1 Introduction

Hura crepitans (sand box tree) is a perennial crop of the family Euphorbiaceae. It is one of the tropical crops grown in North and South America, also common to Africa especially Nigeria where it is abundant. Oil extracted from its seeds contains a volatile colourless liquid called “Hurin”, though a vegetable oil, it is poisonous if ingested, hence the oil remains underutilised.

Recent researches have shown that huracrepitan seed oil has been utilised as an industrial raw material in pharmaceuticals, as well as a prospect for biodiesel production. Muhammed et al. analysed the fatty acid composition (wt%) of huracrepitan seed oils, from the assessment, it indicates that the oil is a good base material for the plastic and paint industries.

Previous works have been reported on the extraction of oil from huracrepitan seed. Optimisation of the process parameters for the extraction of oil from Huracrepitan seed using response surface technique has been reported by Oniya et al., but its comparison with other predictive models and analytically establishing a non-linear relationship between input and output variable of the extraction process have not been reported.

Nowadays, researchers have asserted the potential of soft computing models such as Artificial Neural Networks (ANNs) and Adaptive Neuro-fuzzy Inference System (ANFIS). ANN is a modelling tool inspired by biological neural networks, while ANFIS integrates the strengths of ANN and FL (fuzzy logic). The use of ANN and ANFIS has gained increasing applications where the dependency between dependent and independent variables is either unknown or very complex. They provide more accurate results for process control of a complicated system than conventional mathematical models.

Moreover, a limited number of researchers have modelled the extraction of oil from plant seeds using soft computing techniques.

Onoji et al. compared RSM and ANN in modelling and optimising the rubber seed oil extraction process, both models were effective in describing the parametric effect of the considered operating variables on the extraction, however, ANN described the effect more accurately than RSM model. Olajide et al. applied RSM and ANFIS to optimise oil yield from shea kernels in a hydraulic press, considering moisture content, heating temperature, heating time, applied pressure, and pressing time on oil yield; RSM gave a better prediction performance having $R^2$ of 0.9998, while ANFIS had $R^2$ of 0.9865. Hence both models gave a good prediction.

Eletta et al. modelled and optimised oil extraction from Luffa cylindrica seeds using a binary solvent mixture, the predicted oil yield values from ANN model was more accurate than that of RSM when compared with experimental values. Therefore, this study is aimed at filling the gap in scientific research by predicting oil yield from huracrepitan seed using ANN and ANFIS.

2 Materials and method

2.1 Sandbox (Hura crepitans) seeds

Sandbox (Hura crepitans) seeds/fruits were collected from different locations in Nigeria (Auchi in Edo State and Umudike, Abia State). Dried seeds were collected from under the tree and in some cases, the matured and dried
fruits were harvested from a tree in bulk quantity. The seeds were removed naturally by sun-drying, which makes the pods break on their own. The seeds were sundried for one week, and subsequently, oven-dried 8 h at a temperature of 60 °C until constant weight to minimise moisture content before extraction. The extraction of the oil was carried out at the analysis laboratory of the Department of Chemical Engineering, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria, using the solvent extraction method. It was done with a Soxhlet apparatus of 250 cm³ capacity using n-hexane of analytical grade as the solvent. The parameters were 10–20 g Hura crepitans (sand box) seed, extraction time of 4–6 h, and extraction temperature in the range of 60 °C and 80 °C, and the ratio of the solute (biomass) to solvent (n-hexane) from 0.05 to 0.10. The solvent used was recovered at every interval through a distillation process or the use of a rotary evaporator, and the actual oil obtained was weighed, the experiment was repeated for other parameters, and the percentage yield was calculated using Eq. (1) below:

\[ Y = \frac{M_o}{M_s} \times 100 \]  

where: \( Y \) = oil yield (%); \( M_o \) = mass of oil extracted (g); and \( M_s \) = mass of Hura crepitans seed (g).

