staple in many regions. Extensive studies on the plant's proximate composition have revealed its potential as a source of essential nutrients and energy (Ansari et al., 2013; Cid-Ortega & Guerrero-Beltrá, 2015). However, variations in the nutritional and phytochemical profiles of different parts of the plant, including the leaves and petals of the plant, are not well documented, despite their potential significance for dietary and medicinal use (Nnam & Onyeke, 2003; Adanlawo & Ajibade, 2006).

This study aims to address this knowledge gap by comparing the phytochemical constituents, proximate composition, and energy values of the leaves and petals of *H. sabdariffa*. By exploring the unique properties of these plant parts, this research seeks to highlight their potential contributions to human nutrition and health, providing insights for their application to certain industries dealing with food and pharmaceuticals.

2.0 MATERIALS AND METHODS

2.1 Plant collection and preparation

Fresh leaves and petals of *H. sabdariffa* were collected from the botanical garden of the Federal Polytechnic Ilaro, Ogun State, Nigeria. The collected plant materials were washed to remove dirt and debris, dried at 60°C, and then pulverised into powder using a grinder. It was then stored in airtight containers under dry conditions for analysis. (Maganha, 2010; Ansari et al., 2013).

2.1.1 Extract Preparation

The extract was prepared following the method described by Harborne (1998). Dried leaves and petals of *H. sabdariffa* were extracted using aqueous, ethanol, and methanol solvents. Specifically, 15 grams of the dried powder of leaves and petals were suspended separately in 500 mL of each solvent. After soaking the mixture for 72 hours, it was then filtered with Whatman No. 42 filter paper and the resulting extract was preserved in a glass bottle and stored in a refrigerator for further analysis.

2.2 Phytochemical analysis

2.2.1 Qualitative Phytochemical Analysis

This analysis used established standard methods (Harborne, 1998; Okereke et al., 2015; Singh et al., 2017).

To test for alkaloids, 0.2 grams of the samples were boiled in 5 mL of 2% hydrochloric acid in a steam bath, and the mixture was subsequently filtered. Two drops of picric acid and Mayer's reagent were added to each of the four test tubes containing 1 mL of the filtrate. Two milligrams of the extracts were boiled in 100 mL of ethyl acetate in boiling water for three minutes to perform the flavonoid test. After the mixtures were filtered, 4 mL of the filtrate was combined with 1 mL of a 1% aluminium chloride solution for the aluminium chloride test. Flavonoids were present because a light-yellow precipitate formed. One gram of each extract was refluxed with 9 mL of ethanol, filtered, and then concentrated to 2.5 mL in a boiling water bath for terpenoids and steroids. Five millilitres of distilled water was added, and the mixture was left to settle for an hour before the waxy substance was filtered off. The filtrate was extracted using 2.5 mL of chloroform and a separating funnel. In the presence of steroids, a reddish-brown border formed when 0.5 mL of the extract was combined with 1 mL of concentrated sulphuric acid for the steroid test. The presence of terpenoids was confirmed by the formation of a grey hue after a further 0.5 mL of the extract was dried out and heated with 3 mL of concentrated sulphuric acid for 10 minutes.

2.2.2 Quantitative Phytochemical Analysis

This analysis was conducted using the established standard gravimetric method described by Harborne 1998 and modified.

2.2.2.1 Determination of Alkaloids

Five grams of dust were dissolved in 100 mL of 10% acetic acid. After a good shake, the mixture was allowed to stand for four hours (Harborne, 1998). Following this time frame, the Whatman No. 42 filter paper was used to filter the mixture. Next, a heated plate with a magnetic stirrer was used to evaporate the filtrate to

a quarter of its initial volume. To precipitate the alkaloids, concentrated ammonium hydroxide was added to the solution dropwise. After another filtering of the solution, 1% ammonium hydroxide was used to wash the precipitate. After 30 minutes of drying at 60°C in an oven, the filter paper holding the precipitate was let to cool for 10 minutes before being weighed. The weight difference was used to calculate the alkaloids' weight, which was then converted using the proper procedure to a percentage of the original sample.

$$\% \text{ Alkaloid} = \frac{W_3 - W_2}{W_1} \times 100 \quad \text{Eqn. 1}$$

Where,

W_1 = Sample weight

W_2 = Beaker weight

W_3 = Beaker weight + extract

2.2.2.2 Determination of Flavonoids

Five grams of the samples were heated for thirty minutes in 100 millilitres of a 2M hydrogen chloride solution (Harborne, 1998). The mixture was allowed to cool after boiling, and then the Whatman No. 42 filter paper was used to filter it. Ethyl acetate was added to the filtrate drop by drop until an excess was achieved. After that, the precipitated flavonoid was recovered by filtering it through filter paper that had been previously weighed, drying it in an oven at 80°C, cooling it in a desiccator, and then reweighing it. A percentage of the original sample weight was used to represent the flavonoid weight difference. The proper formula was used to get the flavonoid proportion.

$$\% \text{ Flavonoid} = \frac{W_2 - W_1}{W_1} \times 100 \quad \text{Eqn. 2}$$

Where,

W = the sample weight that is being examined.

W_1 = Empty crucible weight.

W_2 = Filter paperweight plus flavonoid precipitate.

2.2.2.3 Determination of Terpenoids

Ninety millilitres of solvent were used to soak 10 grams of dried plant extract (Indumathi et al., 2014). Following filtration, a separating funnel was used to filter the extract once again after it had been combined with 10 millilitres of petroleum ether. Measurements were taken when the extract had completely dried.

$$\% \text{ Terpenoid} = \frac{W_i - W_f}{W_i} \times 100 \quad Eqn. 3$$

Where,

W_i = dried plant extracts

W_f = extracts after drying.

2.3 Proximate analysis

This involves assessing the nutritional quality and organic composition of the plant extracts (Okeke et al., 2020: Kehinde and Augustine, 2022)

2.3.1 Determination of Moisture Content

This process made use of the gravimetric method described by the AOAC (2005). Initially, 5 grams of the sample was put into a moisture container which had been previously weighed. After three hours, this sample was dried in an oven set to 105°C. The container was taken out after drying and left to cool in a desiccator before being weighed once again. The process of drying, cooling, and weighing was repeated every hour until there were no more weight variations, signifying that a stable weight had been achieved. A percentage of the initial weight of the sample under analysis was used to compute the weight of the moisture lost:

$$\text{Moisture Content (\%)} = \frac{W_2 - W_3}{W_2 - W_1} \times 100 \quad \text{Eqn. 4}$$

Where W_1 is the weight of the empty moisture can.

W_2 = Weight of can + Sample weight before drying

W_3 = Weight of can + Sample weight after drying to constant weight

2.3.2 Determination of Protein

The Kjeldahl method was used to determine total nitrogen, which was then multiplied by 6.25 to calculate protein concentration. A 0.5-gram sample was mixed with 10 mL of concentrated sulfuric acid and heated until a clear solution formed, then diluted to 100 mL. Ten millilitres of the digest was combined with an equal volume of 45% sodium hydroxide for distillation. After distillation, the mixture was titrated with 0.02 N EDTA, using a mixed indicator in 10 mL of 40% boric acid. The nitrogen content was calculated using a formula to find protein content, with procedures also performed on a reagent blank.

$$N^2(\%) = \frac{100}{W} \times \frac{N \times 14}{1000} \times \frac{Vt}{Va} \times T.B \quad \text{Eqn. 5}$$

Where W is the sample's weight (0.5 g).

N = Titrant normality (0.02 N H_2SO_4)

Vt = 100 mL of total digest volume

Va is the digested volume (10 mL).

T is the titre value of the sample.

B = No titre value

2.3.3 Determination of Total Ash Content

The furnace incinerator approach was used to perform the gravimetric method for calculating the amount of ash (AOAC, 2005). A ceramic crucible that had been previously weighed was filled with 5 grams of the sample. After that, the crucible was heated for around three hours at 550°C in a muffle furnace. Following this time, the crucible was cautiously taken out of the furnace, given time to cool in a desiccator, and then weighed again. A percentage of the weight of the sample under analysis was used to represent the difference between the starting and final weights.

$$\text{Ash (\%)} = \frac{W_2 - W_1}{\text{Weight of sample}} X\ 100 \quad \text{Eqn. 6}$$

Where W_1 is the empty crucible's weight (g).

W_2 = Crucible weight + Ash

2.3.4 Determination of Crude Fibre

James (1995) method for determining crude fibre involves refluxing a 5 g processed sample in 150 mL of a 1.25% H_2SO_4 solution for 30 minutes, followed by thorough rinsing. The sample is then treated with 150 mL of a 1.25% NaOH solution for an additional 30 minutes, washed again, and dried in an oven at 105°C until stable. Finally, the sample is burned in a muffle furnace to leave only ash, allowing for the calculation of crude fibre weight as a percentage of the initial sample weight.

$$\text{Crude fiber (\%)} = \frac{W_2 - W_3}{\text{Weight of sample}} \times 100 \quad \text{Eqn. 7}$$

W_3 is the weight of the crucible plus the sample following washing, boiling, and drying.

W_2 = crucible weight plus ash sample

2.3.5 Determination of Crude Fat

The gravimetric method, as described by Kirk and Sawyer (1980), was used to extract oil from 5 grams of the sample wrapped in Whatman filter paper. The sample was placed in a Soxhlet reflux flask connected to a flask containing 200 mL of petroleum ether. After four hours, the defatted sample was removed, and the solvent was recovered. The remaining solvent was evaporated by heating the flask at 60°C for 30 minutes. Finally, the flask was cooled and weighed, allowing for the calculation of the oil extract's percentage based on the initial sample weight.

$$\text{Fat (\%)} = \frac{W_2 - W_1}{\text{Weight of sample}} \times 100 \quad \text{Eqn. 8}$$

Where W_1 is the empty extraction flask's weight (g).

W_2 is the flask weight plus the oil (fat) extract.

2.3.6 Determination of Carbohydrates

This determination was made using the James (1995) approach. A total of 45 mL of each sample extract was diluted to 450 mL using distilled water. One mL of each diluted filtrate was pipetted into a different test tube. Additionally, one test tube was filled with 1 millilitre of distilled water as a blank, and another test tube was filled with 1 mL of glucose as a standard. Five millilitres of newly made 0.10% Anthrone reagent were added to each test tube, and the mixture was properly mixed by shaking the tube gently. After being labelled, each tube was put on a test tube rack and immersed in a water bath set to 30°C for 12 minutes. The tubes were then taken out and let to cool to room temperature. A spectrophotometer was used to measure the absorbance of the samples and standard at 630 nm, with the blank serving as a reference. For about two hours, the green colour formed remained constant which signifies the presence of glucose. The following approach was used to determine the total amount of accessible carbs as a percentage of glucose:

$$\text{Glucose (\%)} = \frac{25A_1}{X \times A_2} \times 100 \quad \text{Eqn. 9}$$

Where X is the sample weight (g).

