Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 3, June 2025, pp. 2393~2403
ISSN: 2302-9285, DOI: 10.11591/eei.v14i3.8901

Unmasking effects of feature selection and SMOTE-Tomek in tree-based random forest for scorch occurrence detection

Margaret Dumebi Okpor[1], Kizito Eluemnor Anazia[2], Wilfred Adigwe[3], Ejaita Abugor Okpako[4], De Rosal Ignatius Moses Setiadi[5], Arnold Adimabua Ojugo[6], Felix Omoruwou[7], Rita Erhovwo Ako[6], Victor Ochuko Geteloma[6], Eferhire Valentine Ugbotu[8], Tabitha Chukwudi Aghaunor[9], Amanda Enaodona Oweimeito[10]

[1]Department of Cybersecurity, Faculty of Computing, Delta State University of Science and Technology, Ozoro, Nigeria
[2]Department of Information Systems and Tech, Faculty of Computing, Delta State University of Science and Technology, Ozoro, Nigeria
[3]Department of Computer Science, Faculty of Computing, Delta State University of Science and Technology, Ozoro, Nigeria
[4]Department of Cybersecurity, Faculty of Computing, University of Delta, Agbor, Nigeria
[5]Department of Informatic Engineering, Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia
[6]Department of Computer Science, College of Computing, Federal University of Petroleum Resources Effurun, Warri, Nigeria
[7]Department of Chemical Engineering, College of Engineering, Federal University of Petroleum Resources Effurun, Warri, Nigeria
[8]Department of Data Science, School of Science, Engineering and Enviroment, University of Salford, Manchester, United Kingdom
[9]School of Data Intelligence and Technology, Robert Morris University, Pittsburg, Pennsylvania, United States
[10]Department of Mathematical Science, Faculty of Science, Edwin Clark University, Kiagbodo, Nigeria

Article Info

Article history:

Received Jun 23, 2024
Revised Dec 27, 2024
Accepted Mar 25, 2025

Keywords:

Feature selection
Machine learning
Polyurethan foam production
Scorch occurrence in foams
Synthetic minority over-sampling technique-Tomek

ABSTRACT

Scorch occurrence during the production of flexible polyurethane foam has been a menace that consistently, jeopardize a foam's integrity and resilience. It leads to foam suppression and compactness integrity failure due to scorch. There is always the increased likelihood of scorching, and makes crucial the utilization of methods that seek to avert it. Studies predict that the formation of foam constituent processes via optimization using machine learning have adequately trained models to effectively identify scorch occurrence during the profiling in the polyurethane foam production. Our study utilizes the random forest (RF) ensemble with feature selection (FS) and data balancing technique to identify production predictors. Study yields accuracy of 0.9998 with F1-score of 0.9819. Model yields 2-distinct cases for (non)-occurrence of scorch respectively, and the ensemble demonstrates that it can effectively and efficiently predict the occurrence of scorch in the production of flexible polyurethane foam manufacturing process.

Corresponding Author:

Margaret Dumebi Okpor
Department of Cybersecurity, Faculty of Computing, Delta State University of Science and Technology
Ozoro, Delta State, Nigeria
Email: okpormd@dsust.edu.ng

1. INTRODUCTION

With the manufacture of polyurethane foams–chlorofluorocarbons have been successfully utilized as blowing agents–even when it results in the hazardous emission of atmospheric constituent effects [1] that has also led to its consequent ban [2]. The quest for an alternative has caused the utilization of water as substitute for polyurethane foam production. The mix of isocyanate and water ripples across the chemical composition, an exothermic reaction causing a rise in the foam's temperature [3] such that during its cure phase–scorch often occurs. Scorch is a yellow-to-brown coloration in polyurethane foam, visually recognized on exposure of a polyurethane foam during its curing phase [4] that makes any adjustments, unwise. Thus, expert quest on how best to adjust the material composition prior formation of scorch, which physically suppresses the foam and significantly reduce its durability and compactness. A scorched foam quickly suppresses [5] and has no elastic recovery. Known causes of scorch includes the existence of non-polymeric components, and oxidation of phenols/amines [6]. Studies suggests that these constituents are responsible for discoloration of scorched foam. While seeking better alternatives–using the proper amount of these constituents can greatly prevent scorch formation. Scorch is a heat-induced change in the production of polyurethane foam. Also, the inadequate exposure to air prevents the dissipation of trapped heat in the formed polyurethane foam [7], so that the foam scorches prior its hardening. Some known impact of scorch effect include [8], [9]: i) low resilience in its compactness, ii) low elasticity cum recovery, iii) porous structure with low resistant to load-carrying capacity, iv) reduced life span with high-volume waste, and v) reduced profitability.

Known remedies to scorch includes: i) use of temperature retardants/suppressant, ii) addition of antioxidants/salts as anti-scorch components, iii) addition of free isocyanate to moderate the fast exothermic reaction that reduces rise in temperature, and iv) use of inhibitors (e.g., halogenated phosphate esters) in proper ratio with hydroquinone and diphenylamine. While, these have benefits, some concerns on emissions, and high operational cost of environmental clean-up often renders unsuitable such solutions. Also, the use of inhibitors exert morphological and chemical changes to the foam; and thought of as lessening discoloration [10] inhibitors contributes to discoloration not due to scorching. Thus, it is evident that a major available option to tackle scorch requires expert skillsets to properly navigate via careful proportions of materials using cost-efficient solution(s) that simulates the process, as it responds to scorch occurrence prior the physical manufacturing. In lieu of ground-truth, generalization may not adequately account for various specifications in the polyurethane materials such as: i) diisocyanates and polyols materials and ii) other environmental conditions within a production plants and scenarios that scorch is typically advanced as resulting from mechanical defects [11].