### 2.2 Experimental design

A Box Behnken experimental design using response surface methodology developed in Design expert version 6.0.8, consisting of three factors and three levels basis, which generated 17 experimental runs, was employed in the process. The three independent factors considered during this study included extraction time, extraction temperature, and seed/solvent ratio; the response is oil yield. The quadratic model developed depicts the interaction between the oil yield (Y) and the coded values of the independent variables A, B, and C (temperature, time, and seed/solvent ratio).

\[ Y = 52.73 + 6.92A + 3.08B - 4.03C - 7.21A^2 - 3.25B^2 - 0.055C^2 - 8.30AB + 5.36AC - 1.03BC \]  

where Y = oil yield (%).

### 2.3 ANN model development

Artificial neural network (ANN) architecture was developed in MATLAB 8.4 (R2014b) software environment where the training, validation, and testing of the ANN model was carried out. The three-layer ANN, (Fig. 1), comprised a tangent sigmoid transfer function (tansig) at hidden layer, a linear transfer function (purelin) at the output layer, and Levenberg–Marquardt back-propagation algorithm with 1000 iterations. The input layer corresponded to the three experimental parameters, which were solute/solvent ratio (g mol⁻¹), temperature (°C), and time (h). The output layer was oil yield. All the data derived from the convective extraction of oil were randomly divided into three groups (training, validation, and testing) with a ratio of 70%, 15%, and 15%, respectively. In this study, ten neurons were used as a default testing to determine the perfect algorithm for the prediction. One to fifteen neurons in the hidden layer and one neuron in the output layer were applied, and the data used were obtained from the multiple factors at a time experiment.

### 2.4 Neuro-fuzzy Exhaustive Search

**Parametric Technique**

ANFIS exhaustive search programming codes were written in MATLAB 8.4 (R2014b) environment and implemented for the selection of the set of one and two variable inputs combination that had the most and the least influence on the oil yield. The Root Mean Squared Error (RMSE) was used as a performance indicator for the exhaustive search technique. The architecture of ANFIS consisted of 5 layers, as shown in Fig. 2. In Fig. 2, square nodes (adaptive nodes) show adjustable parameters that are to be learned, whereas the circle nodes (fixed nodes) are fixed parameters. A common rule set with two fuzzy if-then rules is as follows:

Rule 1: If \( x \) is \( A_1 \) and \( y \) is \( B_1 \), then \( f_1 = p_1x + q_1y + r_1 \) \hspace{1cm} (3a)

Rule 2: If \( x \) is \( A_2 \) and \( y \) is \( B_2 \), then \( f_2 = p_2x + q_2y + r_2 \) \hspace{1cm} (3b)

where \( A_1, B_1 \) are linguistic terms that are user-defined and represent a range of values. The sequence and functions of the layers are as follows:

**Layer 1**: Square node equipped with a node function:

\[ Q_i^1 = \mu_{A_i}(x) \]  

Assuming x and y are the two typical input values fed at the two input nodes, which then transform those values to the membership functions such as triangle, generalised bell-shaped, Gaussian membership, etc., where, \( Q_i^1 \) is the membership function of \( A_i \), and \( x \) is the input parameter to
the node. \( A_i \) is the linguistic label connected with the node function.

Layer 2: This node multiplies the incoming signal and sends the product out. Each node output is the firing strength of a rule:

\[
w_i = \mu_{a_i}(x) \cdot \mu_{a_i}(y), \quad i = 1, 2 \tag{4a}
\]

Layer 3: circle node. Node calculates the ratio of \( i \)-th rule’s firing strength to the sum of all rules’ firing strengths:

\[
w_i' = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \tag{4b}
\]

Layer 4: Square node with node function:

\[
O_i^4 = w_i' f_i = w_i' (p_i x + q_i y + r_i) \tag{5a}
\]

\( p, q, r \) – parameter set (consequent, linear, parameters)

Layer 5: circle node. This node computes the overall output as a summation of all incoming signals.

