Phishing Website Detection via a Transfer Learning based XGBoost Meta-learner with SMOTE-Tomek

Joy Agboi [1,*], Frances Uche Emordi [2], Christopher Chukufunaya Odiakaose [3], Rebecca Okeoghene Idama [4], Evans Fubara Jumbo [5], Amanda Enaodona Oweimieotu [6], Peace Oguguo Ezzeh [7], Andrew Okonji Eboka [8], Anne Odoh [9], Eferhire Valentine Ugbotu [10], Paul Avwerosuo Onoma [11], Arnold Adimabua Ojugo [12], Tabitha Chukwudi Aghaunor [13], Amaka Patience Binitie [14], Christopher Chukwudi Onochie [15], Blessing Uche Nwozor [16], Patrick Ogholuwarami Ejeh [17]
[1] Faculty of Science, Delta State University, Abraka, Delta State, Nigeria
[2, 3, 17] Faculty of Computing, Dennis Osadebay University, Asaba, Delta State, Nigeria
[4] Faculty of Computing, Southern Delta University, Ozoro, Delta State, Nigeria
[5, 6] School of Sciences, Edwin Clark University, Kiagbodo, Delta State, Nigeria
[7, 8, 14, 15] School of Science, Federal College of Education (Technical), Asaba, Nigeria
[9] School of Media and Communications, Pan-Atlantic University, Lekki, Lagos State, Nigeria
[10] Department of Data Science, University of Salford, Manchester, United Kingdom
[11, 12, 16] College of Computing, Federal University of Petroleum Resources, Effurun, Nigeria
[13] Department of Data Intelligence and Tech, Robert Morris University, Pittsburgh, Pennsylvania, USA
Email: [1] agboijoy0@gmail.com, [2] emordi.frances@dou.edu.ng, [3] osegalaxy@gmail.com, [4] idamaro@dsust.edu.ng, [5] evans3447@gmail.com, [6] oweimieotuamanda@edwinclarkuniversity.edu.ng, [7] peace.ezzeh@fcetasaba.edu.ng, [8] ebokaandrew@gmail.com, [9] aodoh@pau.edu.ng, [10] eferhire.ugbotu@gmail.com, [11] kenbridge14@gmail.com, [12] ojugo.arnold@fupre.edu.ng, [13] tabitha.aghaunor@gmail.com, [14] amaka.binitie@fcetasaba.edu.ng, [15] xtoline2@gmail.com, [16] nwozor.blessing@fupre.edu.ng, [17] patrick.ejeh@dou.edu.ng
*Corresponding Author

***Abstract*—The widespread proliferation of smartphones has advanced portability, data access ease, mobility, and other merits; it has also birthed adversarial targeting of network resources that seek to compromise unsuspecting user devices. Increased susceptibility was traced to user's personality, which renders them repeatedly vulnerable to exploits. Our study posits a stacked learning model to classify malicious lures used by adversaries on phishing websites. Our hybrid fuses 3-base learners (i.e. Genetic Algorithm, Random Forest, Modular Net) with its output sent as input to the XGBoost. The imbalanced dataset was resolved via SMOTE-Tomek with predictors selected using a relief rank feature selection. Our hybrid yields F1 0.995, Accuracy 1.000, Recall 0.998, Precision 1.000, MCC 1.000, and Specificity 1.000 – to accurately classify all 3,316 cases of its held-out test dataset. Results affirm that it outperformed benchmark ensembles. The study shows that our proposed model, as explored on the UCI Phishing Website dataset, effectively classified phishing (cues and lures) contents on websites.**

Keywords*—*Phishing Website; SMOTE-Tomek; Data Balancing; Memetic Algorithm; Tree-based Ensembles

I. INTRODUCTION

Digital revolution has ushered in a plethora of tools and processes that seek to advance efficient knowledge exchange among users [1]. The devices ease data processing tasks [2] while offering the benefits of flexibility in the shared resource cum enhanced user-connectivity [3]. With security a major issue, such advances have continued to ignite the interest of adversaries [4]. The proliferation of smartphones with their processing capacities has further eased it as invasive targets, with protocols made more possible with emergent tools [5], [6]. An adversary uses the penetrative tools like malware (spam) [7] to bolster socially-engineered threats that explore subterfuge mode to coordinate their attack at unsuspecting devices in their bid to compromise network infrastructure and resources [8][9]. These attack ensures that data exchange is targeted at exploits on a user's social needs, desires [10] and insatiable traits [11]. Today's businesses are reshaped via fusion of informatics [12] – as a channel to deliver high-end values to consumers, who receive services as rendered [13]. This exchange has today become a trillion-dollar war [14], as businesses must seek new frontiers to curb attacks amongst other issues [15][16], as failure to safeguard these exchanges ushers in the need for cross-cutting research [17][18].

The success of many of these adversarial attacks is hinged on user personality traits, which include online presence, emotional seclusion, insatiable wants, and trust issues [19]. An adversary masks their intent as a trusted ally, to exploit a compromised resource – providing the attacker with a pivot access for further exploits on the infrastructure [20]. The consequent rise in the adoption of smartphones has further eased these attacks and compromises considerably. Phishing simply redirects a user's request to a spoofed website, rippled with malicious content that seeks to expose a targeted user [21] or device without their knowledge [22]. Phishing consist of 3-elements: (a) a lure masks an attacker as a genuine-user, targeting a user's empathy, fear and curiosity [23], (b) a hook is an embedded link in a message [24], and (c) a catch obtains an exposed device's private data [25]. Its success is hinged on its frequency and diversity [26] with unrealistic demands that seek to intimidate a user's psyche with petty gifts [27], [28]. Vulnerability to scam can be due to demographics (i.e. age, gender, and status) shown as in Fig 1 to Fig. 3 [29][30]. Girls between 24-to-42 years were the most phished due to

media presence or social seclusion [31]; There was also the factor of educational status cum societal approval [32]; and users between 18-to-29years were also phished more due to behavioural traits [33][34].

Victimization impacts website's contents and its structure with greater probability an unsuspecting user will fall prey [35]. To identify malicious contents, we must eliminate gaps by [36]: (a) identify lures that increases believability in a user [37], and (b) assess the undetectability and potency of cues to unsuspecting users [38]. Learning models are successfully used to identify attacks, and detect cues and lures [39][40] that leave users as susceptible. They identify data anomalies via learned outliers in a dataset [41] as accomplished via vote, bagging, boosting, and stacked models/schemes [42][43].

MLs are veritable tools to identify attacks [44]. A trained MLs can detect anomalous patterns – even with its dynamic predictors [45]. Learning schemes are grouped into: machine learning (ML) [46][47], deep learning (DL) [48][49], and ensemble learning (EL) [50]. ML's flexibility and robustness help it to learn intrinsic patterns and decode predictors that fastens model design, and ease outliers identification [51]. Its pitfalls are imbalanced dataset and the feature selection mode used [52][53]. DLs utilize recurrent neural networks to capture chaotic, high-dimensional data patterns [54][55].