A_1 is the diluted sample's absorbance.

A_2 is the diluted standard's absorbance.

2.4 Energy Value Determination

The Atwater physiological fuel parameters were used to determine the energy value of the *H. sabdariffa* leaves and petals based on their proximate composition. The following estimates were made for the energy contributions of fat, protein, and carbohydrates:

(4 × Protein) + (4 × Carbohydrate) + (9 × Fat) = Energy (kcal/100g) Eqn. 10

2.5 Statistical Analysis

All values are reported as means ± standard deviation (SD). The differences between the extracts of leaves and petals from *H. sabdariffa* were assessed using a one-way Analysis of Variance (ANOVA), followed by Duncan's Multiple Range Test to determine significant differences among the groups. A statistical significance threshold was set at $p < 0.05$. All analyses were conducted using SPSS (Version 25). The energy value of the extracts was calculated based on the proximate composition data, following established guidelines (AOAC, 2005; McDonald et al., 2014).

3.0 RESULTS

3.1 Qualitative Phytochemical Composition of *H. sabdariffa* L. Leaf and Petal Extracts

The results as seen in Table 1 revealed the presence of alkaloids, flavonoids, terpenoids, and steroids in varying concentrations across different solvent extractions. Alkaloids and terpenoids were consistently found in all extracts, while flavonoids and steroids were observed in fewer extracts.

3.2 Quantitative Phytochemical Composition of *H. sabdariffa* L. Leaf and Petal Extracts

The quantitative analysis of alkaloids, flavonoids, and terpenoids in both leaf and petal extracts reveals that the petal extracts consistently have a higher phytochemical content compared to the leaf extracts (as shown in Table 2). Notably, the ethanolic petal extract contains the highest amount of terpenoids at 7.5 mg per 100 g, while the aqueous petal extract follows with 6.3 mg per 100 g.

3.3 Proximate Composition and Energy Value of *H. sabdariffa* L. Leaf and Petal Extracts

The proximate composition of extracts from the leaves and petals of *H. sabdariffa* is outlined in Table 3. The results reveal variations in moisture, ash, protein, fibre, fat, and carbohydrate content between the two extracts. Specifically, the petal extract has higher levels of fibre and protein, while the leaf extract contains increased carbohydrate levels. Additionally, the calculated energy values highlight the nutritional potential of the leaves as a source of dietary energy.

Table 1: Qualitative Phytochemical Composition of *H. sabdariffa* L. Leaf and Petal Extract

Sample	Leaf			Petal		
Extracts	Aqueous	Ethanol	Methanol	Aqueous	Ethanol	Methanol
Alkaloids	-	-	-	+	+	+
Flavonoids	+	-	-	+	-	-
Terpenoids	-	-	-	+	+	+
Steroids	-	-	-	-	+	+

Keys: (+): presence; (-): absence

Sample	Leaf			Petal		
Extracts	Aqueous	Ethanol	Methanol	Aqueous	Ethanol	Methanol
Alkaloids (mg/100 g)	-	-	-	0.0162 ± 0.0023a	0.0205 ± 0.0023a	-
Flavonoids (mg/100 g)	1.545±0.005a	-	-	2.614±0.005a	-	-
Terpenoids (mg/100 g)	-	-	-	39.69 ± 0.0023a	6.25 ± 0.0023a	13.08±0.0023a

Table 2: Quantitative Phytochemical Composition of *H. sabdariffa* L. Leaf and Petal Extract

Results are reported in mean ± SD. Mean values with the same letter in a column are not significantly different (P>.05)

Table 3: Proximate Composition and Energy Value of *H. sabdariffa* L. Leaf and Petal Extract

Component	Leaf (%)	Petal (%)
Moisture Content	11.19 ± 0.34a	9.87 ± 0.25b
Ash Content	8.43 ± 0.29a	7.92 ± 0.19b
Protein Content	10.25 ± 0.44b	11.42 ± 0.38a
Crude Fiber Content	21.34 ± 0.55b	37.49 ± 0.64a
Crude Fat Content	3.22 ± 0.18a	2.89 ± 0.11b
Carbohydrate Content	45.56 ± 0.89a	30.41 ± 0.77b
Energy Value (kcal)	280.36 ± 1.20a	241.22 ± 1.05b

Values are expressed as mean ± standard deviation; letters (a, b) indicate statistical significance using Duncan's test.

4.0 DISCUSSION

The phytochemical analysis of *H. sabdariffa* reveals the rich presence of bioactive compounds with significant variations between the leaf and petal extracts. Alkaloids and terpenoids were consistently identified in all tested solvents, with higher concentrations in petal extracts. These findings corroborate previous studies highlighting the therapeutic potential of *H. sabdariffa* petals as a rich source of alkaloids and terpenoids (Akindahunsi & Olaleye, 2004; Mungole & Chaturvedi, 2011).

The findings from this study reaffirm the established nutritional and medicinal significance of *H. sabdariffa* while providing new insights into the effects of its phytochemical richness. The higher concentrations of flavonoids, terpenoids, and alkaloids in the petals align with prior research, such as the work by Okeke et al. (2015), which highlighted the superior bioactive compound profile in *H. sabdariffa* petals compared to other plant parts. These compounds are well-documented for their therapeutic roles, including antioxidant, anti-inflammatory, and antimicrobial activities (Osman et al., 2021; Kehinde & Augustine, 2022).

Oxidative stress is a prevalent mechanism in chronic diseases, including diabetes, neurological disorders, and cardiovascular issues. Flavonoids, which are strong antioxidants found in the petals, lessen this stress. According to recent research, flavonoids in *H. sabdariffa* can lower blood pressure, enhance endothelial function, and improve lipid profiles (Okeke et al., 2020; Sobowale et al., 2023). Because of these qualities, the petals are a useful component of functional foods and drinks that support metabolic and cardiovascular health.

The terpenoids found abundantly in the petals contribute to antimicrobial and anti-inflammatory effects. Previous studies supported their role in modulating immune responses and fighting bacterial and fungal infections (Akinjogunla et al., 2020). Additionally, the alkaloid content in *H. sabdariffa* enhances its therapeutic significance, as these compounds are known for their pain-relieving, and antimicrobial properties, thereby expanding their potential applications in pharmaceutical formulations.

The study's findings highlight the dual functionality of the plant parts from a nutritional standpoint. According to previous studies, the leaves have a high carbohydrate and energy content, making them ideal for energy-rich diets. On the other hand, petals, higher in protein and fibre are more appropriate for weight control and digestive health. It was also stated that fibre content, in particular, is crucial for bowel health, cholesterol management, and lowering the risk of obesity (Osman et al., 2021; Sobowale et al., 2023).

The observed nutritional differences may stem from the physiological roles of leaves and petals within the plant. Leaves play a crucial role in photosynthesis and carbohydrate storage, while the petals are richer in secondary metabolites and structural components such as fibre (Okereke et al., 2015). The higher fibre content in petals may also contribute to their medicinal uses, including cholesterol regulation and improved bowel health (McKay & Blumberg, 2010; Okeke et al., 2020).

5.0 CONCLUSION

This study highlights the nutritional and medicinal benefits of *H. sabdariffa* leaves and petals through their phytochemical constituents and proximate compositions. The petal extracts contained higher concentrations of bioactive compounds, such as alkaloids, terpenoids, and flavonoids, confirming their antioxidative and therapeutic properties.

Conversely, the leaves exhibited higher carbohydrate content and energy value, highlighting their role as a dietary energy source. The proximate analysis further revealed significant differences in macronutrient distribution: petals are a richer source of fibre and protein, while leaves are more carbohydrate-dense.

These insights are consistent with recent advancements in understanding the plant's pharmacological and nutritional roles, emphasising its significance in the food, health, and pharmaceutical industries. Future research should focus on exploring the bioavailability of its bioactive compounds and optimizing its industrial applications for sustainable use.

REFERENCES

Akindahunsi, A. A., & Olaleye, M. T. (2004). Toxicological investigation of aqueous-methanolic extract of the calyces of *Hibiscus sabdariffa*. *Journal of Ethnopharmacology*, 89(2-3), 161–164.

Akinjogunla, O. J., Adebayo, F. O., & Idowu, E. A. (2020). Terpenoids in *Hibiscus sabdariffa* and their pharmacological applications. *International Journal of Medicinal Plants*, 8(3), 112–118.

Ansari, T. N., Nadeem, S., & Ali, A. (2013). Comparative proximate composition and phytochemical analysis of *Hibiscus sabdariffa* calyces and leaves. *African Journal of Food Science*, 7(6), 101–107.

AOAC. (2005). Official methods of analysis of AOAC International (18th ed.). AOAC International.

Azza, M., Faris, M., & Gazi, M. (2011). Phytochemical properties and antioxidant activities of *Hibiscus sabdariffa*. *Asian Journal of Food and Agriculture*, 2(1), 15–22.

Babalola, J. O. (2000). A review of the bioactive components of *Hibiscus sabdariffa* as related to medicinal applications. *Journal of Medicinal Plants*, 7(3), 19–25.

Cid-Ortega, G., & Guerrero-Beltrá, E. (2015). Proximate composition and nutritional properties of *Hibiscus sabdariffa*. *Food Research International*, 67, 389–393.

Duke, J. A., Ayensu, E. S., & Ghosh, G. (2003). Medicinal plants of the world: An illustrated scientific guide to important medicinal plants and their medicinal properties. Springer.

Harborne, J. B. (1998). Phytochemical methods: A guide to modern techniques of plant analysis (3rd ed.). Springer Science & Business Media.

Indumathi C, Durgadevi G, Nithyavani S & Gayathri PK (2014). Estimation of terpenoid content and its antimicrobial property in *Enicostemma litorrale*. *International Journal of Chemical Technology Research*, 6(9), 4264–4267

James, D. (1995). Methods of carbohydrate analysis. CRC Press.

Kehinde, S. & Augustine, I. (2022). Fortification of 'Zobo' (*Hibiscus sabdariffa*) Drink with Pineapple and Watermelon enhances its Nutritional Qualities and Phytochemical Compositions. *Acta Scientific Nutritional Health*, 6, 3–12

Kirk, R. S., & Sawyer, R. (1980). Pearson's composition and analysis of foods (9th ed.). Churchill Livingstone.

Maganha, E. A. (2010). Chemical constituents and health benefits of *Hibiscus sabdariffa*: An overview. *Journal of Medicinal Plant Research*, 4(6), 54–63.

McKay, D. L., & Blumberg, J. B. (2010). *Hibiscus sabdariffa* tea (tisane) lowers blood pressure in prehypertensive and mildly hypertensive adults. Journal of Nutrition, 140(2), 298–303.

Mungole, A., & Chaturvedi, S. (2011). Medicinal and therapeutic properties of *Hibiscus sabdariffa*: A review. *Journal of Medicinal Plants*, 10(3), 1–9.