\[
O_5^5 = \text{overall output} = \sum w_i' f_i = \frac{\sum w_i' f_i}{\sum w_i'} \tag{5b}
\]

\[
\text{MSE} = \frac{1}{p} \sum_{p=1}^{p} (d_p - O_p)^2 \tag{5}
\]

\[
\text{RMSE} = \sqrt{\frac{1}{p} \sum_{p=1}^{p} (d_p - O_p)^2} \tag{6}
\]

\[
R^2 = 1 - \frac{\sum_{p=1}^{p} (d_p - O_p)^2}{\sum_{p=1}^{p} (O_p)^2} \tag{7}
\]

3 Results and discussion

3.1 Physicochemical properties analyses of the extracted HCSO

The physicochemical properties of the extracted *huracrepitan* seed oil were analysed and most of the properties conformed to the ASTM standards, the degree of unsaturated triglycerides present in the oil was 46.50 \( I_2 \) \( g^{-1} \), hence, the oil can be modified to biobased resins for plastic and paint industries.

| Physicochemical properties | Result | ASTM standard |
|----------------------------|--------|---------------|
| viscosity at room temperature | 4.58 mm² s⁻¹ | 1.90–6.0 |
| specific gravity | 1.450 | 0.87–0.98 |
| peroxide value | 0.592 mg g⁻¹ | 10 |
| saponification value | 186.70 mgKOH/g | 195–205 |
| specific density | 0.853 g cm⁻³ | 0.897–0.907 |
| free fatty acid value | 0.380 mgKOH/g | 3.5 |
| iodine value | 46.50 gI² g⁻¹ | – |
| colour | golden yellow | – |

3.2 ANN model simulation

Eleven backpropagation (BP) algorithms were compared to select the best suited BP algorithm. For all BP algorithms, a three-layer ANN with a tangent sigmoid transfer function (\textit{tansig}) at the hidden layer and linear transfer function (\textit{purelin}) at the output layer was used. Ten neurons were used in the hidden layer for all BP algorithms, the benchmark comparison displayed loss on the optimality of the estimates/results produced by some BP training algorithms. As shown in Table 3, the Bayesian regularisation was found as the best of all 11 BP having the smallest MSE of 1.01810E-11. However, the \textit{traingd} produced the greatest error of 321.1121. The loss on the optimality of the estimates/results produced by some BP training algorithms can be attributed to the combinatorial nature and non-linear structure of the experimental data. Hence, the complexity analysis of the problem was validated by the results.
of the various training algorithms used in the benchmark comparison.

Table 3 – Comparison of 11 backpropagation with 10 neurons in the hidden layers with oil yield as output

| Algorithm                  | MSE         | $R^2$       | IN  |
|----------------------------|-------------|-------------|-----|
| Levenberg-Marquardt        | 3.44166E-8  | 9.99999E-1  | 1000|
| Bayesian regularisation    | 1.01810E-11 | 9.99999E-1  | 1000|
| scaled conjugate gradient | 1.63549E-0  | 9.78837E-1  | 15  |
| trainlp                    | 0.1115      | 0.9985      | 66  |
| traincgf                   | 0.5667      | 0.99427     | 20  |
| traincgp                   | 0.0916      | 0.99894     | 31  |
| traincgb                   | 0.1860      | 0.99734     | 24  |
| trainbfg                   | 0.3339      | 0.99983     | 47  |
| trainoss                   | 0.2449      | 0.99736     | 42  |
| traingd                    | 321.1121    | 0.22155     | 6   |
| traindx                    | 2.2633      | 0.98083     | 92  |
| traindm                    | 137.2230    | 0.27611     | 6   |