Its poor generalization due to the vanishing gradient problem, restricts its use. But, its variant overcomes this via its gates to control its input, and eases its adaptability to learned changes as long-term dependency [56]. Its inability to handle larger dataset and longer training time required implies the quest for better alternative [57]. Lastly, ELs effectively fuses ML with DL into a stronger learner to enhance performance [58]. It must resolve conflicts of structure and data-encode [59]; while, leveraging the merits of both ML and DL to avoid model overfit as birthed by the underlying models [60][61].

Thus, we explore the XGBoost to achieve such predictive abilities, leveraging its base, weak learners to enhance itself [62]. It will improve its performance via error reduction on its weak (base) learner, and reduce its overall variance and bias in the dataset to improve generalization. It benefits from the comprehensive knowledge of its weaker base learners, to improve its generalization by exploiting the XGBoost scheme [63]. With degraded performance due to an imbalanced data [64][65], we explore the variant SMOTE-Tomek balancing. Our study wishes to [66]: (a) identify phishing lures content on spoofed website, (b) resolve data imbalance via SMOTE-Tomek, and (c) select predictors concerning the target class via the relief rank feature selection.

Resolving data imbalance via oversampling has become imperative in ML, as it accounts for the minor class as crucial [67][68]. It is opposed to under-samplers that often reduces or ignore as meaningless, the minor class in a dataset [69]. Thus, we use the synthetic minority oversample technique (SMOTE) [70], or its variants namely SMOTE-Tomek [71] and SMOTEEN [22]. Our study contributes thus: Section 1 introduces the subject with gaps that motivate the study, (b) Section 2 explores the proposed method – and leans on data collection, pre-processing, dataset split-balance-normalize via SMOTE-Tomek, the stacked model construction, training and validation with XGBoost, and (c) Section 3 – discusses the experimental results obtained as evidence in a broader sense cum context for the stacked ensemble on the phishing website dataset as obtained from UCI.

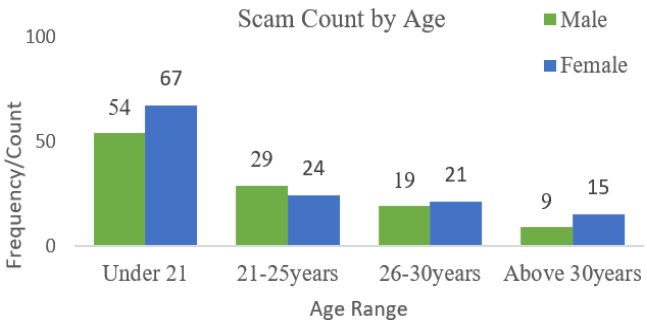

Fig. 1. Scam count by age distribution

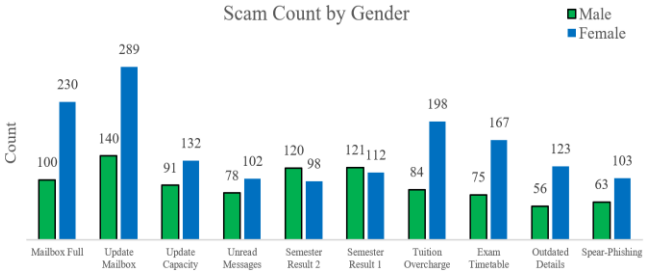

Fig. 2. Scam count by gender

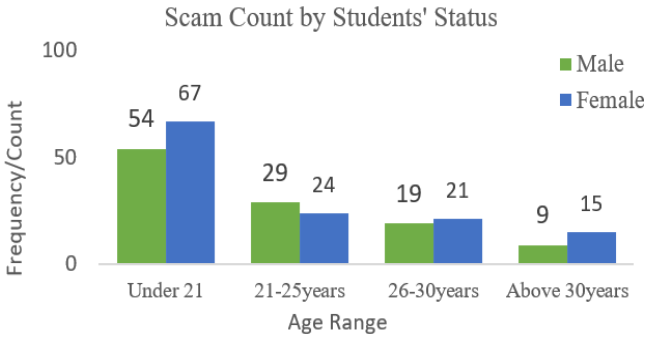

Fig. 3. Students' status by year of study

II. Materials and Methods

The stacking mode is based on Fig. 4 as thus:

- **Step-1–Data Collection:** We explore UCI phishing dataset, that consist of 11,055-records distributed into 5,180-cases in genuine class, and 5,875-cases in phishing class [72][73]. The original dataset plot is seen in Fig. 5 and detailed in Table 1.
- **Step-2–Pre-processing:** cleans up the dataset by expunging redundancies to yield integrity, and removes missing values to yield quality. The optimized data is encoded via a one-hot encoding that transforms categorical data into its equivalent binary forms [74]. Fig. 6 shows the optimized dataset.
- **Step-3–Relief Rank Feature Selection:** We select strictly, only the predictors of relevance to expunge all docile feats and reduce dataset dimensionality, to aid fastened model construction [75]. The relief rank: (a) assumes all features have same weight and influence on accuracy, (b) identifies the nearest sample from the same class as the nearest hit, and the nearest sample from a varying class as the nearest miss, and (c) uses feature value of nearest neighbor to update its weight(s) [76]. It

assesses the correlation of all predictors for a target class as in (1). With a threshold of 8.321 computed, Algorithm 1 ranks features using relief ranking to choose a total of 20 predictors as in Table 1, from the original UCI dataset with the initial 30 features.

$$Y = 100 * \sum|(x_1^2 - x_2^2)^2 + (1 - x_1^2)^2| \quad (1)$$

Algorithm 1: Relief Ranking Feature Selection Approach

1. With dataset: n ← number of train samples), a ← number of features), m ← random train samples used to update W
2. initialize all feature weights W[A]=0.0
3. for i = 1 to m do:
4. randomly select a target instance R
5. find nearest hit 'H' and nearest miss 'M' (instances)
6. for A = 1 to m do:
7. W[A] = W[A] – *diff*(A,R,H)/m + *diff*(A,R,M)/m
8. end for: end for
9. return vector W of feature scores that estimate feat quality

- **Step-4–Data Split/Balance:** First, dataset is split into train (75%) and test (25%). Balancing resample data, interpolating its nearest neighbor to create synthetic data to repopulate a pool, or removing data from its original pool to yield a more balanced dataset. We adapt SMOTE-Tomek via the SMOTE (oversampler) and Tomek-links (under-sampler) as detailed in [77][78]. Fig. 7 shows a balanced plot as resulting from the Algorithm 2 for SMOTE-Tomeks.