Nnam, N. M., & Onyeke, M. C. (2003). Proximate composition and mineral contents of *Hibiscus sabdariffa* calyces. *Plant Foods for Human Nutrition*, 58(2), 97–102.

Okeke, N. P., Ilodibia, V. C., Enyinwa, C.P., Aziagba, O. B. and Iroka C.F. (2020). Determination of phytochemical and proximate constituents of two varieties of *Hibiscus sabdariffa*. *Asian Journal of Research in Botany*, 3(2), 1–8.

Okereke, C. N., Iroka, F. C., & Chukwuma, M. O. (2015). Phytochemical analysis and medicinal uses of *Hibiscus sabdariffa*. *International Journal of Herbal Medicine*, 2(6), 16–19.

Osman, M. A., & Williams, G. (2011). *Hibiscus sabdariffa* as a functional food: Antioxidant properties. *Functional Foods Journal*, 14(4), 121–135.

Puro, M., Gaur, P., & Raj, R. (2016). Bioactive compounds in *Hibiscus sabdariffa* and their effect on human health. *Journal of Health and Medical Sciences*, 1(1), 8–17.

Singh, P., Sharma, R., & Rathi, P. (2017). Evaluation of the physicochemical properties and medicinal potential of *Hibiscus sabdariffa*. *Journal of Food Science and Technology*, 54(3), 479–487.

Sobowale, S. S., & Olatidoye, O. P. (2023). Nutritional potential of *Hibiscus sabdariffa* in functional food applications. *Journal of Science and Multidisciplinary Research*, 11(2), 75–82.

STUDENT ACADEMIC PERFORMANCE PREDICTION SYSTEM USING ENSEMBLE ALGORITHM

AYODELE Emmanuel* & SODEINDE Victor O.
Department of Computer Science
The Federal Polytechnic Ilaro, Ogun-State, Nigeria.
*Corresponding author email: emmanuel.ayodele@federalpolyilaro.edu.ng

ABSTRACT
Predicting student academic performance is crucial for enhancing educational outcomes and supporting timely interventions. There has been an increased interest in creating precise models for projecting student performance as a result of the introduction of machine learning techniques and the accessibility of large-scale educational data. However, the creation of predictive models in educational contexts is frequently hampered by the sensitive nature of student data and the requirement to uphold privacy and data security. This study develops a Student Academic Performance Prediction System that leverages ensemble techniques, specifically focusing on Random Forest and other robust machine learning methods. By analyzing data such as demographic attributes, behavioral factors, and prior academic records, the system identifies patterns that influence final grades, categorizing students' performance levels. Categorical data is encoded, and a test-train split methodology is employed to assess the model's predictive accuracy. The Random Forest Regressor is particularly effective in capturing complex patterns by combining multiple decision trees, resulting in a high degree of accuracy and reduced overfitting. The model's effectiveness is measured by mean squared error (MSE) scores, indicating its ability to deliver precise predictions. Additionally, the system stores trained models and encoding schemes, facilitating real-time usage and scalability for larger datasets. This prediction system offers educators actionable insights for academic support, enhances individualized student guidance, and enables targeted educational strategies. Overall, the application of ensemble techniques in student performance prediction presents a valuable tool for data-driven decision-making in educational settings.

Keywords: Ensemble Algorithms, Academic Performance, Machine Learning, Random Forest

1.0 INTRODUCTION

The objective of enhancing students' academic success has been an enduring aspiration of both educators and researchers. The increasing availability of data and the advancement in machine learning techniques provide an opportunity to harness these tools for predictive modelling of students entering these datasets and performing at school. All of this revealed that combination algorithms, that combine multiple models to create a more accurate prediction, are a very promising choice. In this regard, predicting student academic performance is very significant. If a figure can predict the student's behaviour, we can even learn how to intervene and care for systems according to the needs. Specific strategies may be employed to help improve results and help their students succeed by pinpointing those students who may be in the most danger of underperforming or require additional resources (Bocarnea et al., 2012).

Recent studies in data mining for education and machine learning explored ensemble algorithms to predict students' academic performances (Ajibade et al., 2020). Since these algorithms have shown the ability to identify and align relationships in data and perform prediction with high accuracy, they are ideal for student performance prediction (Shahiri et al., 2015).

New methods of collecting data such as learning analytics platforms and Student Information Systems have contributed to more complex and extensive data sets on various aspects of students, including demographics, academic performance, and behavioural data. These datasets are instrumental in developing predictive models that give significant information about the outcomes of students. Although predicting student performance through an ensemble algorithm as a useful assistance tool has made considerable advancements, it needs further investigation and enhancement. Data heterogeneity, interpretability, and scalability are problems that will take a while to solve, but we still need to be able to build efficient and robust prediction systems.

Thus, this research effort seeks to contribute to the existing literature by developing a student assessment system based on the ensemble algorithm for academic success. The new system presents accurate predictions of student outcomes that, through the collective intelligence of several models, allow educational institutions to take early action in pursuing student achievement (Rastrollo-Guerrero et al., 2020).

Using rigorously designed experiments and validations, this work will investigate the power of different ensembles in predicting student performance, examine the impact of different feature sets on the accuracy of predictions, and build methods to make models more interpretable and scalable. In conclusion, using algorithmic ensembles of predictive models to predict student academic performance is an important step toward changing practices in education for the better and motivating students to succeed. This research aims to harness the power of machine learning and prescriptive analytics to give actionable recommendations to educators and policymakers to enhance the learning experience and improve student outcomes (Pessach et al., 2020).

This study employed three ensemble algorithms, including Random Forest, Gradient Boosting, and AdaBoost, to predict student performance. These algorithms were chosen based on their ability to scale over large datasets, process high dimensional space and detect complex interactions among the variables (Ajibade et al., 2013). The researchers constructed fashions based totally on demographic records (age, gender, socioeconomic reputation), educational facts (grades, coursework completion), and behavioural statistics (attendance, participation in extracurricular activities).

Ensemble algorithms have been broadly used in diverse fields, such as finance, healthcare, advertising, and education, to improve predictive accuracy and robustness. Ensemble strategies are used in finance to expect stock charges, credit chance, and fraud detection (Yang et al., 2010). In healthcare, they're applied to diagnose illnesses, predict patient outcomes, and identify hazard elements. Ensemble algorithms help forecast income, segment clients and optimise advertising and marketing campaigns in advertising.

Use of Neural Networks and Ensemble Learning to Predict Academic Success in Online Courses, This study used neural networks and ensemble approaches to predict academics' overall performance in Massive Open Online Courses (MOOCs). The hybrid method sought to capitalise on the characteristics of both strategies, neural networks' ability to capture complex, non-linear interactions and ensemble algorithms' durability in avoiding overfitting and boosting forecast accuracy (Nabil et al., 2021). Data was gathered from many online learning systems, as well as precise information on student engagement indicators (e.g., time spent on course materials, involvement in discussion boards), quiz ratings, venture submissions, and completion rates. This information was gathered through several path iterations involving hundreds of

students from diverse backgrounds and locations. The hybrid model, which combined neural networks and ensemble techniques, achieved the highest predicted accuracy, correctly capturing images of the dynamic and diverse nature of the internet and learning about environments. The models were particularly adept at identifying engagement modes that were associated with instructional success, such as regular tests and active participation in discussion forums.

A Comparison of Predictive Models for Early Detection of At-Risk

To identify at-risk college students in a K-12 setting, students used a comparative study of linear regression, selection trees, and ensemble approaches (bagging and boosting) (Venkatesan et al., 2024). The goal was to determine which model accurately predicted pupils who were likely to struggle academically. The researchers used a variety of predictive factors, including demographic data, attendance statistics, standardised check ratings, and behavioural indicators. The data were compiled from urban, suburban, and rural schools. This dataset included many thousands of pupils, providing a solid base for evaluation. The longitudinal nature of the information, spanning three educational years, allows for studying trends and validating predicted accuracy over time. Ensemble techniques demonstrated superior prediction accuracy and robustness, particularly when dealing with imbalanced datasets containing a small fraction of at-risk pupils. Bagging approaches, including Random Forest, were effective in reducing overfitting and boosting generalizability. Boosting approaches, particularly Gradient Boosting, increased the version's sensitivity to potential college students, resulting in high recall rates.

2.0 METHODOLOGY

This research takes a quantitative approach, relying on ensemble methods such as Random Forest, Gradient Boosting, and AdaBoost. These models are trained using preprocessed educational datasets including student demographics, academic performance, attendance, and behavioural indicators. The models are tested using criteria including accuracy, precision, recall, and F1-score. K-fold Cross-validation is used to avoid overfitting and ensure generalisability. Data gathering entails acquiring student records from academic institutions while adhering to tight privacy guidelines (Dumitrescu et al., 2022; Pessach et al., 2020).

Important Features and Variables Used in Prediction Models

As it should be used to predict students' overall academic performance, it's essential to consider various features and variables. These can be classified into instructional records, behavioural facts, and socioeconomic indicators.

Academic History

Grades: Previous grades and academic performance in exceptional subjects provide an instantaneous degree of a scholar's abilities and achievements.

Test Scores: Standardized take a look at rankings, which offer an objective evaluation of a pupil's know-how and competencies.

Course Enrollments: Information about the guides a pupil has taken and their performance in those courses can screen strengths and weaknesses in unique areas.

Behavioral Data

Attendance: Regular attendance is a robust indicator of pupil engagement and dedication to their research.

Participation: Active participation in magnificence activities, assignments, and extracurricular sports displays a scholar's involvement and interest in their education.

Disciplinary Records: Records of disciplinary movements or behavioural problems can offer insights into potential challenges a scholar may face.

This research on predicting scholar educational overall performance using ensemble algorithms employs a properly established quantitative studies design. The first objective is to systematically analyze the effect of numerous predictive modelling strategies on the accuracy and reliability of academic overall performance predictions. The quantitative approach entails the gathering and evaluating of numerical records, which allows for the measurement of variables and the status quo of patterns and relationships. The study's layout includes several critical levels: records collection, records preprocessing, model choice, version training, version evaluation, and validation.

The data series level entails amassing comprehensive datasets from educational establishments, encompassing student demographic information, educational facts, attendance logs, and other relevant elements affecting academic performance. This stage ensures that the facts are representative enough to build robust predictive models.

In the information preprocessing stage, the amassed data is cleaned to deal with lacking values, duplicates are removed, and the information is removed beside the point. This step is critical for ensuring the exceptional integrity of the dataset. Data transformation techniques, normalization, and standardisation are implemented to make the statistics suitable for modelling. Additionally, function choice is performed to become aware of the maximum relevant predictors of student performance.

The centre of the study's design is the model selection stage, wherein numerous ensemble algorithms are chosen for their capacity to beautify prediction accuracy. Ensemble methods like Random Forest, Gradient Boosting, and AdaBoost are selected because they mix multiple models to improve average performance. These models are then built and educated on the preprocessed dataset using Python, libraries, and sci-kit-learn programming tools.