We model these dynamics and complex production processes as variables using machine learning (ML) schemes–so as to provision adequate insight with trial-and-error simulations for a variety of scenarios in the chemical manufacturing phases [12]. Learning seeks to aggregate the learned intrinsic patterns using a classifier. Successfully known implemented ML approaches include: logistic regression [13], K-nearest neighbors (KNN) [14], random forest (RF) [15], and deep learning [16]. MLs have their inherent drawbacks–use of feature selection (FS) and data balancing in their quest for prediction accuracy has remain a crucial feat. Thus, we utilize the RF ensemble with relief ranking and synthetic minority over-sampling technique (SMOTE)-Tomek links (synthetic minority oversampling technique) for data balancing on dataset from the Winco Foam Limited in Benin-City (Nigeria). The ensemble choice is attributed to its capacity to reduce overfit, address the imbalanced nature in datasets, and thus, yield enhanced performance model accuracy.

2. MATERIALS AND METHODS

2.1. Tree-based ensemble(s)

A common ML scheme is the tree-based approaches – where each decision tree (DT) yields a collect of if-then-else rules used in majority voting that allows it to predict observed classes. Each tree explores a recursive top-down mode, partitioned using binary-approach for its predictors with variables grouped into successful homogenous distribution of the dependent variable y. A DT alone is easily understood; But yield model overfit and degraded performance when classifying unknown labels. Tree-based models learn by constructing individually-trained DTs that aggregates their results into a stronger model. Trees learn via: i) bagging that iteratively generates a training-set that aggregates results to reduce variance and bias via voting, and ii) boosting reduces bias by sequentially aggregating the performance of many weak classifiers onto a strong learner with enhanced accuracy so that each successor learner's outcome accounts for the inherent weakness in the previous base learner [17]. Both modes, enhances prediction accuracy by mitigating variance and bias, to reduce errors with misclassified outcomes. With boosting mode, its accuracy is achieved by sequentially training each learner to via feedback–correct the weaknesses in its base weaker predecessor [18]. Popular boosting ensembles includes the stochastic, gradient and adaptive boosting ensembles. Its prediction as in (1) is achieved by combining the outcome of its weak learners to yield higher weights for incorrectly classified cases [19], [20].

$$L^t = \sum_{i=1}^{n} l\left(Y_i^t, \hat{Y}_i^{t-1} + f_k(x_i)\right) + \Omega(f_t) \tag{1}$$

While for bagging approach–each DT grows independently from the root/parent trees–constructed via bootstrap summation to traverse all sample data during prediction via majority vote mode [21], [22]. The RF adds an extra (randomness) layer to bagging approach, to change how each tree is constructed. With DTs, each node is split among all predictor(s) that are randomly chosen at the node. Its recursive structure captures interaction effects between variables. Thus, tree-based models have proven successful for a variety of tasks; with most hill-climbing models stuck at local minima, the RF weight combines various local minima to yield an ensemble method that minimizes risk in the choice of a (wrongly) chosen local minima [23].

2.2. Feature selection

A major challenge to learning classifiers include: i) finding the right dataset with appropriate feats, which estimates to ground-truth (i.e., target parameter) and ii) selection of appropriate features, which when extracted (is near enough to the target parameter) to lead to ground-truth and devoid the model of overfitting and poor generalization; and thus, helps to effectively predict its underlying distribution. Beside selection of properly formatted dataset, our heuristic choice must be able to select crucial estimation predictors to devoid the model over-parameterization–and leads to poor generalization (i.e., model overfit and overtrain). FS helps to curb this as a dimensionality reduction technique–to remove irrelevant predictors and helps the model to overcome challenges of dimensionality [24] so that its training will leads to enhanced performance. It is critical for tasks where cost and attribute's measure are important–to streamline the dataset, equipping and fastening a model's construction that assists with interpreting the intrinsic feats of the dataset. Every classifier that achieves good performance on training dataset, do not often blend well on set test-dataset, and it may result in model overfit. In many cases, dataset is split into *k*-fold(s) and engaged in both training and test phases. Thus, FS is executed as means to aid dimensionality reduction [25], [26].

FS modes is classified into: i) filter scheme that leverages inherent features of the dataset distribution to select those predictor variables (parameters) that are considered appropriate for classification task and ii) wrapper scheme that assess predictor qualities [27] making it computationally, less cost-effective when compared to the filter mode. This is because, chosen predictors are more inclined to learn in lieu of the adopted classifier [28], [29]. Thus, assessing its goodness of fit is assessed with its efficiency cum efficacy within the chosen model. Evaluating the FS where the target class (i.e., real relevant features) is known can be quite easy. However, for realtime data can be tedious, as target class (i.e., ground truth) is very much unavailable for train with real-world, realtime dataset.

2.3. Method and framework

For our proposed method, we adopt the tree-based RF ensemble as in the Figure 1 which portends the following steps:

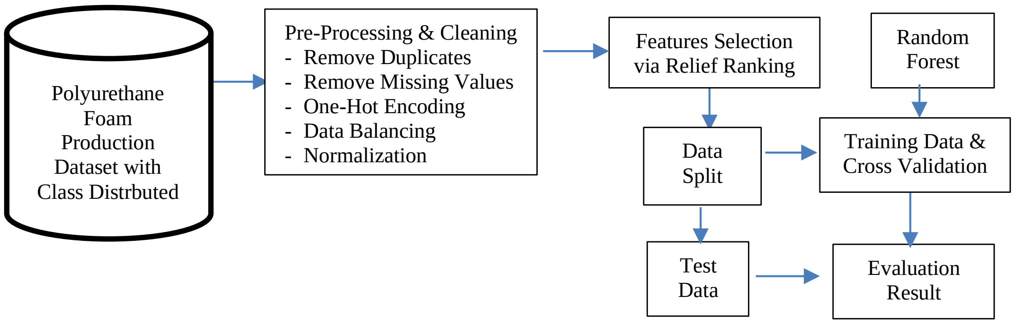

Figure 1. Proposed RF ensemble tree-based algorithm for scorch prediction

a. Step 1–data collection: dataset as retrieved from Winco Foams, Edo State in Nigeria–consist of 15-features with 8540-records as thus [30], [31]: polyurethane_thruput, TDI_thruput, calcium_thruput, calcium_dial, water_thruput, polyurethane_dials, TDI_dial, water_dial, calcium_qnty, TDI_qnty, polyurethane_qnty, water_qnty, polyurethane_water. time_production, and scorch as in Table 1. Dataset was retrieved via the Google Play Scrapper Library. Figure 2 shows the class distribution for the scorched (minority) and unscorched (majority) classes.