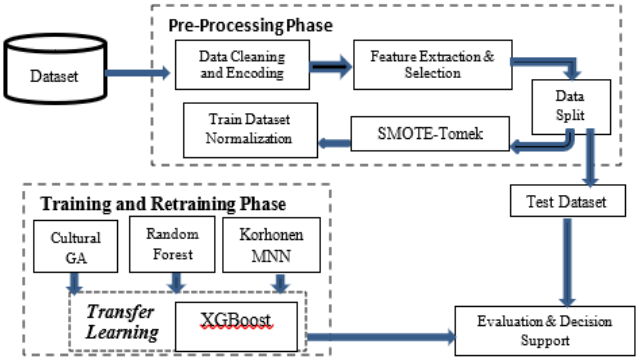

Fig. 4. Proposed stacking ensemble approach with XGBoost as meta-regressor

Table 1. Ranking of Features Engineered Using Wrapper Mode

Parameters	Description	Data Type	Selected
shortening_service	Whether a URL shortening service like bit.ly is used (1=Yes, -1=No)	char	Yes
double_slash_redirecting	Presence of "//" in the URL path (1: Yes, -1: No)	char	Yes
having_IP_Address	Whether the URL has an IP Address instead of a domain name (1=Yes, 1=No)	alphanumeric	No
having_At_Symbol	Presence of "@" symbol in the URL (1: Yes, -1: No)	char	No
having_Sub_Domain	Number of subdomains in the URL (1: More than one, 0: One, -1: None)	char	Yes
URL_lenght	Length of the URL (1=long, 0=medium, -1=short)	integer	Yes
domain_registration_length	Length of time domain has been registered (1: over a year, -1: Less than a year)	integer	Yes
Prefix_Suffix	Presence of "-" in the domain part of the URL (1: Yes, -1: No)	char	No
SSLfinal_State	Whether the website uses HTTPS with a valid SSL certificate (1: Yes, -1: No)	char	Yes
Favicon	Whether the favicon is loaded from the same domain (1: Yes, -1: No)	char	Yes
port	Use of non-standard ports (1: Yes, -1: No)	alphanumeric	No
HTTPS_token	Presence of "HTTPS" token in the URL (1: Yes, -1: No)	char	Yes
Request_URL	Percentage of external links in the source code of the website (1: High, -1: Low)	alphanumeric	No
URL_of_Anchor	Percentage of external anchor links on the website (1: High, -1: Low)	char	Yes
Links_in_tags	Percentage of external links in tags (e.g., meta, script) (1: High, -1: Low)	char	Yes
SFH	Form Handler, where form data is submitted (1: External, 0: Internal, -1: Same)	alphanumeric	Yes
Submitting_to_email	Whether the form submits data to an email address (1: Yes, -1: No)	alphanumeric	Yes
Abnormal_URL	Whether the URL is abnormal (1: Yes, -1: No)	alphanumeric	No
Redirect	Number of redirects (1: More than one, -1: Less than one)	alphanumeric	No
on_mouseover	Whether changing status bar content on mouseover (1: Yes, -1: No)	char	No
RightClick	Whether right-click is turned off on the website (1: Yes, -1: No)	char	Yes
popUpWindow	Whether pop-up windows are present (1: Yes, -1: No)	char	Yes
Iframe	Whether iframe is used on the website (1: Yes, -1: No)	char	No
age_of_domain	Age of the domain (1: More than 6 months, -1: Less than 6 months)	integer	No
DNSRecord	Whether the DNS record exists (1: Yes, -1: No)	boolean	Yes
web_traffic	Web traffic rank (1: High, 0: Medium, -1: Low)	alphanumeric	Yes
Page_Rank	Google PageRank (1: High, -1: Low)	integer	Yes
Google_Index	Whether Google indexes the site (1: Yes, -1: No)	integer	Yes
Links_pointing_to_page	Number of links pointing to the page (1: High, 0: Medium, -1: Low)	alphanumeric	Yes
Statistical_report	Whether the website is reported as a phishing site (1: Yes, -1: No)	integer	Yes

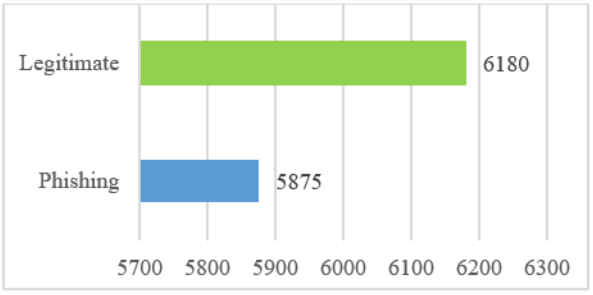

Fig. 5. Original dataset plot

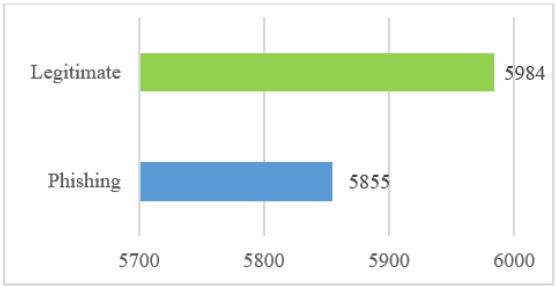

Fig. 6. Preprocessing applied to the dataset

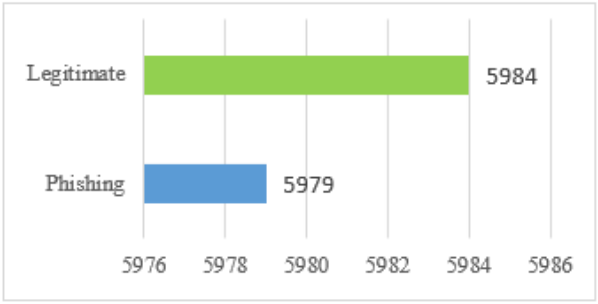

Fig. 7. SMOTE-Tomek data balancing

Algorithm 2: SMOTE-Tomek's Links Data balancing approach

1. //stratified split of dataset with train-70% and test-30% subsets
2. from sklearn.select import train_test_split, StratifiedShuffleSplit
3. xy_train, = train_test_split (testSize=0.3, stratify=y, random_state=42)
4. x_val, x_test, y_val, y_test = train_test_split(x_temp, y_temp, test_size=0.3, stratify=y_temp, random_state=42)
5. from minor_class, choose random data-point *//start SMOTE_mode*
6. compute: rel_dist from rnd_selected_data and k_nearest_neighbor
7. choose rnd_val = random_value(0,1): rnd_val * rel_distance;
8. **if** simSamples = obtained **then** minorClassNew = minorClas + simSample
9. repeat steps 2-to-4 until threshold_minor_class_new = reached;
10. select rnd_minor_class(data) //start Tomek (under-sampler) approach
11. find k_nearest_neighbor(randomized_data)
12. **if** knn.selected = minor_class_new **then** TomekLink created
13. stop TomekLink procedure: end

- **Step-5 – Stacked-Ensemble** fuses 3-base learners with the XGB meta-regressor, explained as:

1) **Cultural Genetic Algorithm uses these belief spaces as:** (a) **normative** values to which predictors are bound, (b) **domain** equip predictors with knowledge about task, (c) **temporal** ensures each predictor knows the solution, and (d) **spatial** ensures each [79] predictor with its topology. It uses an influence function to set its (upper and lower) bounds which lies between (0,1) in its quest for optimal as in (2); and (3) – allows knowledge transfer between its belief space(s) and the pool, and to alter each predictor to conform with its belief space [80]. Each $b_i \in \{1,0\}$ is a chromosome gene [81]. Table 2 is the CGA design.

$$f(x) = L_{lower} + x' \frac{L_{upper}}{2^N - 1} \quad (2)$$

$$x' = \sum_{i=0}^{N} b_i 2^i \quad (3)$$

Table 2. CGA Design and Configuration

Features	Value	Description
max_nos_gen	120	Maximum number of generations
nos_individuals	30	Number of solutions in a generation
selection_type	int	1-rank, 2-elitism, 3-steady state, 4-tourney, 5-stochastic universal sampling
offspring_created	int	Offspring: 1-crossover, 2-mutation
req_fit_function	10	Minimal number of samples needed
learning_rate	0.1	Determines the step size in learning
random_state	25	The seeds for reproduction
max_nos_gens	120	Epochs or max number of generations

2) **Random Forest** successively grows its decision trees independently via a bootstrap sample, in bagging mode [82]. It uses a binary split on its extra layer to extend the randomness on how its trees are constructed, so that its best nodes are selected randomly to capture intricate feats in the dataset [83]. Its inability to handle diversity in categorical data often results in its poor performance [84]. Thus, we tune the hyperparameters to reduce model overfitting [85]. Expressed in (4), with $normfi$ as normalized feature importance for i in tree j in (5). T is the total number of trees, and fi_i is the importance of a feature i about ground-truth, and ni_j is nodal feature importance as in (6) that yields Gini value [86]. Table 3 shows the Random Forest design configuration.