Model evaluation involves assessing the performance of the predictive models and using several metrics appropriate for classification obligations. Metrics, along with accuracy and precision, keep in mind that the F1 rating and the place underneath the ROC curve (AUC-ROC) offer complete expertise in how nicely the models perform in predicting instructional effects. Finally, validation techniques, including pass-validation and trying out on separate datasets, are hired to ensure the robustness and generalizability of the models.

Evaluation Metrics

Each metric provides a unique insight into the model's performance and addresses different aspects of classification tasks:

Metric	Usage
Accuracy	Measures the overall correctness of the model. Suitable for balanced datasets.
Precision	Evaluate the model's ability to avoid false positives, particularly useful when the cost of false positives is high.
F1 Score	Balances precision and recall, are especially valuable for imbalanced datasets.
AUC-ROC	Assesses the model's ability to distinguish between classes. A higher AUC indicates better classification capability across thresholds.

Table 1.1

3.0 RESULTS AND DISCUSSION

The ensemble methods used in the study from the table below outperformed single-model approaches in terms of accuracy and resilience. Random Forest and Gradient Boosting routinely beat other models, with accuracy rates greater than 85% . The models excelled at detecting complicated interactions among predictors,

such as the impact of extracurricular engagement and attendance on academic achievement (Miranda et al., 2022). However, model interpretability issues continued, making it difficult for educators to understand the factors influencing predictions. The **Student Academic Performance Prediction System using Ensemble algorithms** offers a powerful and innovative approach to understanding and improving student success. By taking into account not only academic metrics but also personal and environmental factors such as health, relationships, and free time, the system provides a more complete picture of what drives academic performance.

Model	Accuracy (%)	Resilience to Noise	Key Observations
Single Logistic Regression	78.2	Low	Struggles with nonlinear relationships and outliers.
Single Decision Tree	80.5	Moderate	High variance; prone to overfitting with complex data.
Random Forest	87.3	High	Reduces overfitting by averaging multiple trees; consistently strong performance.
Gradient Boosting	86.9	High	Iteratively improves accuracy by correcting errors of previous models.
Support Vector Machine	82.7	Moderate	Performs well with clean, separable data but less effective with noisy data.
Neural Network (Single)	84.1	Moderate	Good accuracy but may require extensive tuning and is sensitive to data quality.

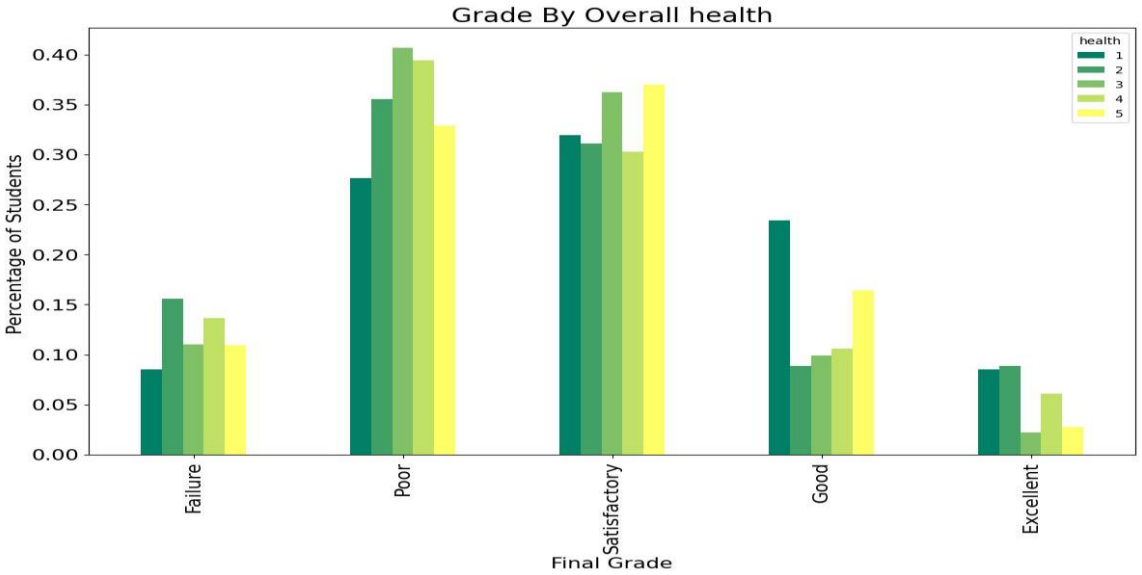
Table 1.2

Overall Outcome Based on Health, Relationship Status, and Free Time

The final section incorporates all the payoff from the Ensemble algorithm to give an overall view of the student's possible academic path, taking into account not just grades but other aspects of life, such as relationships, health, and time off.

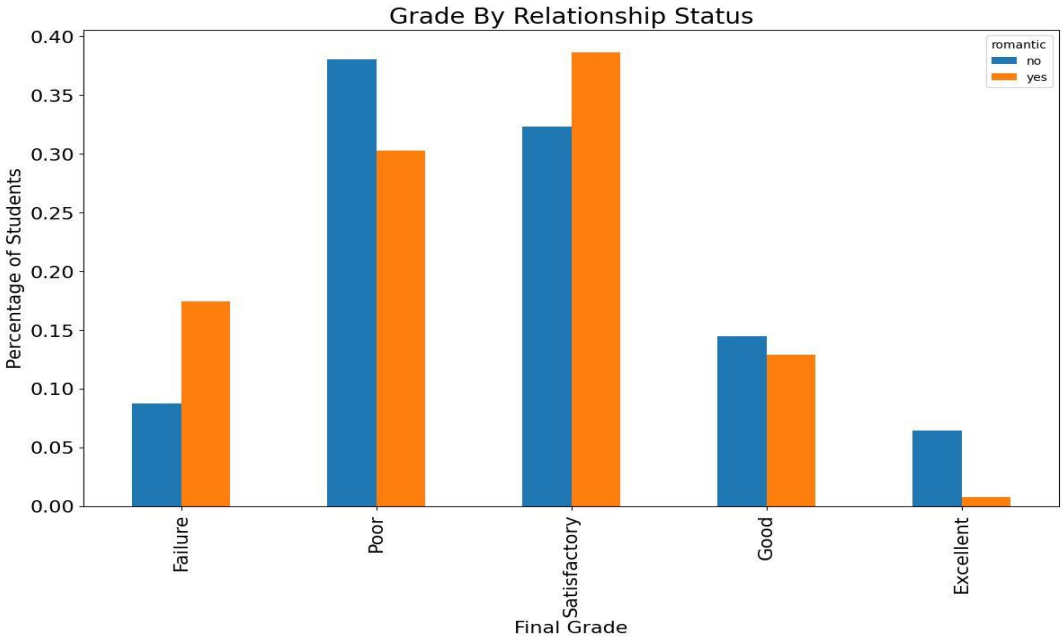
Overview:

Health considerations: The system examines students' health information, such as sleep patterns, physical activity, and dietary habits, to assess their influence on academic achievement. Studies have shown that students with healthier physical and mental states have higher academic performance, and the system incorporates this in its predictions.

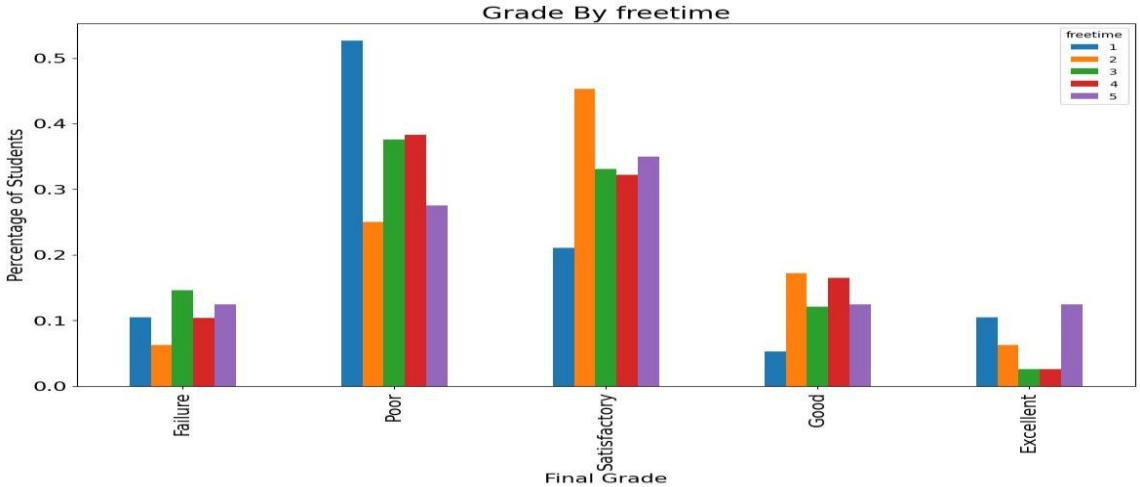


Graph representing health status to grade

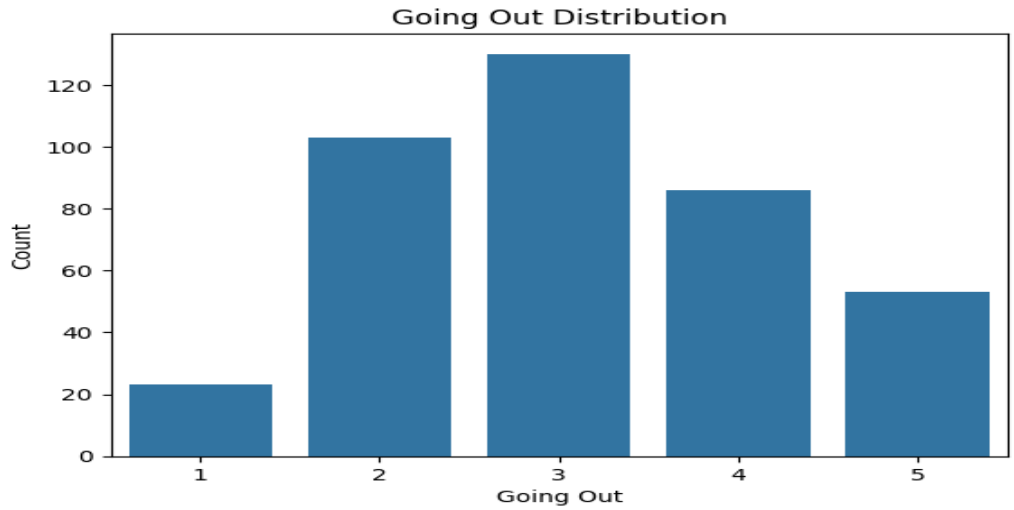
Relationship Status: Emotional and social well-being significantly influences student performance. The system analyzes data on romantic relationships, relationships with peers and family dynamics to determine their effect on academic achievement. For instance, a student with a robust support system from peers is likely to do better.