Table 1. The Winco Foam company dataset description

Items	Poly thru	Calc thru	TDI thru	Water thru	Poly dial	Calc dial	TDI dial	Water dial	Qnty poly	Qnty calc	Qnty TDI	Qnty water	Prod time	Scorch
Min	45.000	10.000	50.250	4.3500	6.6000	15.000	67.000	270.00	291.00	43.650	213.40	17.190	0.0008	3.8800
Mid	75.000	14.005	55.420	4.5600	14.050	18.800	68.000	280.00	1350.0	262.02	1042.5	89.645	0.0008	20.002
Max	80.000	16.000	55.610	4.7000	23.610	20.800	71.000	318.00	3000.0	923.85	2625.0	228.40	0.0008	50.000
Mean	71.996	13.142	54.643	4.4988	13.469	18.280	68.111	278.28	1485.5	293.86	1141.5	96.276	0.0800	20.908
Std	9.3113	1.9250	1.4917	0.0865	3.6109	2.1598	0.8544	10.053	729.86	175.89	573.91	46.816	0.0004	10.501

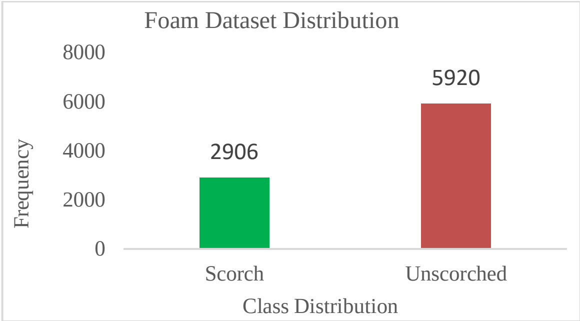

Figure 2. Class distribution for the Winco Foam dataset

b. Step 2–preprocess and clean: seeks to remove duplicate values to improve data quality, and remove missing values to avoid redundancy. We utilize one-hot encoding method that converts categorical values into a suitable form for the ML models; since, ML schemes cannot handle category data directly, it creates a binary equivalence of the dataset by converting categorical variable into their binary form.
c. Step 3–data balancing seeks to redistribute the data-points in the dataset to ensure an almost equitable distribution between major-and-minor classes. With a variety of modes available, we adopt the SMOTE-Tomek as thus: i) identifies majority-class, ii) interpolates to create synthetic data-points via Tomek-link under-sample mode for a majority class, iii) adjusts data-points to those of its immediate neighbors so that new data-points overlaps, and iv) it repopulates the original dataset with generated synthetic labels to yield a balanced dataset as in Figure 3(a) which shows dataset balancing using SMOTE; while Figure 3(b) shows the dataset balancing using SMOTE-Tomek links.

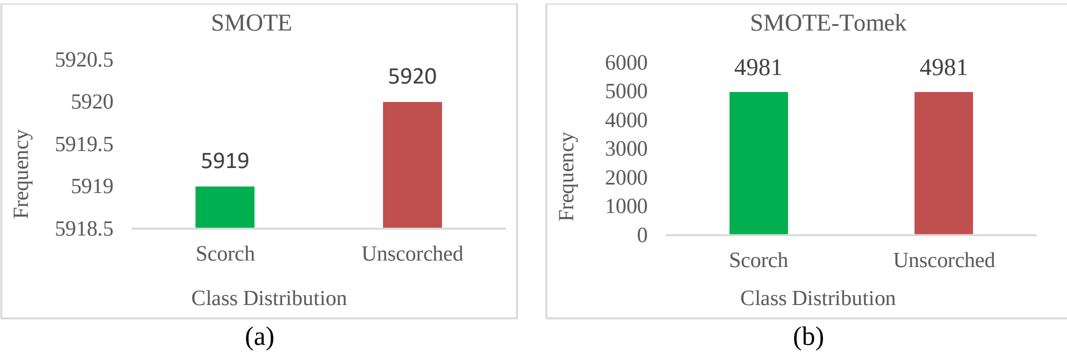

Figure 3. Data balancing approach; (a) SMOTE applied to dataset and (b) SMOTE-Tomek links applied

d. Step 4–normalization: explores transformative mode to the ‘often’ skewed dataset and ensure a nearly balanced class(es) distribution of the revised, repopulated dataset. The chosen feats are normalized using the standard normalization scaler as in (2), which reverts data features with a distribution mean value of 0 and a distribution standard deviation of 1. Also, x is the original value, μ is the mean, σ is the standard deviation, and z is our normalization process.

$$z = \frac{(x - \mu)}{\sigma} \quad (2)$$

e. Step 5–relief ranking wrapper FS mode: FS selects and extract what data is input x and determine labels to be estimated as ensemble output y. It removes all irrelevant/docile feats in estimating for ground-truth; and thus, reduce the predictor dimensions in the task dataset [32] to fasten construction of the ensemble for better performance [33] especially with cost as a major factor. How fit the ensemble is determined by the efficiency of the chosen feature(s) in estimating ground truth (i.e., target class). We use the relief ranking wrapper [34] mode to unveil how relevant, and ascertain how its occurrence fits with the target class. Original dataset consists of 13-features, we categorized the correlation of parameters to target (scorch) class [35], [36]. With the computed threshold is 11.323–a total of 7 predictors were selected as:

i) polyurethane_thruput, ii) calcium_throughput, iii) time_production, iv) quantity_water, v) water_ throughput, vi) quantity_polyurethane, and vii) scorch as in Table 2. These were examined in lieu of their correlated contribution to the classification task [37].