Table 3. Rf Design Configuration

Features	Value	Description
n_estimators	150	Number of trees constructed
learning_rate	0.25	Step size learning for update
max_depth	5	Max depth of each tree
min_sample_split	10	Minimal samples needed
random_state	25	The seeds for reproduction
eval_metric	error, logloss	Performance evaluation metrics
eval_set	x,val, y_val	Train data for evaluation
bootstrap	True	sets bootstrap aggregation used

$$normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_j} \quad (4)$$

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ features\ i} ni_j}{\sum_{k \in all\ nodes} ni_k} \quad (5)$$

$$ni_j = w_j c_j - w_{left(j)} c_{left(j)} - w_{right(j)} c_{right(j)} \quad (6)$$

3) **Korhonen Modular Neural Network (KMNN)** yields a deep, modular learning model that computes its output using the tan-sigmoid function. It splits a network into smaller units for enhanced dependence and improved efficacy of its units [87]. This improves its computational efficiency, reduces time convergence, and lets it handle more tasks effectively in parallel [88]. Its diversity grants each unit independent training to make KMNN more robust and flexible, with improved generalization [89][90]. Table 4 details the KMNN design configuration.

Table 4. Korhonen Modular NN Configuration

Features	Value	Description
eval_perf_set	MSE	Evaluation metrics at training
hidden_layers	10	Number of hidden layers adopted
training_percent	50	k-fold dataset used for training
transfer_hidden	tan-sigmoid	Transfer (activation) learning function
learning_rate	0.25	Step size learning to update the ensemble
number_layer	10	Minimal number of samples needed
data_division	stratified	k-fold dataset for construction
train_net_algo	LMBP	Training mode by a neural network
bkpg_momentum	auto	Backpropagation-in-momentum learn

4) **XGBoost** meta-regressor leans on the predictive output of its base models, expanding its goal function through its regularizer term $\Omega(f_t)$ and loss function $l(Y_i^t, \hat{Y}_i^t)$ [91] to ensure its solution remains within the bounds for its improved accuracy via its tuned hyperparameters [92] as in Table 5 and (7).

$$L_i = \sum_{i=1}^{n}(l(Y_i^t, \hat{Y}_i^t) + f_k(x_i)) + \Omega(f_t) \quad (7)$$

Table 5. XGB Regressor Design and Configuration

Features	Value	Description
n_estimators	250	Number of trees constructed
max_depth	5	Max depth of each tree
eval_set	x,val, y_val	Train dataset to evaluate performance
learning_rate	0.25	Step size learning to update XGBoost
eval_metric	error, logloss	Performance evaluation metrics
random_state	25	The seeds for reproduction

- **Step 6 – Train/Cross Validation:** is initialized with default configuration as in Table 2 to Table 5 to tune hyperparameters. Each tree is iteratively constructed and trained to ensure the collective knowledge is used in identifying intricate patterns. Training blends synthetic with original data that guarantees comprehensive learning, while improving its adaptability for a variety of configurations [93][94].

III. Result Findings and Discussion

A. Results, Findings, and Discussion

For a comprehensive evaluation devoid of overfitting, we use a 5-fold cross-validation on the 70% train-subset obtained via SMOTE-Tomek balancing, and a final evaluation carried out via a held-out test (30%) dataset as in Table 6. Proposed hybrid yields an average accuracy of 99.34% with a Precision of 99.6%, a Recall of 98.64%, an F1 of 99.2%, a Specificity of 99.66%, an MCC of 97.7% and an AUC-ROC of 99.6%.

From Table 6, the high value resulting in the MCC scores implies that the model accurately and consistently handles the minority class with data balancing performed; while the Specificity value of 99.66% reached indicates that the model effectively recognizes phishing, malicious websites that agree with [95]. The held-out test (30%) assesses the model's generalization ability with unseen data. The results showed an accuracy of 99.7%, precision 100%, recall of 99.8%, and F1 of 99.5%. The AUC value of 99.7% implies that the model was able to differentiate between the benign and malignant records. Also, a Specificity of 100% indicates that no benign (phishing) record was misclassified.

Table 6. Evaluation Without Feature Selection

Models	5-Fold Cross-Validation (Training)					Held-Out Test Set
	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
Accuracy	0.991	0.981	0.997	0.998	1.000	0.997
Recall	0.981	1.000	0.975	0.976	1.000	0.998
Precision	1.000	0.984	1.000	0.996	1.000	1.000
F1	0.991	0.989	0.995	0.985	1.000	0.995
MCC	0.982	0.963	0.955	0.985	1.000	0.986
Specificity	1.000	1.000	0.985	0.998	1.000	1.000
AUC-ROC	0.999	0.999	0.986	0.996	1.000	0.997

Fig. 8 is the AUC-ROC with a 99.73%, and shows the model's capability to differentiate the negative and positive classes. The proposed model accurately identified all 3,591 of the test data. With only a misclassified case and no false positives recorded [96] – Its specificity of 1.000 implies that no phishing content was misclassified. This is critical to avoid misclassification when detecting phishing. Proposed model enhances phishing website detection performance on both the training data and the held-out test set [97].

Fig. 9 implies the ensemble correctly classified all test datasets with perfect accuracy. The utilization of both feature selection, SMOTE-Tomek balancing, and data normalization did not degrade performance [98][99]. Rather, it focuses on critical feats for model construction to successfully detect spoofed websites with reduced errors that will secure user(s) resources and enhance experience.