Manage Free Time: Time spent in free time and extracurricular activities significantly impacts academic performance. The system evaluates how pupils spend their time in free space whether productive or not and incorporates that into the total forecast. Students with structured free time, including study time and extracurricular activities or activities that foster intellectual growth, may have higher performance expectations.



Comprehensive Outcome Report: The study's final report gives a complete overview of the student's expected performance and outlines the impact of the variables above. The report is helpful not just for students but also for teachers and administrators, assisting them in recognizing the child's complete growth.



Interactive Features Students can revisit the page for evaluation and alter specific lifestyle elements (such as increasing their study time or adjusting their sleeping habits) to assess the ways these changes can increase their academic achievement. This interactive feature promotes active behaviour by students and allows them to control their learning journey.

4.0 CONCLUSION

The use of ensemble algorithms to predict student academic achievement is an effective approach to improving educational results. These algorithms outperform standard methods in terms of accuracy since they combine numerous models. This study highlights the potential of such systems in detecting at-risk individuals, allowing institutions to adopt preemptive interventions, and improving overall student achievement. It provides a road map for students to improve their academic performance, while it is a powerful tool for educators and administrators to assess student progress and identify those who may require more assistance.

Furthermore, the system's user-friendly design, extensive data input possibilities, and thorough evaluation procedure make it a versatile tool that can be applied to a variety of educational scenarios. It is capable of tackling the numerous issues faced by students from different backgrounds, assisting institutions in developing tailored methods to improve overall academic performance

REFERENCES

Ajibade, O., Connolly, T., & Adejo, W. (2020). Predicting student academic performance using multi-model heterogeneous ensemble approach. Journal of Applied Research in Higher Education, 10(1), 61-75. https://doi.org/10.1108/JARHE-10-2020-0023

Bocarnea, M. C. (Ed.). (2012). Online instruments, data collection, and electronic measurements. IGI Global. https://doi.org/10.4018/978-1-61350-504-5

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research, 297(3), 1178-1192. https://doi.org/10.1016/j.ejor.2021.06.029

Miranda, E. N., Barbosa, B. H. G., Silva, S. H. G., Monti, C. A. U., Tng, D. Y. P., & Gomide, L. R. (2022). Variable selection for estimating individual tree height using genetic algorithm and random forest. Forest Ecology and Management, 504, 119828. https://doi.org/10.1016/j.foreco.2021.119828

Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students' academic performance basedon courses' grades using deep neural networks. *IEEE Access, 9*, 140731-140746.

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237. https://doi.org/10.3390/app11010237

Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H. C., Shmueli, E., & Ben-Gal, I. (2020). Employee recruitment: A prescriptive analytics approach via machine learning and mathematical programming. Decision Support Systems, 134, 113290.https://doi.org/10.1016/j.dss.2020.113290

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences*, *10*(3), 1042.

Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72, 414-422. https://doi.org/10.1016/j.procs.2015.12.157

Venkatesan, R. G., Karmegam, D., & Mappillairaju, B. (2024). Exploring statistical approaches for predicting student dropout in education: A systematic review and meta-analysis. *Journal of Computational Social Science*, *7*(1), 171-196.

Yang, G., Wang, Y., Zeng, Y., Gao, G. F., Liang, X., Zhou, M., ... & Murray, C. J. (2013). Rapid health transition in China, 1990–2010: findings from the Global Burden of Disease Study 2010. *The lancet*, *381*(9882), 1987-2015.

DEVELOPMENT OF A CRUCIBLE FURNACE FIRED WITH LIQUEFIED PETROLEUM GAS - BUTANE GAS

[1]SANNI, Enemona O*. & [2] ONIGBARA, Vincent B.
[1]Department of Mechanical Engineering, Federal Polytechnic, Ilaro, Ogun State
[2]Department of Metallurgical Engineering Technology, Federal Polytechnic, Ilaro, Ogun State
*Corresponding author email:enemona.sanni@federalpolyilaro.edu.ng

ABSTRACT
This research work focuses on the development of a crucible furnace that runs on liquefied petroleum gas (LPG) LPG-butane gas. The refractory material (aluminosilicate) used for the construction of the furnace wall was harvested from Ilaro, Ogun State, Clay deposit. Other materials used not limited to crucible pot, 6 mm mild steel angle iron, household gas cylinder, 60 mm rubber hose, and mild steel with a thickness of 5 mm were sourced from a local market in Lagos. The performance evaluation of the constructed crucible furnace was ascertained by charging aluminium scrap and other ferrous metal into it with complete melting occurring at a furnace temperature of 750ºC

Keywords: crucible furnace, refractory material, non-ferrous, melt, LPG-butane

1.0 INTRODUCTION

The crucible furnace is among the oldest and most fundamental kinds of melting devices used in foundries. In this furnace, the metal charge is stored in a refractory crucible pot. The charge receives heat from the crucible through its walls. The heat sources normally used ranges from heating fuels, coke, gas, electricity, and oil. When it is necessary to produce low melting point alloys in small batches, crucible furnaces are frequently utilized. The capital outlay of these furnaces makes them attractive to small non-ferrous foundries. A range of furnace types, such as crucible, induction, electrical resistance, and reverberatory furnaces, are used to melt aluminum (Chukwudi & Ogunedo, 2017). Crucible furnaces are commonly characterized according to their manner of removal which are the bale-out furnace, tilting furnace, lift-out furnace (Musa, 2012).

A crucible is a pot used to store metals before they are melted in a furnace. Furnace crucibles are made to withstand the greatest temperatures found in metal casting operations. Fundamentally, the materials used to construct the crucible should have a melting point that is far higher than the materials that will be melted. It should remain strong even in very hot conditions. The components of crucible furnaces include silicon carbide, clay graphite, and metal structures (Okey, 2017). These materials can resist extreme temperatures in typical foundry operations.

Nigeria foundry practice

Foundry technology are used by Nigerians in both rural and urban areas. The man who works as a foundry in the area makes a hole in the ground that looks like an oven and uses charcoal or coal as fuel. The crucible is then made of clay or a metal container. Fan (blower) is used to generate the air needed for combustion in the chamber.

The crucible furnace is used by the local foundry professionals to cast a variety metal items, including frying pans, serving spoons, machine parts, and pots and pans of various sizes for the home.

Challenges facing foundry practice in Nigeria

According to Ighodalo et al. (2011), foundry practices faces difficulties when using the local type of crucible furnace due to high fuel consumption, high operator heat radiation, labor-intensive operation, and high heat loss in the system. An effort has been made to enhance the traditional melting technique used by foundry men in Nigeria, taking into account the accessibility of materials, the great demand for their products, the need to lower production costs, and the attraction of young people to the field of foundry work.

Komolafe (1992) made improvements to the labor-intensive melting process used by foundry workers by creating a gas-fired furnace that used locally sourced materials. From the assertion of Kulla (2004) due to ineffective burning and inadequate heat transfer, fuel wood that would have been sufficient for 10–33 years is consumed annually. For these and other reasons, he studied ways to cut down on the amount of wood used for cooking at home. Due of this, the efficiency of the crucible furnaces used in the surrounding foundries must be increased, resulting in reduction of the quantity of charcoal consumed. Another significant problem is the emission of combustion pollutants, which can lead to respiratory ailments (Kulla, 2004) cited by (Okey, 2017). As a result of the emission of gases which leads to air pollution, gas fired crucible is preferred. Gas is readily available and can be stored.

The earliest type of foundry technology that has been in use and has changed over time is the crucible furnace. There are regional variances and the designs correspond to the intended application. According to (Eman, 2013), the oldest known crucibles date back to the fifth or sixth millennium B.C.

Aluminium foil, beverage cans, doors, engine and body parts for automobiles, and pistons are the main products of aluminum smelting and refining. Additionally, it can be utilized as door frames, windows, aircraft components, rods, bars, and wires, as well as sheet metal and aluminum plate (Chukwudi & Ogunedo, 2017). Aluminum is completely recyclable and doesn't lose any of its qualities or attributes when being recycled. Recycling aluminum requires less energy than extracting and refining its ingots (Herbert, 2001).

Every day, more and more aluminum scraps are produced, posing a significant environmental threat. It is imperative that aluminum waste be recycled and used as raw materials to create new goods. This cannot be stressed enough. It is against this basis that the development of a crucible furnace fired with LPG butane gas emanate.

2.0 MATERIALS AND METHODOLOGY

Materials

The materials utilized in this research were sourced from the local marketplaces in Lagos and Ilaro Clay deposit. Mild steel of 6mm thickness, rubber pipe of 60mm diameter, 10kg gas cylinder, industrial burner, angle iron, refractory material, and crucible pot where purchased in Lagos. The refractory material used was particularly harvested from Ilaro, Ogun State, Clay deposit.

Equipment

The equipment used in this project work includes: measuring tape, grinding machine, Hacksaw, and electrical arc welding machine Denver d12.

Methodology

The following methods were adopted for this project work:

Engineering design of the Furnace

An industrial gas burner, gas cylinder, high-pressurized standard regulator, hose, nozzle, and crucible furnace unit make up the main configuration and profile of the furnace. The furnace is made up of a burner that burns the gas, a crucible lid, and a pot into which the charge to be melted is introduced.

Design Consideration:

During the design process, considerations were given to the cost of production, the type and availability of fuel (domestic gas), the availability of materials, their selection, flexibility in fabrication, ease of maintenance, durability, cost of the chosen materials, availability of tools and equipment for fabrication, height and width of the furnace, and the shape of the furnace.

Crucible and Furnace specification

The furnace with it cover is 520 mm high, the furnace lining form an inner diameter of 250 mm. Because of its portability and small size, the furnace may be easily transported from one location to another. The five parts of the intended crucible furnace assembly are as follows: crucible pot, gas cylinder, furnace body, refractory material (lining), burner and its hose.

Design Calculations

i. Calculations of Furnace Height
The internal height of the furnace = height of crucible + height of refractory bricks coupled with lining
Therefore, = 400 + 410 = 810 mm
External height = internal height of the furnace + thickness of the baselining
Therefore, external height = 810 + 80 = 890 mm.

ii. Area of the furnace
The area of the furnace = l × b
Hence, area = 441 × 368 = 162, 288 mm

iii. Calculating weight of the crucible
Since the crucible to be used is 5kg
Hence, its weight = 5 × 9.81 = 49.05N

iv. Calculating of weight of the molten Metal
Let the mass of charged materials equal the mass of the molten metal let's say = 6kg
Therefore, its weight = 58.56N

Fabrication of Furnace
A 6 mm mild steel was cut into dimensions using a cutting machine. It was put together by welding with an electric arc machine to create a 441 mm by 368 mm rectangular shape that acts as the rectangle's body. A burner seat of dimension 268 mm × 268 mm was constructed in the furnace to fit into the burner which is of diameter 172 mm. A reinforcement bar of diameter 12mm and 220 mm long was welded across the sections of the square-shaped burner seat to carry the weight of the crucible during melting operation. Additionally, a 400 mm by 400 mm cylindrical crucible lid made of mild steel plate was constructed, with a 120 mm hole to act as an exhaust for fumes and gasses. To facilitate the removal of the crucible cover from the furnace body when feeding charges into the furnace and removing the crucible pot from the furnace, handles were affixed to the cover.