Table 2. Ranking of attributes score using the Chi-Square

Features	Selected (Yes/No)	*X2*-Value
polyurethane_thruput	Yes	13.364
calcium_thruput	Yes	15.419
TDI_thruput	No	0.9562
water_thruput	Yes	20.012
polyurethane_dial	No	0.2489
calcium_dial	No	24.701
tdi_dial	No	84.920
water_dial	No	83.721
quantity_polyurethane	Yes	88.222
quantity_calcium	No	0.2589
quantity_TDI	No	30.298
water_quantity	Yes	18.006
time_production	Yes	23.092
scorch	Yes	160.929

f. Step 6–data split and model initialization: to track each feature in lieu of our target class, we split dataset into 75% (train), and 25% (test) subsets. The RF model is a tree-based, widely-used, bagging ensemble–that yields accuracy via combining at its output using majority vote for its weak trees to yield a powerful learner [38], [39]. Constructed from many DTs–its voting ensures that all base classifiers have same weight. With its bootstrap (random) sampling, its trees yield a higher weight if selected attributes ensures that all DTs yield the same ability to make decisions. Thus, RF is devoid of overfitting, poor generalization and handles complex contiguous datasets effectively – leaning on decisions of its many weak, base trees to yield a stronger classifier as thus: i) it splits the original dataset using row-and-feature sampling so that dataset is made up of select columns/rows with replacement, ii) it creates a DT for each subset selected/assigned, and iii) each DT yields an output via majority voting and/or averaging [40]. Table 3 shows the construction parameters for the adopted RF ensemble.

Table 3. The RF ensemble design and configuration

Features	Values	Description
learning_rate	0.25	Step size learning for update
n_estimators	150	Number of trees constructed
max_features	5	Maximum number of features to construct of the RF tree ensemble
max_depth	5	Max depth of each tree
eval_metric	error, logloss	Performance evaluation metrics
min_weight_fraction_leaf	0.1	Tree's structure based on weight assigned to each sample
warm_start	FALSE	Ensure tree does not restart
random_state	25	The seeds for reproduction
eval_set	x,val, y_val	Train data for evaluation
min_sample_split	10	Minimal samples needed
verbose	TRUE	Determines if ensemble evaluation metric is printed at training
min_sample_leaf	auto	Number of feats to be considered
bootstrap	TRUE	Ensures bootstrap aggregation use

g. Step 3–training as applied–seeks to estimate learned skills on unseen data. It evaluates the model's accuracy on performance on how well the ensemble has learned intricate underlying (interest) features as they impact on change via resampling. The (stratified) k-fold is used to rearrange labels and ensure that each fold yields a good representation of dataset. Our ensemble learns from scratch, and iteratively constructs the DTs for the RF-model. Each tree is trained via bootstrap resampling on the enhanced train dataset, which enhances the DT's collective knowledge to identify intricate trends inherent the dataset. Train dataset blends actual examples that guarantees the tree's comprehensive learning experience; and thus, improves the adaptive flexibility of the ensemble to a variety of settings within a dataset [41].

3. FINDINGS AND DISCUSSION

3.1. Training/hyper-predictors tuning

Table 3 yields the before/after performance evaluation with the application of FS and SMOTE-Tomek, and agrees with [42], [43] on the outlier effects of data not present from the outset. Studies agrees that RF outperformed other benchmark models with accuracy, recall, precision, and F1.

Table 3. Performance ‘before/after’ SMOTE-Tomek and wrapper mode FS

Ensembles	Before				After			
	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall
KNN	94.35	77.47	92.64	66.57	92.1	92.28	90.18	94.48
LR	92.19	97.18	93.57	95.82	98.05	98.05	98.05	98.05
Support vector machine (SVM)	90.08	50	94.57	33.98	81.45	80.32	85.41	75.81
Naïve Bayes (NB)	95.08	83.03	83.62	82.45	91.25	90.74	96.16	85.9
Our proposed RF method	98.02	98.02	96.89	99.01	99.19	98.19	98.28	98.1

Our ensemble outperforms the benchmarks such that prior the application of FS and data balancing, our RF model yields a 98.02% accuracy with 92.19% for LR, 94.35% for KNN, 95.08% for NB, and 90.08% for SVM respectively. Furthermore, RF yields F1 0.9919 having applied FS with SMOTE-Tomek mode; while, LR, KNN, NB, and SVM yielded an accuracy of 98.05%, 92.10%, 91.25%, and 81.45% respectively. Thus, the utilization of both FS [44], [45] and SMOTE-Tomek assured the classifier of improved accuracy. Our proposed model yields an F1-score and accuracy of 99.19% and 98.19% respectively.

3.2. Discussion of findings

Our proposed ensemble effectively identifies scorch data accurately in the chosen dataset, and is proven to efficiently reduce skewness, variance and bias indicative in the adapted method. Figure 4 shows the confusion matrix, which agrees with this–as the ensemble yields a more robust scheme for the hidden cum underlying (interest) predictors within a task’s train-dataset being considered.

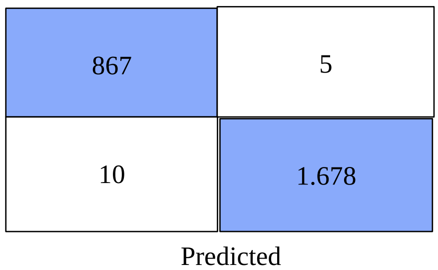

Figure 4. Confusion matrix for the proposed ensemble

Our study provides evidence of SMOTE-Tomek outperforming SMOTE with greater influence in lieu of target class (and/or ground-truth). It enhanced the ensemble’s generalization by identifying intrinsic patterns that influences for improved performance and enhanced efficiency for differentiating between false/true negatives-and-positives. We successfully utilized various schemes with k-fold (stratified) re-train (cross-validation) resultant performance on the occurrence/presence of scorch in production of polyurethane foam. Our ensemble outperforms others for the chosen dataset with an accuracy of 98.19% – and yielding 2-distinct predictions for the occurrence of scorch as in Table 4. Our proposed model successfully predicts the occurrence of scorch with all predictor mix. This, will significantly cut down the scorch occurrence bin that unveils a variety of intertwined relationships between its features and water.