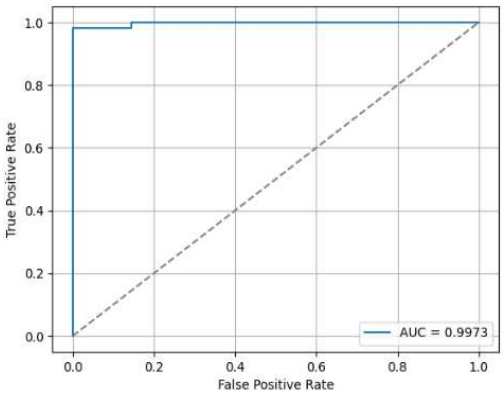

Fig. 8. ROC result of the held-out test dataset

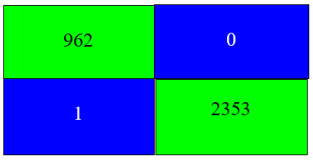

Fig. 9. Confusion Matrix

B. Comparison

As we explore the high performance of our proposed ensemble with the dataset to demonstrate its robust flexibility, adaptability, robustness, and prediction ability, we also benchmark it against previous methods that have used the same dataset [100]. Thus, we benchmark our ensemble's similar design constructs on various datasets for various domain tasks, as in Table 7 [101]. We focus on the held-out test dataset performance as it presents a more realistic indication of the model's generalization capabilities. These are summarized using the metric Table 7:

Table 7. Evaluation Without Feature Selection

	SEM + DBN Ref [25]	DHH + GRU Ref [102]	BiGRU + FSOR Ref [103]	LSTM + CNN [104]	GBM + PSO Ref [95]	Our Model
Accu.	0.973	0.919	1.000	0.992	0.969	0.997
Recall	0.974	0.959	1.000	0.989	0.976	0.998
Precis.	0.982	0.948	1.000	0.992	0.947	1.000
F1	0.976	0.973	1.000	0.985	0.974	0.995
Spec	-	0.926	-	0.991	-	1.000
ROC	0.938	-	1.000	0.987	0.958	0.997

The proposed model underperforms against [103] due to its use of BiGRU deep learning scheme with hybrid feature selection; However, other benchmark model underperformed in comparison to our proposed model, across metrics on the test dataset – achieving its high accuracy (99.7%), precision (100%), recall (99.8%) and specificity (100%) – showing best generalization with low false-positives, which is crucial in phishing detection especially with complex lures used by adversaries [105] in their evolving exploit methods. Models leverage deep learning capabilities – their performance can be seen to be slightly lower in metrics, and the lack thereof of specificity indicates that they are less robust; whereas, our model can be seen to maintain high sensitivity performance, even with its transfer learning architectures. We used the SMOTE-Tomek scheme to address class imbalances [57].

IV. Conclusion

This study presents a hybrid fusion of supervised CGA with unsupervised KMNN, and tree-based (RF and XGB) to classify websites via the UCI phishing website dataset. The model achieved high-performing discriminative ability by fusing statistically selected features using the relief ranking mode. It used SMOTE-Tomek at training successfully to mitigate imbalance in classes to yield enhanced recall and F1. Its final classification with the XGB kernel achieved a 99.7% accuracy with 100% precision on test data. The comparative analysis with benchmarks showed our method's superior generalization and data balance. Thus, our study contributes a light framework yet effective model that avoids complex training, handles larger dataset complexities, and proffers interpretability with high performance. Future work may extend this hybrid strategy to multiclass or multimodal datasets and test alternative fusion or dimensionality reduction.

References

[1] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *Journal of Computing Theories and Applications*, vol. 1, no. 4, pp. 407–420, 2024, https://doi.org/10.62411/jcta.10323.

[2] D. A. Obasuyi *et al.*, "NiCuSBlockIoT: Sensor-based Cargo Assets Management and Traceability Blockchain Support for Nigerian Custom Services," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 15, no. 2, pp. 45–64, 2024, https://doi.org/10.22624/AIMS/CISDI/V15N2P4.

[3] H. Alamleh, A. A. S. Alqahtani, and B. Al Smadi, "Secure Mobile Payment Architecture Enabling Multi-factor Authentication," *2023 Systems and Information Engineering Design Symposium, SIEDS 2023*, pp. 19–24, 2023, https://doi.org/10.1109/SIEDS58326.2023.10137778.

[4] A. A. Ojugo *et al.*, "Forging a learner-centric blended-learning framework via an adaptive content-based architecture," *Science in Information Technology Letters*, vol. 4, no. 1, pp. 40–53, 2023, https://doi.org/10.31763/sitech.v4i1.1186.

[5] L. De Kimpe, M. Walrave, W. Hardyns, L. Pauwels, and K. Ponnet, "You've got mail! Explaining individual differences in becoming a phishing target," *Telematics and Informatics*, vol. 35, no. 5, pp. 1277–1287, 2018, https://doi.org/10.1016/j.tele.2018.02.009.

[6] P. O. Ejeh *et al.*, "Counterfeit Drugs Detection in the Nigeria Pharma-Chain via Enhanced Blockchain-based Mobile Authentication Service," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 12, no. 2, pp. 25–44, 2024, https://doi.org/10.22624/aims/maths/v12n2p3.

[7] B. O. Malasowe, M. I. Akazue, E. A. Okpako, F. O. Aghware, A. A. Ojugo, and D. V. Ojie, "Adaptive Learner-CBT with Secured Fault-Tolerant and Resumption Capability for Nigerian Universities," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 135–142, 2023, https://doi.org/10.14569/IJACSA.2023.0140816.

[8] Ifoko, A.M. *et al.*, "StreamBoostE: A Hybrid Boosting-Collaborative Filter Scheme for Adaptive User-Item Recommender for Streaming Services," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 10, no. 2, pp. 89–106, 2024, https://doi.org/10.22624/aims/v10n2p8.

[9] E. A. Otorokpo *et al.*, "DaBO-BoostE: Enhanced Data Balancing via Oversampling Technique for a Boosting Ensemble in Card-Fraud Detection," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 12, no. 2, pp. 45–66, 2024, https://doi.org/10.22624/aims/maths/v12n2p4.

[10] Ifoko, A.M. *et al.*, "CoDuBoTeSS: A Pilot Study to Eradicate Counterfeit Drugs via a Blockchain Tracer Support System on the Nigerian Frontier," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 10, no. 2, pp. 53–74, 2024, https://doi.org/10.22624/aims/bij/v10n1p6.

[11] A. Ibor, E. Edim, and A. Ojugo, "Secure Health Information System with Blockchain Technology," *Journal of the Nigerian Society of Physical Sciences*, vol. 5, no. 2, p. 992, 2023, https://doi.org/10.46481/jnsps.2023.992.

[12] G. Sasikala *et al.*, "An Innovative Sensing Machine Learning Technique to Detect Credit Card Frauds in Wireless Communications," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–12, 2022, https://doi.org/10.1155/2022/2439205.

[13] M. D. Okpor *et al.*, "Unmasking effects of feature selection and SMOTE-Tomek in tree-based random forest for scorch occurrence detection," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 3, pp. 2393–2403, 2025, https://doi.org/10.11591/eei.v14i3.8901.

[14] P. Onoma *et al.*, "Attrition rate prediction using a frequency-recency-monetization-based smoteen-boosted approach," *International Journal of Advanced Computing and Intelligent System*, vol. 3, no. 1, pp. 1–11, 2025, https://msis-press.com/paper/ijacis/3/1/20.

[15] A. A. Ojugo and A. O. Eboka, "Comparative Evaluation for High Intelligent Performance Adaptive Model for Spam Phishing Detection," *Digital Technologies*, vol. 3, no. 1, pp. 9–15, 2018, https://doi.org/10.12691/dt-3-1-2.

[16] 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," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 4, pp. 455–473, 2025, https://doi.org/10.62411/faith.3048-3719-61.