Table 4 – Varying of neurons (1–15) using the Bayesian Regularisation algorithm for oil yield

| Neuron | MSE         | $R^2$       | IN  |
|--------|-------------|-------------|-----|
| 1      | 1.14389E-6  | 9.99999E-1  | 1000|
| 2      | 1.23605E-8  | 9.99999E-1  | 1000|
| 3      | 9.96781E-10 | 9.99999E-1  | 1000|
| 4      | 4.81631E-10 | 9.99999E-1  | 1000|
| 5      | 1.21241E-11 | 9.99999E-1  | 1000|
| 6      | 3.41983E-11 | 9.99999E-1  | 1000|
| 7      | 8.53059E-12 | 9.99999E-1  | 1000|
| 8      | 3.88252E-11 | 9.99999E-1  | 1000|
| 9      | 1.00148E-10 | 9.99999E-1  | 1000|
| 10     | 1.93080E-12 | 9.99999E-1  | 914 |
| 11     | 4.29907E-11 | 9.99999E-1  | 1000|
| 12     | 5.63192E-13 | 9.99999E-1  | 372 |
| 13     | 1.76670E-11 | 9.99999E-1  | 1000|
| 14     | 6.24266E-12 | 9.99999E-1  | 1000|
| 15     | 1.84144E-11 | 9.99999E-1  | 1000|

3.3 Optimisation of the ANN structure

The optimal architecture of the ANN model and its parameter variation were determined based on the minimal value of the MSE of the training and prediction set from the Bayesian regularisation algorithm. In optimisation of the network, one neuron was used in the hidden layer as an initial guess. With an increase in the number of neurons, the network gave several local minimum values and different MSE values were obtained for the training set. Table 4 shows that the minimum MSE (5.63192E-13) was attained with 12 neurons for oil yield. With a further increase in the number of neurons beyond the point, the minimum MSE was attained for each of the outputs, had the best correlation coefficient, $R^2$, which is closest to one.

Table 5 – Exhaustive search result of one-input variable ANFIS model for oil yield

| No. of Input | Input variable | RMSE Training | RMSE Checking | Ranking |
|--------------|----------------|---------------|---------------|---------|
| 1            | sol/solvent    | 5.6438        | 6.3536        | 3       |
| 1            | temperature    | 4.3086        | 4.2550        | 1       |
| 1            | time           | 5.5401        | 6.2678        | 2       |

Table 6 – Exhaustive search result of two-input variable ANFIS model for oil yield

| No. of Input | Input Variable | RMSE Training | RMSE Checking | Ranking |
|--------------|----------------|---------------|---------------|---------|
| 2            | sol/solvent-temp| 3.3301        | 3.5756        | 2       |
| 2            | sol/solvent-time| 4.8848        | 4.7303        | 3       |
| 2            | temperature/time| 2.7252        | 2.6252        | 1       |

The ANFIS models using different input variable combinations were investigated with exhaustive search method to determine the input variable that has the greatest effect on the extraction using RMSE as the performance indicator. Table 5 shows exhaustive ANFIS model result with a single input variable, and it was observed that temperature possessed the least RMSE; this indicated that this input variable was the most relevant variable to the response. Also, it can be seen from Table 6, that temperature and time were the best two input variables that mostly affected the oil yield. However, the two inputs were considered for FIS structure.

To obtain the best prediction of huracrepitan seed oil extraction, the developed ANFIS structure was simulated at various input membership functions (mf), such as gauss mf, gauss2 mf, gbell mf, tri mf, trap mf, psg mf, and dsig mf. The correlation coefficient ($R^2$) and the root mean square error (RMSE) were used as the statistical criteria to evaluate the degree of reliability of the network.

3.4 Prediction efficiency of ANFIS model for oil yield

Tables 7 and 8 summarise the different input membership function type for linear and constant output MF . The tables also give the RMSE computed for all the model structures considered, on the training and the corresponding correlation coefficient ($R^2$) between the measured and computed
output in which RMSE and $R^2$ are the statistical criteria to judge the performance of the model. The effects of different input membership functions (MF), such as \textit{gbell}, \textit{gauss}, \textit{gauss2}, \textit{trap}, \textit{pi}, \textit{dsig}, and \textit{psig} on oil yield were tested and verified with a single output MF type “linear” and “constant” by training to determine the best input MF.