Table 4. Predicted values–‘1’ indicates the occurrence/presence of scorch

Poly thru	TDI thru	Poly dial	Calc thru	Qnty Calc	Calc dial	Water thru	Qnty poly	TDI dial	Water dial	Qnty TDI	Prod time	Qnty water	Scorch
75	55.50	11.3	11.25	225	75	4.564	1500	68	280	91.28	20	1110	0(No)
75	55.42	11.3	11.25	71.77	75	4.564	478.5	68	280	101.1	6.38	353.58	1(Yes)

With retrieved data restricted to only daily generation–this resulted in limited availability cum access as infrastructure generated large volume dataset. Its documentation if not quick utilized, is lost. The actions of data balancing, normalization and others sought to minimize outlier effects [46]. Addition of the universal (water) solvent impacted a correlation and suggests/unveils a variety of relations with other components that is also agreed with in [47].

4. CONCLUSION

Advances in technological development and the widespread adoption of technology-driven business strategies, businesses can now operate more efficiently, productively, and profitably. Despite the enormous amount of data generated daily, the polyurethane foam production industry still lags behind in developing data analytics tools. This study is a positive step and should be improved upon. With dataset used, not a number (NaN) values implies no relationship; as such, preprocessing scheme should be adopted to prevent model overfitting, and column flattening prior the deployment of ML scheme, which agrees with. These correlation numbers are illogical at first glance. But ML can successfully glean insightful knowledge therein. And these algorithms can trace the entangled connections and interpret how each variable interacts with the others and influences the column (variable) that we want to predict.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Margaret Dumebi Okpor		✓			✓				✓					✓
Kizito Eluemnor Anazia		✓	✓			✓				✓				✓
Wilfred Adigwe	✓		✓		✓	✓		✓		✓	✓		✓	✓
Ejaita Abugor Okpako		✓	✓			✓		✓	✓	✓	✓	✓		✓
De Rosal Ignatius Moses Setiadi	✓	✓		✓	✓		✓			✓	✓	✓	✓	
Arnold Adimabua Ojugo	✓	✓		✓	✓				✓	✓		✓	✓	
Felix Omoruwou	✓	✓				✓	✓		✓			✓	✓	
Rita Erhovwo Ako		✓		✓	✓				✓		✓	✓		✓
Victor Ochuko Geteloma		✓				✓				✓	✓			✓
Eferhire Valentine Ugbotu	✓		✓	✓		✓				✓	✓	✓	✓	✓
Tabitha Chukwudi Aghaunor		✓	✓		✓	✓			✓		✓		✓	✓
Amanda Enaodona Oweimeito	✓				✓	✓	✓	✓		✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from [Winco Foam Liited, Benin City, Edo State]. Restrictions apply to the availability of these data, which were used under license for this study.

Data are available [Engr. Dr. Felix Omoruwou: +2348039492008] with the permission of [Winco Foams Limited, Benin City – Edo State, Nigeria]. Some aspects of the data supporting the findings of this study are available within the article. No new data were created or analyzed in this study.

REFERENCES

[1] F. O. Aghware *et al.*, "Effects of Data Balancing in Diabetes Mellitus Detection : A Comparative XGBoost and Random Forest Learning Approach," *NIPES - J. Sci. Technol. Res.*, vol. 7, no. 1, pp. 1–11, 2025, doi: 10.37933/nipes/7.1.2025.1.

[2] E. Yakub, S. E. Agarry, F. Omoruwou, and C. N. Owabor, "Comparative study of the batch adsorption kinetics and mass transfer in phenol-sand and phenol-clay adsorption systems," *Part. Sci. Technol.*, vol. 38, no. 7, pp. 801–811, Oct. 2020, doi: 10.1080/02726351.2019.1616862.

[3] A. Taravat and F. D. Frate, "Development of band ratioing algorithms and neural networks to detection of oil spills using Landsat ETM+ data," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, 2012, doi: 10.1186/1687-6180-2012-107.

[4] N. C. Ashioba *et al.*, "Empirical Evidence for Rainfall Runoff in Southern Nigeria Using a Hybrid Ensemble Machine Learning Approach," *J. Adv. Math. Comput. Sci.*, vol. 12, no. 1, pp. 73–86, 2024, doi: 10.22624/AIMS/MATHS/V12N1P6.

[5] S. Levchik, M. P. Luda, P. Bracco, P. Nada, and L. Costa, "Discoloration in fire-retardant flexible polyurethane foams," *J. Cell. Plast.*, vol. 41, no. 3, pp. 235–248, 2005, doi: 10.1177/0021955X05053523.

[6] N. V. Gama, A. Ferreira, and A. Barros-Timmons, "Polyurethane foams: Past, present, and future," *Materials (Basel).*, vol. 11, no. 10, 2018, doi: 10.3390/ma11101841.

[7] I. Ehsan *et al.*, "Internet of Things-Based Fire Alarm Navigation System: A Fire-Rescue Department Perspective," *Mob. Inf. Syst.*, vol. 2022, pp. 1–15, Sep. 2022, doi: 10.1155/2022/3830372.

[8] R. L. Gray and R. E. Lee, "Scorch inhibitors for flexible polyurethanes," *Plastics Additives: An AZ reference,* pp. 567–575, 1998, doi: 10.1007/978-94-011-5862-6_63.

[9] M. A. Rahman, G. A. Francia, and H. Shahriar, "Leveraging GANs for Synthetic Data Generation to Improve Intrusion Detection Systems," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 429–439, Feb. 2025, doi: 10.62411/faith.3048-3719-52.