[17] C. Li, N. Ding, H. Dong, and Y. Zhai, "Application of Credit Card Fraud Detection Based on CS-SVM," *International Journal of Machine Learning and Computing*, vol. 11, no. 1, pp. 34–39, 2021, https://doi.org/10.18178/ijmlc.2021.11.1.1011.

[18] A. O. Eboka *et al.*, "Pilot study on deploying a wireless sensor-based virtual-key access and lock system for home and industrial frontiers," *International Journal of Informatics and Communication Technology*, vol. 14, no. 1, pp. 287–297, 2025, https://doi.org/10.11591/ijict.v14i1.pp287-297.

[19] D. Huang, Y. Lin, Z. Weng, and J. Xiong, "Decision Analysis and Prediction Based on Credit Card Fraud Data," in *ACM International Conference Proceeding Series*, pp. 20–26, 2021, https://doi.org/10.1145/3478301.3478305.

[20] I. A. Anderson and W. Wood, "Habits and the electronic herd: The psychology behind social media's successes and failures," *Consumer Psychology Review*, vol. 4, no. 1, pp. 83–99, 2021, https://doi.org/10.1002/arcp.1063.

[21] A. Razaque *et al.*, "Credit Card-Not-Present Fraud Detection and Prevention Using Big Data Analytics Algorithms," *Applied Sciences (Switzerland)*, vol. 13, no. 1, p. 57, 2023, https://doi.org/10.3390/app13010057.

[22] V. O. Geteloma *et al.*, "Enhanced data augmentation for predicting consumer churn rate with monetization and retention strategies: a pilot study," *Applied Engineering and Technology*, vol. 3, no. 1, pp. 35–51, 2024, https://doi.org/10.31763/aet.v3i1.1408.

[23] V. O. Geteloma *et al.*, "AQuamoAS: unmasking a wireless sensor-based ensemble for air quality monitor and alert system," *Applied Engineering and Technology*, vol. 3, no. 2, pp. 70–85, 2024, https://doi.org/10.31763/aet.v3i2.1409.

[24] S. Adamu, A. Iorliam, and Ö. Asilkan, "Exploring Explainability in Multi-Category Electronic Markets: A Comparison of Machine Learning and Deep Learning Approaches," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 4, pp. 440–454, 2025, https://doi.org/10.62411/faith.3048-3719-58.

[25] S. Alnemari and M. Alshammari, "Detecting Phishing Domains Using Machine Learning," *Applied Sciences (Switzerland)*, vol. 13, no. 8, p. 4649, 2023, https://doi.org/10.3390/app13084649.

[26] M. Ahmed, K. Ansar, C. B. Muckley, A. Khan, A. Anjum, and M. Talha, "A semantic rule based digital fraud detection," *PeerJ Computer Science*, vol. 7, pp. 1–21, 2021, https://doi.org/10.7717/PEERJ-CS.649.

[27] D. R. I. M. Setiadi, A. Susanto, K. Nugroho, A. R. Muslikh, A. A. Ojugo, and H. S. Gan, "Rice Yield Forecasting Using Hybrid Quantum Deep Learning Model," *Computers*, vol. 13, no. 8, pp. 1–18, 2024, https://doi.org/10.3390/computers13080191.

[28] M. A. Haque *et al.*, "Cybersecurity in Universities: An Evaluation Model," *SN Computer Science*, vol. 4, no. 5, 2023, https://doi.org/10.1007/s42979-023-01984-x.

[29] A. A. Ojugo, C. O. Obruche, and A. O. Eboka, "Quest For Convergence Solution Using Hybrid Genetic Algorithm Trained Neural Network Model For Metamorphic Malware Detection," *ARRUS Journal of Engineering and Technology*, vol. 2, no. 1, pp. 12–23, 2021, https://doi.org/10.35877/jetech613.

[30] T. Muralidharan and N. Nissim, "Improving malicious email detection through novel designated deep-learning architectures utilizing entire email," *Neural Networks*, vol. 157, pp. 257-279, 2023, https://doi.org/10.1016/j.neunet.2022.09.002.

[31] S. F. Tan and G. C. Chung, "An Evaluation Study of User Authentication in the Malaysian FinTech Industry With uAuth Security Analytics Framework," *Journal of Cases on Information Technology*, vol. 25, no. 1, pp. 1–27, 2023, https://doi.org/10.4018/JCIT.318703.

[32] M. D. Okpor *et al.*, "Pilot Study on Enhanced Detection of Cues over Malicious Sites Using Data Balancing on the Random Forest Ensemble," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 2, pp. 109–123, 2024, https://doi.org/10.62411/faith.2024-14.

[33] A. A. Ojugo *et al.*, "CoSoGMIR: A Social Graph Contagion Diffusion Framework using the Movement-Interaction-Return Technique," *Journal of Computing Theories and Applications*, vol. 1, no. 2, pp. 163–173, 2023, https://doi.org/10.33633/jcta.v1i2.9355.

[34] M. I. Akazue *et al.*, "FiMoDeAL: pilot study on shortest path heuristics in wireless sensor network for fire detection and alert ensemble," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3534–3543, 2024, https://doi.org/10.11591/eei.v13i5.8084.

[35] V. Umarani, A. Julian, and J. Deepa, "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," *Journal of the Nigerian Society of Physical Sciences*, vol. 3, no. 4, pp. 385–394, 2021, doi: https://doi.org/10.46481/jnsps.2021.308.

[36] J. Agboi *et al.*, "Lung Cancer Detection using a Hybridized Contrast-based Xception Model on Image Data: A Pilot Study," *MSIS - international Journal of Advanced Computing and Intelligent System*, vol. 4, no. 1, pp. 1–11, 2025, https://msis-press.com/paper/ijacis/4/1/21.

[37] S. K. Shandilya, "Paradigm Shift in Adaptive Cyber Defense for Securing the Web Data: The Future Ahead," in *Journal of Web Engineering*, vol. 21, no. 4, pp. 1371-1376, 2022, https://doi.org/10.13052/jwe1540-9589.21416.

[38] 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," *Journal of Computing Theories and Applications*, vol. 1, no. 2, pp. 201–211, 2023, https://doi.org/10.33633/jcta.v1i2.9462.

[39] A. A. Ojugoa *et al.*, "Evidence of Students' Academic Performance at the Federal College of Education Asaba Nigeria: Mining Education Data," *Knowledge Engineering and Data Science*, vol. 6, no. 2, p. 145, 2023, https://doi.org/10.17977/um018v6i22023p145-156.

[40] L. R. Zuama, D. R. I. M. Setiadi, A. Susanto, S. Santosa, H.-S. Gan, and A. A. Ojugo, "High-Performance Face Spoofing Detection using Feature Fusion of FaceNet and Tuned DenseNet201," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 4, pp. 385–400, 2025, https://doi.org/10.62411/faith.3048-3719-62.

[41] A. A. Ojugo and O. Nwankwo, "Spectral-Cluster Solution For Credit-Card Fraud Detection Using A Genetic Algorithm Trained Modular Deep Learning Neural Network," *JINAV: Journal of Information and Visualization*, vol. 2, no. 1, pp. 15–24, 2021, https://doi.org/10.35877/454ri.jinav274.