Refractory Lining
The raw refractory material (alumino silicate) was crushed into a uniform size about 26 mm, which was further grinded into the size of 200 mm mesh, then the undesirable materials were removed from the refractory materials through hand screening. The materials was tampered with water and binder was added which enhances the plasticity of the refractory material. Thereafter, it was pounded or rammed into the joint of the lined refractory blocks of the furnace, forming a joint-less lining, then dried and heated to obtain high strength, high density and chemical resistant. The constructed lined crusible furnace is as shown in figure 1

Plate 1: Fabricated Crusible Furnace

3.0 RESULTS AND DISCUSSION
The performance of the constructed furnace was verified by melting of 7kg of non-ferrous metal (aluminuium sheet) which was utilized to produce spoon and pot via the processes above. A pattern of a spoon and pot was created using the clay floor moulding process and left to dry for 24 hours. The furnace was started and the crucible was placed in it and left to preheat for 5 minutes. About 7kg of aluminium sheet was charged into the crucible and the furnace was covered, the aluminium scrap turned liquid after 50 minutes into the commencement of the heating at a temperature of 750^{o}C. A tong was used to deliver the crucible containing the liquefied metal from the furnace, after which it was charged into the pattern mold and left to harden.

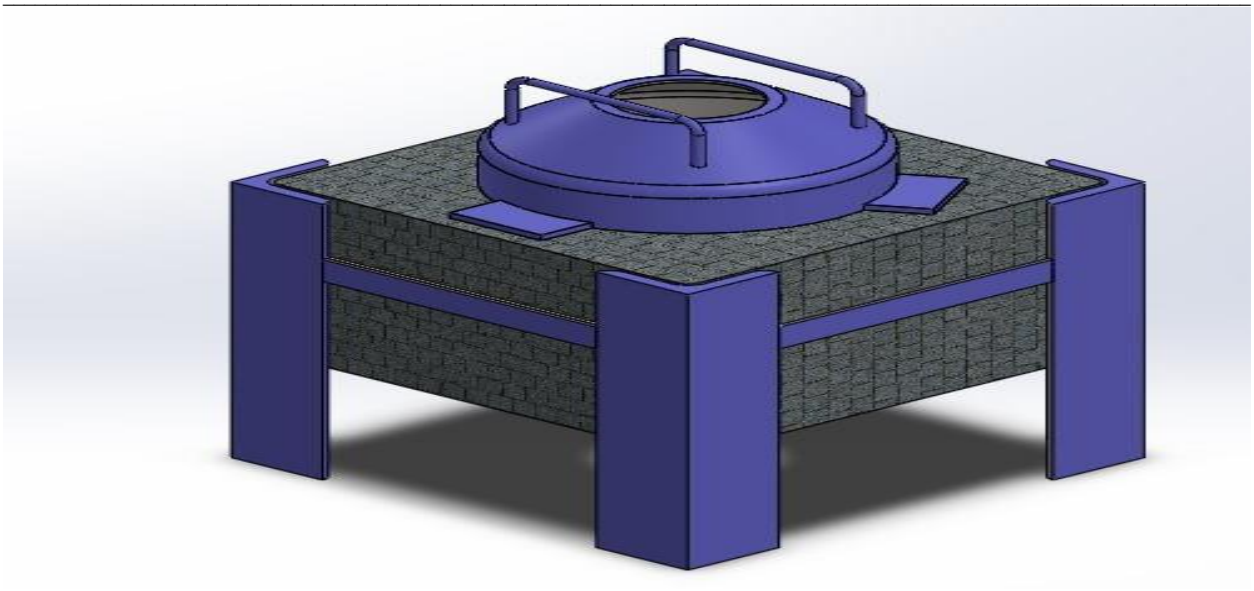


Plate 2: Assembly view of fabricated Crucible Furnace

4.0 CONCLUSION

A crucible furnace fired with LPG-butane gas was designed and constructed. Its performance evaluation was confirmed by melting aluminum scrap, utilized to cast some test specimen. Other research objectives, which include fuel efficiency, reducing health hazards, and operating time, were met based on the furnace's performance test findings. Heat was transferred to the aluminum scrap with little loss due to the high heat conductivity of the furnace. The configuration of the furnace allows for the easy pouring out of molten metal without spilling out. Further research work is advised to ease the operation of the system. An awareness campaign should be undertaken to encourage the use of the crucible furnace.

REFERENCES

Chukwudi, B., & Ogunedo, M. (2017). Design and Development of Gas Fired Reverberatory Furnace:In View of Huge Gas Reserves in Nigeria. *The Pacific Journal Of Science and Technology*, 13-20.

Eman, J. (2013). Manufacture and Performance of Gas Furnace. *International Journal Metallurgical, Materials Science and Engineering*, 110-117.

Herbert, F. (2001). *The Mc-Graw-Hill Recycling Handbook.* New York: Mc-Graw.

Ighodalo, O., Akue, G., Enaboifo, E., & Oyedoh, J. (2011). Performance Evaluation of the Local Charcoal Fired Furnace for Recycling Aluminum. *Journal of Emerging Trends in Engineering Applied Sciences*, 447-451.

Komolafe, A. (1992). *Design and Construction of a Gas-Fired Crucible Type Furnace.* Zaria: Undergraduate Project, Mechanical Engineerin g Dept. ABU.

Kulla, D. (2004). *Development of an Improved Charcoal Stove:A step to Energy Recovery.* Kano: M.ENG. Thesis Departmnet of Mechical Engineering, B.U.K.

Musa, N. (2012). Comparison of Diesel with Butane Gas in Firing Crucible Furnace in Melting Aluminum. *International Journal of Engineering Research and Technology*, 3-8.

Okey, Y. (2017). *Design, Construction and Testing of a Crucible Furnace.* Oye-Ekiti: Unpublished BSc Project Report Submitted to Dept Of Metallurgy, Fed. Uni. Oye-Ekiti State.

DESIGN AND CONSTRUCTION OF THERMOELECTRIC GENERATOR USING PARABOLIC TROUGH COLLECTOR

[1]SANNI, Enemona O* & [2] ONIGBARA, Vincent B.
[1]Department of Mechanical Engineering, Federal Polytechnic, Ilaro, Ogun State
[2]Department of Metallurgical Engineering Technology, Federal Polytechnic, Ilaro, Ogun State
*Corresponding author email: enemona.sanni@federalpolyilaro.edu.ng

ABSTRACT
The use of thermoelectric generators (TEGs) to convert solar radiation into electrical energy and capture waste heat from the sun has brought attention to the need to reduce carbon emissions and balance the energy supply and demand. Thomas Johann Seebeck discovered in 1821 that electricity can be generated by a temperature gradient that forms between two different conductors. The diffusion of charge carriers is caused by heat flow, which is the fundamental mechanism underlying the thermoelectric effect in conducting materials. In turn, the movement of charge carriers between the hot and cool areas produces a voltage differential. A thermoelectric generator (TEG), or a Seebeck generator, is a solid-state device that uses the Seebeck effect to directly convert heat flow, or temperature differential, into electrical energy. This experimentation on the design and construction of a 1kW thermoelectric generator using solar energy with a parabolic trough collector is composed of a parabolic trough collector framed from a plain mirror with an aperture area of m 0.27m2 that was used to concentrate sunlight onto a copper receiver plate with an area of 10×10 cm^2. A power of 0.9kW was achieved by connecting 7 numbers of thermoelectric modules (TEM) generating 2.4V and 469 mA at an average temperature gradient of 60^{o}C.

Keywords: Thermoelectric-generator, Seebeck-effect, Conductor, Thermoelectric-module, Temperature-gradient.

1.0 INTRODUCTION

In light of the current global energy crisis, energy harvesting is essential. The available energy can meet the majority of the world's energy needs, but they originate from non-renewable sources (Dresselhaus and Thomas, 2001). The non-renewable energy sources known as fossil fuels are running out quickly since the amount of energy consumed by each person is rising daily. Furthermore, one of the main causes of the issue is population expansion (Lertsatitthanakorn et al., 2012). Fossil fuel days are limited and fast diminishing in deposits around the planet and alternate forms of energy production have to be researched (Shafiee & Topal, 2009) .

It is imperative to find a way to capture energy and convert it efficiently without risking the depletion of other resources. More than just being a non-renewable energy source, the fundamental issue with fossil fuels has always been their extreme pollution and global warming (Midilli, 2006). As a result, green energy options need to be carefully studied.

According to (Hassen, 2000) the remedy is to manage energy efficiently to resolve these lingering issues. Given the current environmental problem and the world's energy needs, a gradual and seamless shift to green energy is essential. There is nothing more alluring than the idea, development, and application of a plentiful, easily accessible green energy source. Thermal energy has several benefits and is a very appealing energy source (Heremans, 2002). Waste thermal energy from industrial processes, automobile exhaust, and residential heating is constantly present (Snyder and Toberer, 2008). If the temperature differential in the environment is effectively captured, energy harvesting may result.

Thermo-electrics is the study of the transformation of thermal energy into electrical energy and vice versa (Goldsmid, 2009). The initial discovery of this phenomenon was made by Seebeck in the year 1821. The idea of using the temperature differential between two surfaces to produce energy, known as the Seebeck effect, has great promise as a clean, alternative energy source. Although this idea has been around since the turn of the century, it has recently received a lot of attention (Dresselhaus , 2007).

Despite these devices' limited efficacy and high cost, thermoelectrics is a field that is revolutionizing energy gathering due to its many advantages over conventional approaches. Green energy comes from thermoelectric devices. They have no moving parts, emit no emissions, and make no byproducts when they operate, equally, they do not generate noise during operation (Poudel, 2007).

If this technology is to be implemented, many of the frequent questions that need to be answered include no maintenance, high reliability, and long lifespan. All of these criteria are partially addressed by thermoelectric devices now on the market.

The ability to manufacture large and small devices with thermoelectric technology is another huge benefit. These devices have a significant impact, particularly in this era of nanotechnology. By utilizing the Peltier or Seebeck effects, respectively, thermoelectrics can be utilized to control temperature or to generate electricity.