[10] J. Demassa, "Polyol Stabilization and the Introduction of a New PUR Slabstock Foam Antioxidant," *Cell. Plast.*, vol. 41, no. December, pp. 15–24, 2013.

[11] U. A. Amran *et al.*, "Production of Rigid Polyurethane Foams Using Polyol from Liquefied Oil Palm Biomass: Variation of Isocyanate Indexes," *Polymers (Basel).*, vol. 13, no. 18, p. 3072, Sep. 2021, doi: 10.3390/polym13183072.

[12] Y. Nabata, A. Mamada, and H. Yamasaki, "Study of the curing process giving the rigid polyurethane foam by dynamic viscoelastic method," *J. Appl. Polym. Sci.*, vol. 35, no. 1, pp. 155–166, Jan. 1988, doi: 10.1002/app.1988.070350114.

[13] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using GA algorithm for feature selection," *J. Big Data*, vol. 9, no. 1, p. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.

[14] E. A. L. M. Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich, and P. Sankaran, "A systematic review of literature on credit card cyber fraud detection using machine and deep learning," *PeerJ Comput. Sci.*, vol. 9, p. e1278, Apr. 2023, doi: 10.7717/peerj-cs.1278.

[15] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," in *2018 IEEE 15th Int. Conf. Netw., Sen. Cont. (ICNSC)*, IEEE, Mar. 2018, pp. 1–6, doi: 10.1109/ICNSC.2018.8361343.

[16] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *J. Big Data*, vol. 8, no. 1, p. 151, Dec. 2021, doi: 10.1186/s40537-021-00541-8.

[17] A. E. Oweimieotu, M. I. Akazue, A. E. Edje, and C. Asuai, "Designing a Hybrid Genetic Algorithm Trained Feedforward Neural Network for Mental Health Disorder Detection," *Digit. Innov. Contemp. Res. Sci. Eng. Technol.*, vol. 12, no. 1, pp. 49–62, 2024, doi: 10.22624/AIMS/DIGITAL/V11N4P4.

[18] G. Cho, J. Yim, Y. Choi, J. Ko, and S. H. Lee, "Review of machine learning algorithms for diagnosing mental illness," *Psychiatry Investig.*, vol. 16, no. 4, pp. 262–269, 2019, doi: 10.30773/pi.2018.12.21.2.

[19] J. P. Ntayagabiri, Y. Bentaleb, J. Ndikumagenge, and H. El Makhtoum, "OMIC: A Bagging-Based Ensemble Learning Framework for Large-Scale IoT Intrusion Detection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 401–416, Feb. 2025, doi: 10.62411/faith.3048-3719-63.

[20] J. P. Ntayagabiri, Y. Bentaleb, J. Ndikumagenge, and H. El Makhtoum, "A Comparative Analysis of Supervised Machine Learning Algorithms for IoT Attack Detection and Classification," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 395–409, Feb. 2025, doi: 10.62411/jcta.11901.

[21] A. Satpathi *et al.*, "Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India," *Sustain.*, vol. 15, no. 3, p. 2786, Feb. 2023, doi: 10.3390/su15032786.

[22] A. Angdresey, L. Sitanayah, and I. L. H. Tangka, "Sentiment Analysis for Political Debates on YouTube Comments using BERT Labeling, Random Oversampling, and Multinomial Naïve Bayes," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 342–354, Jan. 2025, doi: 10.62411/jcta.11668.

[23] B. N. Supriya and C. B. Akki, "Sentiment prediction using enhanced xgboost and tailored random forest," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 191–199, 2021, doi: 10.12785/ijcds/100119.

[24] M. I. Akazue, C. E. Asuai, A. E. Edje, E. U. Omede, and E. Ufiofio, "Cybershield : Harnessing Ensemble Feature Selection Technique for Robust Distributed Denial of Service Attacks Detection," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 03, pp. 1211–1224, 2023.

[25] S. Adamu, A. Iorliam, and Ö. Asilkan, "Exploring Explainability in Multi-Category Electronic Markets: A Comparison of Machine Learning and Deep Learning Approaches," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 440–454, Mar. 2025, doi: 10.62411/faith.3048-3719-58.

[26] L. Bonde and A. K. Bichanga, "Improving Credit Card Fraud Detection with Ensemble Deep Learning-Based Models: A Hybrid Approach Using SMOTE-ENN," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 383–394, Feb. 2025, doi: 10.62411/jcta.12021.

[27] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, doi: 10.1088/1742-6596/1168/2/022022.

[28] A. Maureen, O. Anthonia, E. Omede, J. P. A. Hampo, J. Anenechukwu, and C. Hampo, "Use of Adaptive Boosting Algorithm to Estimate User's Trust in the Utilization of Virtual Assistant Systems," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 1, pp. 502–509, 2023.

[29] S. E. Brizimor *et al.*, "WiSeCart: Sensor-based Smart-Cart with Self-Payment Mode to Improve Shopping Experience and