[42] J. Yao, C. Wang, C. Hu, and X. Huang, "Chinese Spam Detection Using a Hybrid BiGRU-CNN Network with Joint Textual and Phonetic Embedding," *Electronics (Switzerland)*, vol. 11, no. 15, p. 2418, 2022, https://doi.org/10.3390/electronics11152418.

[43] A. Ojugo and A. O. Eboka, "An Empirical Evaluation On Comparative Machine Learning Techniques For Detection of The Distributed Denial of Service (DDoS) Attacks," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 2, no. 1, pp. 18–27, 2020, https://doi.org/10.35877/454ri.asci2192.

[44] M. I. Akazue *et al.*, "Handling Transactional Data Features via Associative Rule Mining for Mobile Online Shopping Platforms," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, pp. 530–538, 2024, https://doi.org/10.14569/IJACSA.2024.0150354.

[45] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *Journal of Computing Theories and Applications*, vol. 2, no. 1, pp. 86–101, 2024, https://doi.org/10.62411/jcta.10562.

[46] P. A. Onoma *et al.*, "Investigating an Anomaly-based Intrusion Detection via Tree-based Adaptive Boosting Ensemble," *Journal of Fuzzy Systems and Control*, vol. 3, no. 1, pp. 90–97, 2025, https://doi.org/10.59247/jfsc.v3i1.279.

[47] P. A. Onoma *et al.*, "Voice-based Dynamic Time Warping Recognition Scheme for Enhanced Database Access Security," *Journal of Fuzzy Systems and Control*, vol. 3, no. 1, pp. 81–89, 2025, https://doi.org/10.59247/jfsc.v3i1.293.

[48] J. Raphel and P. Vinod, "Heterogeneous opcode space for metamorphic malware detection," *Arabian Journal for Science and Engineering*, vol. 42, no. 2, pp. 537-558, 2017, https://doi.org/10.1007/s13369-016-2264-6.

[49] D. R. I. Moses Setiadi *et al.*, "Single Qubit Quantum Logistic-Sine XYZ-Rotation Maps: An Ultra-Wide Range Dynamics for Image Encryption," *Computers, Materials and Continua*, vol. 83, no. 2, pp. 2161–2188, 2025, https://doi.org/10.32604/cmc.2025.063729.

[50] A. P. Binitie *et al.*, "Stacked Learning Anomaly Detection Scheme with Data Augmentation for Spatiotemporal Traffic Flow," *Journal of Fuzzy Systems and Control*, vol. 2, no. 3, pp. 203–214, 2024, https://doi.org/10.59247/jfsc.v2i3.267.

[51] A. Adimabua Ojugo, "Dependable Community-Cloud Framework for Smartphones," *American Journal of Networks and Communications*, vol. 4, no. 4, p. 95, 2015, https://doi.org/10.11648/j.ajnc.20150404.13.

[52] S. Srivatsan, S. K. Bamrah, and K. S. Gayathri, "Determining the Severity of Dementia Using Ensemble Learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13773 LNCS, pp. 117–135, 2022, https://doi.org/10.1007/978-3-031-24094-2_8.

[53] A. O. Eboka *et al.*, "Resolving Data Imbalance Using a Bi-Directional Long-Short Term Memory for Enhanced Diabetes Mellitus Detection," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 1, pp. 95–109, 2025, https://doi.org/10.62411/faith.3048-3719-73.

[54] B. O. Malasowe, F. O. Aghware, M. D. Okpor, E. B. Edim, R. E. Ako, and A. A. Ojugo, "Techniques and Best Practices for Handling Cybersecurity Risks in Educational Technology Environment (EdTech)," *NIPES - Journal of Science and Technology Research*, vol. 6, no. 2, pp. 293–311, 2024, https://doi.org/10.5281/zenodo.12617068.

[55] A. A. Ojugo and A. O. Eboka, "Extending Campus Network Via Intranet and IP-Telephony for Better Performance and Service Delivery: Meeting Organizational Goals," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 1, no. 2, pp. 94–104, 2019, https://doi.org/10.35877/454ri.asci12100.

[56] M. S. Ataa, E. E. Sanad, and R. A. El-khoribi, "Intrusion detection in software defined network using deep learning approaches," *Scientific Reports*, vol. 14, no. 1, pp. 1–15, 2024, https://doi.org/10.1038/s41598-024-79001-1.

[57] B. O. Malasowe, A. E. Okpako, M. D. Okpor, P. O. Ejeh, A. A. Ojugo, and R. E. Ako, "FePARM: The Frequency-Patterned Associative Rule Mining Framework on Consumer Purchasing-Pattern for Online Shops," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 15, no. 2, pp. 15–28, 2024, https://doi.org/10.22624/aims/cisdi/v15n2p2-1.

[58] R. E. Yoro and A. A. Ojugo, "Quest for Prevalence Rate of Hepatitis-B Virus Infection in the Nigeria: Comparative Study of Supervised Versus Unsupervised Models," *American Journal of Modeling and Optimization*, vol. 7, no. 2, pp. 42–48, 2019, https://doi.org/10.12691/ajmo-7-2-2.

[59] A. N. Safriandono, D. R. I. M. Setiadi, A. Dahlan, F. Z. Rahmanti, I. S. Wibisono, and A. A. Ojugo, "Analyzing Quantum Feature Engineering and Balancing Strategies Effect on Liver Disease Classification," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 1, pp. 51–63, 2024, https://doi.org/10.62411/faith.2024-12.

[60] N. Ben Yahia, M. Dhiaeddine Kandara, and N. Bellamine BenSaoud, "Integrating Models and Fusing Data in a Deep Ensemble Learning Method for Predicting Epidemic Diseases Outbreak," *Big Data Research*, vol. 27, 2022, https://doi.org/10.1016/j.bdr.2021.100286.

[61] C. C. Odiakaose *et al.*, "Hypertension Detection via Tree-Based Stack Ensemble with SMOTE-Tomek Data Balance and XGBoost Meta-Learner," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 3, pp. 269–283, 2024, https://doi.org/10.62411/faith.3048-3719-43.

[62] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, Sep. 2018, doi: https://doi.org/10.1016/j.jbi.2018.07.014.

[63] R. E. Ako *et al.*, "Pilot Study on Fibromyalgia Disorder Detection via XGBoosted Stacked-Learning with SMOTE-Tomek Data Balancing Approach," *NIPES - Journal of Science and Technology Research*, vol. 7, no. 1, pp. 12–22, 2025, https://doi.org/10.37933/nipes/7.1.2025.2.

[64] N. Islam *et al.*, "Towards Machine Learning Based Intrusion Detection in IoT Networks," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1801–1821, 2021, https://doi.org/10.32604/cmc.2021.018466.

[65] E. V. Ugbotu *et al.*, "Transfer Learning Using a CNN Fused Random Forest for SMS Spam Detection with Semantic Normalization of Text Corpus," *NIPES - Journal of Science and Technology Research*, vol. 7, no. 2, pp. 371–382, 2025, https://doi.org/10.37933/nipes/7.2.2025.29.