1.1 SEEBECK EFFECT

It was believed until recently that Thomas Seebeck made the discovery of the Seebeck effect. It is currently believed that Alessandro Volta discovered the Seebeck effect 27 years before Thomas Seebeck (Pastorina, 2009). In 1794, Alessandro Volta performed experiments in which he formed a u shape out of an iron rod. One end of the rod was immersed in hot water to heat it up. The muscles of a dead frog leg flexed when a current was run through it via an electrical connection to the unevenly heated rod. This is thought to be the first instance of the Seebeck effect being demonstrated.

Thomas Johann Seebeck made the new discovery that electricity can be generated by a temperature gradient that forms between two dissimilar conductors in 1821. The diffusion of charge carriers is caused by heat flow, which is the fundamental mechanism underlying the thermoelectric effect in conducting materials. In turn, the movement of charge carriers between the hot and cool areas produces a voltage differential. (Baranowski et al., 2013)

When two dissimilar metals, A and B, are linked, a temperature differential between the junctions (J1 and J2) causes a net voltage to be generated (Figure 1). One connection will be hotter than the other since it is heated to a higher degree than the other.

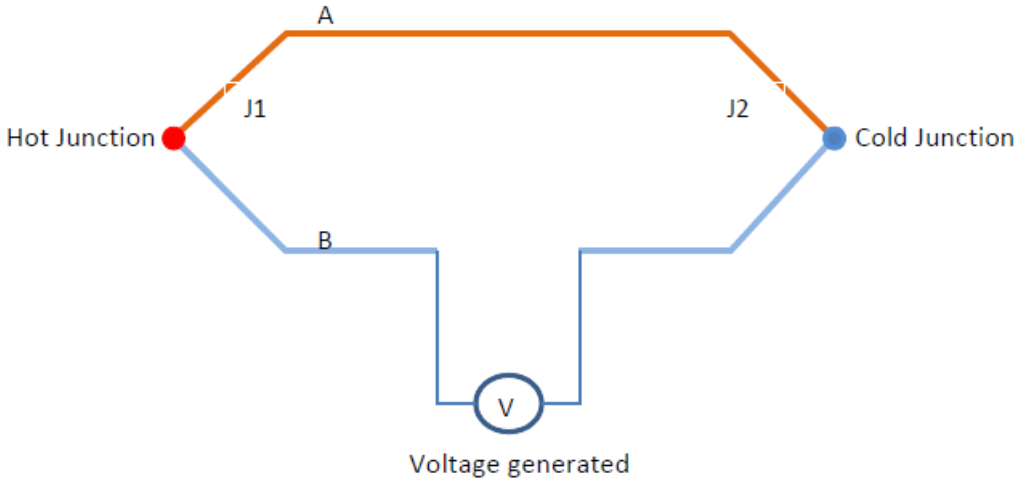


Figure 1: Seebeck effect using dissimilar metal junctions

An additional characteristic of distinct bulk materials is the Seebeck effect. This idea serves as both the foundation for power generation and a means of measuring temperatures. The Seebeck coefficient, or the voltage generated by the temperature gradient divided by the temperature difference, is given by equation 1.

$$S = -\frac{dV}{dT} \quad (1)$$

where, S – coefficient of Seebeck, V is voltage and dT is change in temperature

Although the Seebeck coefficient is measured in volts per degree, it is more frequently expressed in micro-volts per degree Kelvin (μV/K). The sign of S is positive for p-type materials and negative for n-type materials.

1.2 PELTIER EFFECT

The Seebeck coefficient is commonly stated in micro-volts per degree Kelvin (μV/K), even though it is measured in volts per degree. For p-type materials, the sign of S is positive; for n-type materials, it is negative.

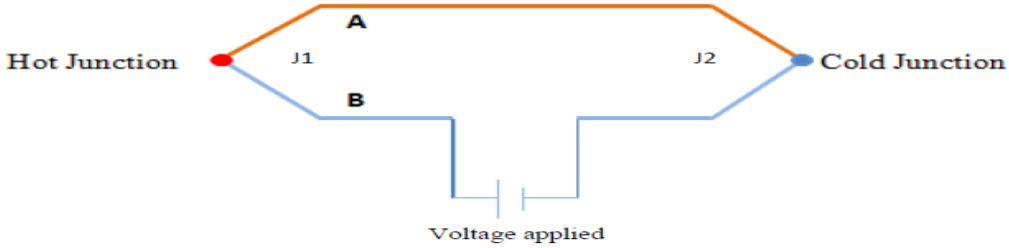


Figure 2: Effect Peltier

The Peltier effect is used in thermoelectric refrigerators, heated automobile seats, and cooling systems in electronics. The Peltier effect is used by thermoelectric devices to cool computer microchips and processors that generate a lot of heat over long periods of time.

1.3 THERMOELECTRIC MATERIALS

Thermoelectric materials convert temperature changes into electric voltage, allowing them to directly convert heat into electricity. These materials must have high electrical conductivity (σ) and low heat conductivity (κ) in order to be deemed suitable thermoelectric materials.Due to poor thermal conductivity, when one side is heated, the other stays cold, which helps create a high voltage in a temperature gradient. The amount that electron flow varies in response to a temperature differential across a material is indicated by the Seebeck coefficient (S) (Dresselhaus , 2007).

1.4 THERMOELECTRIC MODULES

A thermoelectric module is a circuit that uses thermoelectric components, or materials that directly convert heat into electricity. Two distinct thermoelectric materials, an n-type semiconductor (which has negative charge carriers) and a p-type semiconductor (which has positive charge carriers), are connected at their ends to create a thermoelectric module. Direct electric current flows through the circuit when there is a temperature difference between the ends of the materials. (Lertsatitthanakorn et al., 2012)

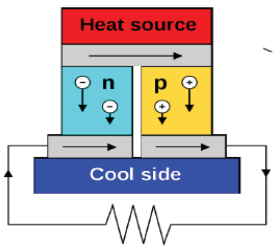


Figure 3: Electric circuit of Thermoelectric

In power generation applications, thermoelectric modules are utilized since they have to function in harsh mechanical and temperature environments. Since the modules operate in an exceptionally high temperature gradient, they are subjected to severe thermally produced stresses and strains over lengthy periods of time. Additionally, a lot of heat cycles might create mechanical wear in them.

In order for the joints and materials to withstand these extreme mechanical and thermal stresses, careful selection is required. Additionally, the two thermoelectric materials need to be positioned in the module such that they are thermally coupled in parallel yet electrically connected in series. The geometry of a thermoelectric module's design has a big impact on how efficient it is.

Thermoelectric Generator Using Parabolic Trough Mirror as a Collector

The parabolic trough mirror has several purposes, but its main one is to collect solar energy from a large surface area and focus it into a smaller area. For best results and economic output, the focal length of the parabolic trough mirror must be determined. The intersection of all the rays is the focus point, and it is there that the most heat is generated (Dresselhaus, 2007).

A parabolic trough mirror's focal length is located in the middle of its diameter and at a certain distance from the mirror's surface. The focal length (20) can be calculated using the formula (2). As seen in Figure 4, focal length is indicated by f, mirror depth is represented by d, and mirror diameter is indicated by D.

$$F=\frac{D2}{(4*4)d} \quad (2)$$

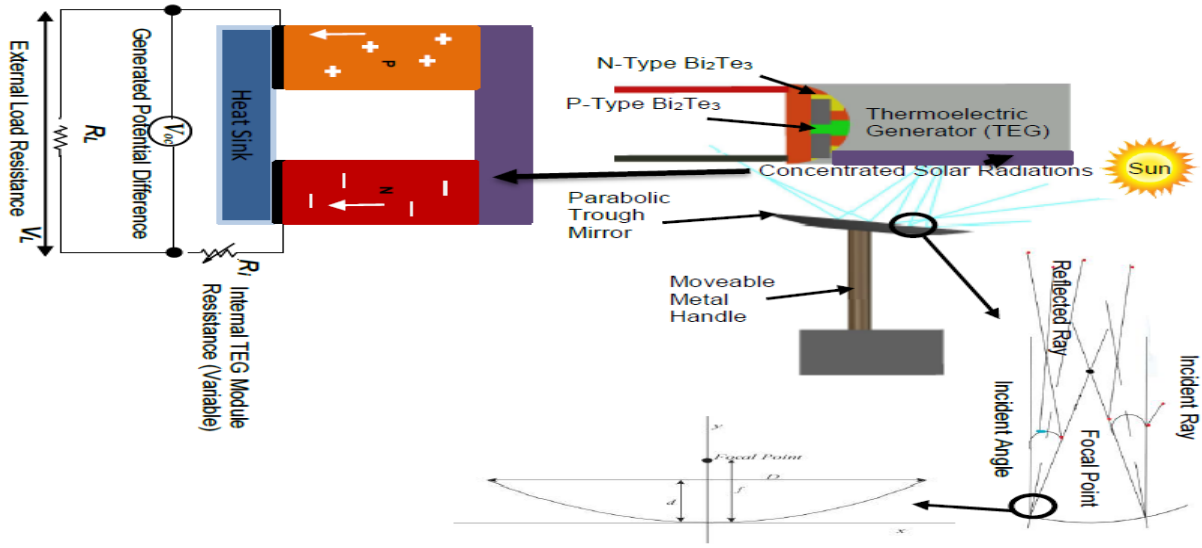


Figure 4: Illustration of a Parabolic Mirror with a Concentrator

A parabolic mirror's focal point is where the reflected rays converge when an incident ray strikes it (Figure 4). As a result, all of the reflected light passes through the focus point and concentrates there, where the mirror's maximum temperature is gathered for application to the TEG's hot side. Because the focal length lies on the focal axis of the mirror, all parallel rays that arrive at its surface are reflected and pass through the focus point. There is equality between the angles of incidence and reflection (Snyder and Toberer, 2008)

Solar Collector Types

Many types of solar collectors are used to collect solar energy. Experimental investigations relevant to altering elements such as the working medium, surrounds, and size are taken into consideration in order to optimize the design (Goldsmid, 2009). Analysis of energy has been done on several kinds of solar collectors. The majority of them were in the area of solar collectors using flat plates. The second most popular field of research concerned hybrid thermal and photovoltaic collectors. Additionally, a few investigations on evacuated tube collectors and parabolic trough collectors have been conducted. Numerous researchers have examined and refined parabolic dish collectors, demonstrating their excellent energy efficiency. Compound parabolic collectors, heat pipe collectors, and hollow receivers are other collector forms that are not as frequently studied (Midilli, 2006).

This study is aimed at the design and construction of a 1kW Thermoelectric generator using a parabolic trough collector.

.

Figure 5: Photograph of Parabolic Trough Collector

2.0 METHODOLOGY

A model of a parabolic trough solar collector was created and evaluated. The design consists of a hardwood support frame, copper sheeting for the system, glass strips for the reflector sheet, and galvanized sheeting for solar radiation absorption.

Dimension Parabolic Trough Collector

Simple parabolic equations were utilized in the collector's design. The aperture width (W) of the system is 0.65 meters, and the focal length (f) is 0.88 meters from the vertex (V) according to the reflector's design. The Cartesian equation for the design system was derived from equation (3) below (Handayani & Ariyanti, 2012).