It was observed that MSE ranged from 0.42385 to 0.48885 with corresponding $R^2$ ranging from 0.3429 to 0.34383, as depicted in Table 7. The ranges of $R^2$ values are not close to 1 and MSE value of \textit{tri} membership function is the closest to zero for the linear output membership function, the \textit{pi} had the best prediction for $R^2$. For the constant output membership function, the \textit{gauss2} \textit{mf} had the best prediction for $R^2$, while the \textit{tri} \textit{mf} had the lowest MSE value. It was observed that the MSE ranged from 0.69012 to 2.103 with corresponding $R^2$ ranging 0.35199–0.36945, as depicted in Table 8.

The $R^2$ values are closer to 1 than the linear output membership function, while MSE values are close to zero, but that of linear output membership function is closer. From a statistical performance point of view, the prediction of oil yield for \textit{huracrepitan} seed is poor compared to the other existing investigation on ANFIS modelling. \textsuperscript{20–22}

### 3.5 Prediction model comparison

The evaluation of the predictive capabilities of ANN and ANFIS for oil yield was assessed using statistical parameters such as MSE (mean square error) and $R^2$ (correlation coefficient). The results showed that the Bayesian regularisation algorithm optimised at 12 neurons yielded the best prediction of 0.9999 ($R^2$) and 5.63192E-13(MSE) for ANN model, while \textit{gauss2} \textit{mf} had the best $R^2$ prediction of 0.36945 and \textit{tri} \textit{mf} had the lowest (MSE) of 0.42331 for ANFIS model, respectively. These statistical comparison results suggest that ANN performed better than ANFIS model.

### 4 Conclusions

This study provided some insight into the effects of process conditions variation of temperature, time, and solute/solvent ratio on \textit{huracrepitan} seed oil extraction. The results obtained from the physicochemical properties of the oil revealed potential to serve as a raw material for the synthesis of bio-based materials/biopolymers. This deduced inference could be attributed to the high iodine value obtained for the oil. The parametric analysis using exhaustive search showed that temperature had the highest influence on the oil yield. The work also showed that the combination of temperature and time had the most significant effect on the oil yield. The statistical comparison for ANN and ANFIS results suggest that ANN predicted oil extraction from \textit{huracrepitan} seed better than ANFIS.

### ACKNOWLEDGEMENTS

The authors acknowledge the team of Unit Operations Laboratory, Department of Chemical Engineering, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria for their contribution.

### List of abbreviations

- ANN: artificial neural network
- ANFIS: adaptive neuro-fuzzy inference system
- MF: membership function

### References

- M. A. Foworola, A. A. Akindahunsi, Nutritional Quality of Sandbox Tree (\textit{Hura crepitans} L.), J. Med. Food. \textbf{10} (2007) 139–164, doi: https://doi.org/10.1089/jmf.2005.062.
- O. O. Oniya, J. O. Oyelade, O. Ogunkunle, D. O. Idowu, Optimization of Solvent Extraction of Oil from Sandbox Kernels (\textit{Hura crepitans} L.) by a Response Surface Method, Energy Policy Res. \textbf{4} (2017) 36–43, doi: https://doi.org/10.1080/23815639.2017.1324332.
- A. Adewuyi, P. O. Awolade, R. A. Oderinde, \textit{Hura crepitans} Seed Oil: An Alternative Feedstock for Biodiesel Production, J. Fuel. \textbf{97} (2014) 1086–1091, doi: https://doi.org/10.1155/2014/464590.
Svjedočenje o učinku prinosa ulja su: ANN (R2 = 0,999, MSE = 5,63192E-13), ANFIS (R2 = 0,36945, MSE = 0,42331). Rezultati su pokazali da je ANN dao bolje predviđanje od ANFIS-a.