Inventory Management," *Adv. Multidiscip. Sci. Res. J. Publ.*, vol. 10, no. 1, pp. 53–74, Mar. 2024, doi: 10.22624/AIMS/SIJ/V10N1P7.
[30] A. Çetin and S. Öztürk, "Comprehensive Exploration of Ensemble Machine Learning Techniques for IoT Cybersecurity Across Multi-Class and Binary Classification Tasks," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 371–384, Feb. 2025, doi: 10.62411/faith.3048-3719-51.
[31] K. K. San, H. H. Win, and K. E. E. Chaw, "Enhancing Hybrid Course Recommendation with Weighted Voting Ensemble Learning," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 337–347, Jan. 2025, doi: 10.62411/faith.3048-3719-55.
[32] H. Lu and C. Rakovski, "The Effect of Text Data Augmentation Methods and Strategies in Classification Tasks of Unstructured Medical Notes," *Res. Sq.*, vol. 1, no. 1, pp. 1–29, 2022, doi: 10.21203/rs.3.rs-2039417/v1.
[33] M. Bayer, M. A. Kaufhold, B. Buchhold, M. Keller, J. Dallmeyer, and C. Reuter, "Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 1, pp. 135–150, 2023, doi: 10.1007/s13042-022-01553-3.
[34] A. Bahl *et al.*, "Recursive feature elimination in random forest classification supports nanomaterial grouping," *NanoImpact*, vol. 15, p. 100179, Mar. 2019, doi: 10.1016/j.impact.2019.100179.
[35] F. Mazhar, N. Aslam, A. Naeem, H. Ahmad, M. Fuzail, and M. Imran, "Enhanced Diagnosis of Skin Cancer from Dermoscopic Images Using Alignment Optimized Convolutional Neural Networks and Grey Wolf Optimization," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 368–382, Jan. 2025, doi: 10.62411/jcta.11954.
[36] A. O. Kolawole, M. E. Irhebhude, and P. O. Odion, "Human Action Recognition in Military Obstacle Crossing Using HOG and Region-Based Descriptors," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 410–426, Feb. 2025, doi: 10.62411/jcta.12195.
[37] A. M. Ifioko *et al.*, "CoDuBoTeSS: A Pilot Study to Eradicate Counterfeit Drugs via a Blockchain Tracer Support System on the Nigerian Frontier," *J. Behav. Informatics, Digit. Humanit. Dev. Res.*, vol. 10, no. 2, pp. 53–74, 2024, doi: 10.22624/AIMS/BHI/V10N2P6.
[38] M. Gratian, S. Bandi, M. Cukier, J. Dykstra, and A. Ginther, "Correlating human traits and cyber security behavior intentions," *Comput. Secur.*, vol. 73, pp. 345–358, Mar. 2018, doi: 10.1016/j.cose.2017.11.015.
[39] O. Okolo, B. Y. Baha, and M. D. Philemon, "Using Causal Graph Model variable selection for BERT models Prediction of Patient Survival in a Clinical Text Discharge Dataset," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 455–473, Mar. 2025, doi: 10.62411/faith.3048-3719-61.
[40] N. B. Yahia, M. D.Kandara, and N. B. BenSaoud, "Integrating Models and Fusing Data in a Deep Ensemble Learning Method for Predicting Epidemic Diseases Outbreak," *Big Data Res.*, vol. 27, no. November, 2022, doi: 10.1016/j.bdr.2021.100286.
[41] Rukshan Pramoditha, "k-fold cross-validation explained in plain English," *Towar. Data Sci.*, 2020.
[42] G. TekalignTujo, G. Dileep Kumar, D. ElifeneshYitagesu, and B. MeseretGirma, "Predictive Model to Predict Seed Classes using Machine Learning," *Int. J. Eng. Res. & Technol.*, vol. 6, no. 08, pp. 334–344, 2017.
[43] Q. Li *et al.*, "An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–15, 2017, doi: 10.1155/2017/9512741.
[44] M. I. Akazue, I. A. Debekeme, A. E. Edje, C. Asuai, and U. J. Osame, "UNMASKING FRAUDSTERS : Ensemble Features Selection to Enhance Random Forest Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 201–212, 2023, doi: 10.33633/jcta.v1i2.9462.
[45] P. O. Adebayo, F. Basaky, and E. Osaghae, "Leveraging Variational Quantum-Classical Algorithms for Enhanced Lung Cancer Prediction," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 307–323, Dec. 2024, doi: 10.62411/jcta.10424.
[46] R. E. Ako *et al.*, "Pilot Study on Fibromyalgia Disorder Detection via XGBoosted Stacked-Learning with SMOTE-Tomek Data Balancing Approach," *NIPES - J. Sci. Technol. Res.*, vol. 7, no. 1, pp. 12–22, 2025, doi: 10.37933/nipes/7.1.2025.2.
[47] F. U. Emordi *et al.*, "TiSPHiMME: Time Series Profile Hidden Markov Ensemble in Resolving Item Location on Shelf Placement in Basket Analysis," *Digit. Innov. Contemp. Res. Sci.*, vol. 12, no. 1, pp. 33–48, 2024, doi: 10.22624/AIMS/DIGITAL/v11N4P3.

BIOGRAPHIES OF AUTHORS

Margaret Dumebi Okpor received her B.Sc. and M.Sc. (Computer Science) in 1997 and 2014 respectively from the University of Benin; and Ph.D. in 2023 from the Ignatius Ajuru University of Education in Port-Harcourt, Rivers State. She lectures with the Department of Cybersecurity at the Delta State University of Science and Technology Ozoro. Her research interests include AI-driven identity management and access control, data science, and machine learning. She is also a member of the Nigerian Computer Society, and the Computer Professionals of Nigeria (CPN). She can be contacted at email: okpormd@dsust.edu.ng.

Kizito Eluemnor Anazia received his B.Sc. in 2001 from Ambrose Alli University, Ekpoma in Edo State; his M.Sc. in 2011 from the University of Port Harcourt, Rivers State, and Ph.D. in 2021 from Nnamdi Azikiwe University, Anambra State. He currently lectures with the Department of Information Systems and Technology at the Delta State University of Science and Technology, Ozoro. He is a member of the Nigerian Computer Society (NCS), Computer Professionals of Nigeria (CPN), Cyber Security Experts Association of Nigeria (CSEAN) and Information Technology Systems & Security Professionals (IT&SP). His research interest includes machine learning, data science, e-commerce, database management, and analysis. He can be contacted at email: anaziake@dsust.edu.ng.

Wilfred Adigwe received B.Eng. in Computer Science in 2000 from the Enugu State University of Science and Technology; M.Sc. in 2010 from the Nnamdi Azikiwe University Awka; and Ph.D. in Computer Science in 2018 from the Nnamdi Azikiwe University Awka. He lectures with the Department of Computer Science, Delta State University of Science and Technology Ozoro. His research interests include data science, cybersecurity, and machine learning. He is a member of Computer Professionals of Nigeria, Nigeria Computer Society and Cyber Security Experts of Nigeria (CSEAN). He can be contacted at email: adigwew@dsust.edu.ng.