$X^2=4fy$ (3)

Equation (4) below was used to determine the height of the parabola in terms of the focal length and aperture diameter..

$$hc = \frac{w2}{16f} \quad (4)$$

$$\text{Tan}\frac{\Psi rim}{2} = \frac{W}{4f} \quad (5)$$

Where Ψrim is the angle of the rim

Figures 6 and 7 illustrate how the cross section for the parabolic trough was created utilizing a geometrical relation of the parabolic section. The parabolic trough, which has an effective aperture area of 0.2711 square meters and measures 1.844 meters in length and 0.50 meters in aperture width, was made using sheets of planar mirror. The values of the aperture area, curve surface area, focal length, parabola height, aperture width, receiver outer diameter, and concentrator length were determined using the above equations. The collector height hc and the focal length f are nearly equivalent. The manufactured system's specifications are listed below.

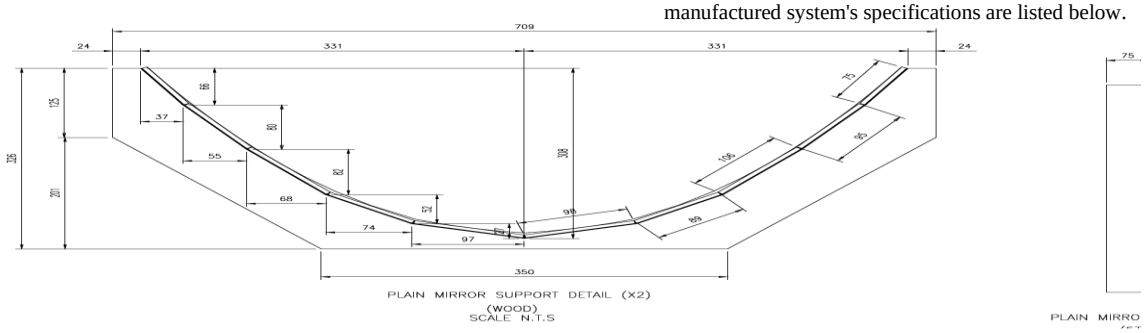


Figure 6: Drawing of produced Parabolic Trough

Figure 7: Produced Parabolic Trough Collector

Table 1: Specifications for the produced Parabolic Trough Collector

Constraint	Identity	Result
Aperture area (m2)	Aa	0.2711
Length (m)	L	1.844
Curve area (m2)	As	1.3655
Focal length	F	0.88

Numerical Approach employed in predicting the Electrical Potential of the TEG Under Use

The output electrical power of the thermoelectric module (TEM) increases with the temperature differential between its focused solar-heated hot side and cold side.

(Saim et al., 2020). To keep one side colder than the other, an efficient cooling system was needed because the temperature differential is difficult to achieve. The greater temperature differential indicates a relatively low total efficiency for the TEG. Thermoelectric materials are mostly responsible for setting the working temperature range of the device, and a rise in overall efficiency is largely dependent upon them. This range is often classified into three categories: The base of commercial TEGs is Bismuth (Bi) composite with Tellurium (Te), Antimony (Sb), and/or Selenium (Se) for less than 450K; lead (Pb) and Telluride (Te)

composites for 600K-850K; and germanium (Ge) and silicon (Si) composites for 900K-1300K (Saim et al, 2020).
Because of its appropriate working temperature range of <450 K, Bi2Te3-based TEG is used. They are frequently packed between heat-conductive components and connected in series to achieve appropriate power, providing heat from the concentrated solar heated source and cooling from the heat sink.

Experimental Procedure

TEG (SP1848-27145), measuring 40 mm by 40 mm by 3.4 mm, was utilized in these studies. It had both P and N-type elements (Bi2Te3), which were coupled to copper lead and insulated with Teflon. To find the right process for converting focused radiation into electrical power, several tests were carried out. To prevent thermal damage to the TEG from heat applied directly to its hot side and to reduce heat loss by reflection, a heat-resistant (220°C) black paint was applied to a flat metal connected to the TEG after it had been polished with sandpaper. During the first testing, a modest cooling fan was employed to support the hot plate with an external heat sink to keep the cold side temperature below 308 K (Dresselhaus, 2007). The hot plate can reach temperatures of up to 419 K, and its electrical performance was evaluated in a range of thermo-electrical scenarios. A heat-resistant (220°C) black paint was applied to a flat metal connected to the TEG after it had been polished with sandpaper. During the first testing, a modest cooling fan was employed to support the hot plate with an external heat sink to keep the cold side temperature below 308 K (Dresselhaus, 2007). The hot plate can reach temperatures of up to 419 K, and its electrical performance was evaluated in a range of thermo-electrical scenarios. The open-circuit voltage (Voc) test is carried out if the circuit is not linked to an external load resistor (RL). The short-circuit current test was performed after shorting the output leads with a copper conductor. When RL=Re, the maximum power performance was measured using a variable resistor of 100 Ω, and the TEG leads were connected to the external load resistor (RL) for the full-load test. The digital multimeter was used in these initial tests to examine the hot and cold sides of the TEG by connecting it to its input leads (Baranowski et al., 2013). A negative voltage measurement indicated the chilly side and a positive voltage reading indicated the heated side. In this way, the hot and cold sides of TEG are distinguished. Once the sides are identified, thermal glue is used to attach the TEG to the black-coated surface, with the hot side facing the surface that will be exposed to high heat. Once a fan and heat sink are attached to the cold side of the TEG to maintain its cool, it is prepared for experiments.

3.0 RESULTS AND DISCUSSION

Performance evaluation was carried out on the produced Thermoelectric Generator in order to ascertain its efficiency and effectiveness. The results of the readings generated for two consecutive days are shown in Tables 2, 3 and Figure 8, respectively.

Table 2: Results for a test carried out on day 1

Cold Temperature (ºC)	Hot Temperature (ºC)	Temperature Gradient (ºC)	Voltage (V)
24	44	20	0.97
37	77	40	1.8
40	100	60	2.4
42	122	80	3.6
50	150	100	4.8

Table 2: Result of tests carried out on day 2

Cold Temperature (ºC)	Hot Temperature (ºC)	Temperature Difference (ºC)	Voltage(ºC)
22	43	21	1.02
38	80	42	1.89
40	105	65	2.6
43	132	89	4.00
48	153	105	5.04

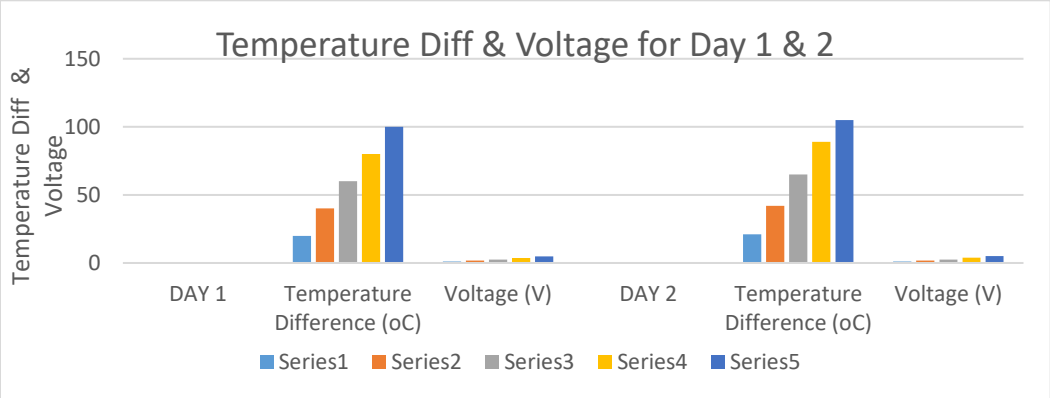


Figure 8: Graph of Temperature Difference and Voltage for Day 1 and 2

3.1 DISCUSSION

Based on the outcome attained for each increase in the temperature differential between the hot and cold sides. This illustrates that voltage exchange increases with thermal energy production. Therefore, this analogy can be used as a model for converting thermal energy into additional electrical power. The Seebeck effect applies the first law of thermodynamics to charging devices by converting temperature difference into voltage supply. The test conducted on the first day yielded a maximum temperature of 150°C for hot temperatures and 50°C for cold temperatures, respectively, at a voltage of 4.8V. However, the results of the test conducted on day two show that the maximum temperature at 5.04V was 153°C for hot temperatures and 48°C for cold temperatures.

4.0 CONCLUSION

A thermoelectric generator using Parabolic Trough Collector was designed and constructed. The constructed generator produces 2.4 V and 469 mA at an average temperature of 60ºC. More so, an average of 0.9kW of power was produced by the constructed thermoelectric generator.

REFERENCES

Baranowski, L., Toberer, E., & Synder, S. (2013). The Misconception of Maximum Power and Power Factor in Thermoelectrics. *Journal of Applied Physics*, 116-123.

Dresselhaus, M. (2007). New Directions for Low-Dimensional Thermoelectric Materials. *Springer*, 100-110.

Dresselhaus, M., & Thomas, I. (2001). Alternative energy technologies. *International Journal of Engineering*, 332-337.

Goldsmid, H. (2009). Introduction to Thermoelectricity. *Springer*, 52-60.

Handayani, N., & Ariyanti, D. (2012). Potency of Solar Energy Applications in Indonesia. *International Journal of Renewable Energy Development*, 34-38.

Hansen, J. (2000). Global warming in the twenty-first century: An alternative scenario. *Proceedings of the National Academy of Sciences*, (pp. 9874-9881).

Heremans, J. (2002). *Thermoelectric Power of Bismuth Nanocomposites.* India: Physical Review Letters.

Lertsatitthanakorn, C., Rungsiyopas, M., Therdtothin, A., & Soponronarit, S. (2012). Performance study of a double-pass thermoelectric solar air collector with flat plate reflectors. *Journal of Electronic Materials*, 997-1000.

Midilli, A. (2006). Green energy strategies for sustainable development. *Journal of Energy Policy*, 3622-3632.

Pastorina, G. (2009). Allessandro Volta And His Role in Thermoelectricity. *Journal of Thermoelectricity*, 1604-1634.

Poudel, B. (2007). *A study on thermoelectric properties of Nanostructured Bulk Materials.* Boston: Boston College.

Saim, M., Maekele, M., Takao, K., Ali, R., Shanwen, Z., Ahmed, A., . . . Essam, M. (2020). Experimental and Theoritical Evaluation of Parabolic Trough Mirror as Solar Thermal Concentrator to Thermoelectric Generator. *International Journal of Solar and Vacuum Technology*, 126-135.

Shafiee, S., & Topal, E. (2009). When will fossil fuel reserves be diminished . *Journal of Energy Policy*, 182-188.

Snyder, & Toberer, E. (2008). *Complex thermoelectric materials.* London: National Materials.