Ejaita Abugor Okpako received B.Sc., M.Sc., and Ph.D. (all in Computer Science) from the University of Port Harcourt, Nigeria. He is the Acting Dean, Faculty of Computing at the University of Delta, Agbor, Nigeria. His areas of interest include artificial intelligence, data science, cyber security, big data, and software engineering. He served previously as the Director of ICT, Edwin Clark University, Kiagbodo, Delta State, Nigeria. He has published various articles in both referred Journals and Conference proceedings. He can be contacted at email: ejaita.okpako@unidel.edu.ng.

De Rosal Ignatius Moses Setiadi received his Bachelor's degree in Informatics Engineering from Soegijapranata Catholic University in 2010, followed by his Master's degree and Doctoral degree in Informatics Engineering from Dian Nuswantoro University in 2012 and 2024, respectively. He is currently a Lecturer and Researcher at the Faculty of Computer Science, Dian Nuswantoro University. His research interests include image encryption, cryptography, steganography, watermarking, and machine learning. He serves as an academic editor for Security and Communication and the Journal of Computer Networks and Communications, a member of the Editorial Board for Computers, Materials & Continua and the Technology, Education, Management, Informatics Journal, and Editor-in-Chief of the Journal of Future Artificial Intelligence and Technologies. He has been recognized as a Top 2% Scientist in the World (2022–2024) by Elsevier B.V. His interest includes image cryptography, steganography, and machine learning. He can be contacted at email: moses@dsn.dinus.ac.id.

Arnold Adimabua Ojugo received his B.Sc., M.Sc., and Ph.D. from Imo State University Owerri, Nnamdi Azikiwe University Awka, and Ebonyi State University Abakiliki in 2000, 2005 and 2013 respectively. He is a Professor with the Department of Computer Science, Federal University of Petroleum Resources Effurun – with research interest(s) in: intelligent systems computing, data science, and cybersecurity. He is a researcher, and editorial board member of: frontiers in big data, The International Journal of Modern Education in Computer Science IJMECS, and Progress for Intelligent Computation and Application. He is a member of the Nigerian Computer Society, Council of Computer Professionals of Nigeria, and International Association of Engineers. He can be contacted at email: ojugo.arnold@fupre.edu.ng.

Felix Omoruwou received his B.Eng. and Ph.D. in 1998 and 2023 respectively, from the University of Benin in Edo State; And his M.Eng. in 2015 from the University of Port Harcourt, Rivers State. He is a Senior Lecturer with the Department of Chemical Engineering, Federal University of Petroleum Resources Effurun with research interest(s) in environmental and biochemical engineering, and process modelling and simulation. As a researcher with over 10-years experience in both teaching and consulting – he is a corporate member of the Nigerian Society of Chemical Engineers (NSChE) and a registered engineer with the Council for the Regulation of Engineering in Nigeria (COREN). He can be contacted at email: omoruwou.felix@fupre.edu.ng.

Rita Erhovwo Ako received her B.Sc. Industrial Mathematics in 2000 from the Delta State University Abraka in Delta State, Nigeria; M.Sc. Computer Science in 2005 from the University of Ibadan in Oyo State; M.Sc. Internet-Computer and System Security in 2006, and Ph.D. Computer Science in 2013 respectively from the University of Bradford, Bradford, United Kingdom. She is currently a Senior Lecturer with the Department of Computer Science at The Federal University of Petroleum Resources Effurun. She has several publications to her credit with research interests in: artificial intelligence, cybersecurity, e-commerce, embedded systems, and risk management. She is a member of the Nigerian Computer Society. She can be contacted at email: ako.rita@fupre.edu.ng.

Victor Ochuko Geteloma received his B.Sc. in Computer Science from the Federal University of Petroleum Resources Effurun, Delta State, Nigeria in 2015; M.Sc. in Computer Science in 2019 from the Covenant University, Ogun State. He currently Lectures with the Department of Computer Science at the Federal University of Petroleum Resources Effurun. He has several publications to his credit. His research interests include: cyber security, cloud computing, e-government, technology adoption, and digital inclusion. He is also a member of the prestigious Nigerian Computer Society (NCS). He can be contacted at email: geteloma.victor@fupre.edu.ng.

Eferhire Valentine Ugbotu received his B.Sc. in 2017 and M.Sc. in 2022 (in Computer Science) from the Federal University of Petroleum Resources Effurun, Delta State in Nigeria; He also received a second M.Sc. (Data Science) in 2024 from the Department of Data Science at the University of Salford in United Kingdom. He currently is a Research Assistant with the University of Salford in United Kingdom. His research interests are in: data science with machine learning approaches, IoTs, and cybersecurity. He can be contacted at email: eferhire.ugbotu@gmail.com.

Tabitha Chukwudi Aghaunor received her B.Sc. and M.Sc. in Computer Science from the University of Benin in Edo State in 2014 and 2017 respectively. She is currently a PostGraduate student at the School of Data Intelligence and Technology of the Robert Moriss University Pittsburg, Pennsylvania in United States. She also lectures with the Department of Computer Science at the University of Maritime Okerenkoko. She is a member of the Nigeria Computer Society (NCS) and the Council for the Registration of Computer Professionals of Nigeria (CPN). She can be contacted at email: tabitha.aghaunor@gmail.com.

Amanda Enaodona Oweimeito received her M.Sc. (Computer Science) in 2013 from Liverpool Hope University in Liverpool, United Kingdom; Her M.Phil. from the Delta State University Abraka – and she is currently undergoing her Doctoral Studies in Computer Science at the Federal University of Petroleum Resources Effurun, Delta State in Nigeria. She also lectures with the Department of Mathematical Science at the Edwin Clark University, Kiagbodo, Nigeria. Her research interest sinclude data science with machine learning approaches, IoT, robotics, and bioinformatics. She can be contacted at email: amandaoweimieotu@gmail.com.