[66] D. R. I. M. Setiadi *et al.*, "Integrating Hybrid Statistical and Unsupervised LSTM-Guided Feature Extraction for Breast Cancer Detection," *Journal of Computing Theories and Applications*, vol. 2, no. 4, pp. 536–552, 2025, https://doi.org/10.62411/jcta.12698.

[67] C. C. Odiakaose *et al.*, "Investigating Data Balancing Effects for Enhanced Behavioural Risk Detection in Cervical Cancer Using BiGRU: A Pilot Study," *NIPES - Journal of Science and Technology Research*, vol. 7, no. 2, pp. 319–329, 2025, https://doi.org/10.37933/nipes/7.2.2025.24.

[68] T. C. Aghaunor *et al.*, "Enhanced Scorch Occurrence Prediction in Foam Production via a Fusion SMOTE-Tomek Balanced Deep Learning Scheme," *NIPES - Journal of Science and Technology Research*, vol. 7, no. 2, pp. 330–339, 2025, https://doi.org/10.37933/nipes/7.2.2025.25.

[69] O. Sinayobye, R. Musabe, A. Uwitonze, and A. Ngenzi, "A Credit Card Fraud Detection Model Using Machine Learning Methods with a Hybrid of Undersampling and Oversampling for Handling Imbalanced Datasets for High Scores," in *Communications in Computer and Information Science*, 2023, vol. 1818 CCIS, pp. 142–155, 2023, https://doi.org/10.1007/978-3-031-34222-6_12.

[70] A. A. Ojugo, P. O. Ejeh, C. C. Odiakaose, A. O. Eboka, and F. U. Emordi, "Predicting rainfall runoff in Southern Nigeria using a fused hybrid deep learning ensemble," *International Journal of Informatics and Communication Technology*, vol. 13, no. 1, pp. 108–115, 2024, https://doi.org/10.11591/ijict.v13i1.pp108-115.

[71] F. O. Aghware *et al.*, "Effects of Data Balancing in Diabetes Mellitus Detection: A Comparative XGBoost and Random Forest Learning Approach," *NIPES - Journal of Science and Technology Research*, vol. 7, no. 1, pp. 1–11, 2025, https://doi.org/10.37933/nipes/7.1.2025.1.

[72] A. A. Ojugo and D. O. Otakore, "Redesigning Academic Website for Better Visibility and Footprint: A Case of the Federal University of Petroleum Resources Effurun Website," *Network and Communication Technologies*, vol. 3, no. 1, p. 33, 2018, https://doi.org/10.5539/nct.v3n1p33.

[73] M. D. Okpor *et al.*, "Comparative Data Resample to Predict Subscription Services Attrition Using Tree-based Ensembles," *Journal of Fuzzy Systems and Control*, vol. 2, no. 2, pp. 117–128, 2024, https://doi.org/10.59247/jfsc.v2i2.213.

[74] A. Jolicoeur-Martineau, J. J. Li, and C. M. T. Greenwood, "Statistical modeling of GxE," *Prenatal Stress and Child Development*, vol. 58, no. 11, pp. 433–466, 2021, https://doi.org/10.1007/978-3-030-60159-1_15.

[75] H. Said, B. B. S. Tawfik, and M. A. Makhlouf, "A Deep Learning Approach for Sentiment Classification of COVID-19 Vaccination Tweets," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, pp. 530–538, 2023, https://doi.org/10.14569/IJACSA.2023.0140458.

[76] N. R. Pratama, D. R. I. M. Setiadi, I. Harkespan, and A. A. Ojugo, "Feature Fusion with Albumentation for Enhancing Monkeypox Detection Using Deep Learning Models," *Journal of Computer Theory and Applications*, vol. 2, no. 3, pp. 427–440, 2025, https://doi.org/10.62411/jcta.12255.

[77] A. A. Ojugo and D. A. Oyemade, "Predicting Diffusion Dynamics Of The Coronavirus In Nigeria Through Ties-Strength Threshold On A Cascading SI-Graph," *Technology Reports of Kansai University*, vol. 62, no. 08, pp. 126–132, 2020, https://doi.org/TRKU-13-08-2020-10998.

[78] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. W. Irianandа, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *Journal of Future Artificial Intelligence Technology*, vol. 1, no. 1, pp. 23–38, 2024, https://doi.org/10.62411/faith.2024-11.

[79] C. C. Odiakaose, *et al.*, "Hybrid Genetic Algorithm Trained Bayesian Ensemble for Short Messages Spam Detection," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 12, no. 1, pp. 37–52, 2024, https://doi.org/10.22624/aims/maths/v12n1p4.

[80] C. Odiakaose *et al.*, "DeLEMPaD: Pilot Study on a Deep Learning Ensemble for Energy Market Prediction of Price Volatility and Direction. Computing, Information Systems," *2024-www.isteams.net/cisdijournal.net CISDI Journal Reference Format Christopher Odiakaose*, vol. 15, no. 1, pp. 47–62, 2024, https://doi.org/10.22624/AIMS/CISDI/V15N1P4.

[81] A. A. Ojugo, E. Ben Iwhiwhu, O. Kekeje, M. O. Yerokun, and I. J. B. Iyawa, "Malware Propagation on Social Time Varying Networks: A Comparative Study of Machine Learning Frameworks," *International Journal of Modern Education and Computer Science*, vol. 6, no. 8, pp. 25–33, 2014, https://doi.org/10.5815/ijmecs.2014.08.04.

[82] F. U. Emordi, *et al.*, "TiSPHiMME: Time Series Profile Hidden Markov Ensemble in Resolving Item Location on Shelf Placement in Basket Analysis," *Advances in Multidisciplinary & Scientific Research Journal Publications*, vol. 12, no. 1, pp. 33–48, 2024, https://doi.org/10.22624/aims/digital/v11n4p3.

[83] B. F. Alkhalil, Y. Zhuang, K. T. Mursi and A. O. Aseeri, "Game-Theory-Based Analysis of Key Factors Influencing Saudi Consumer Trust in E-commerce," *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1-7, 2024, https://doi.org/10.1109/HORA61326.2024.10550488.

[84] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *Journal of Information Security and Applications*, vol. 55, p. 102596, 2020, https://doi.org/10.1016/j.jisa.2020.102596.

[85] F. Omoruwou, A. A. Ojugo, and S. E. Ilodigwe, "Strategic Feature Selection for Enhanced Scorch Prediction in Flexible Polyurethane Form Manufacturing," *Journal of Computing Theories and Applications*, vol. 1, no. 3, pp. 346–357, 2024, https://doi.org/10.62411/jcta.9539.

[86] A. R. Muslikh, D. R. I. M. Setiadi, and A. A. Ojugo, "Rice Disease Recognition Using Transfer Learning Xception Convolutional Neural Network," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1535–1540, 2023, https://doi.org/10.52436/1.jutif.2023.4.6.1529.

[87] P. Kumari and S. Mittal, "Fraud Detection System for Financial System Using Machine Learning Techniques: A Review," *2024 11th International Conference on Reliability, Infocom*

Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1-6, 2024, https://doi.org/10.1109/ICRITO61523.2024.10522197.

[88] A. A. Ojugo and C. O. Obruche, "Empirical Evaluation for Intelligent Predictive Models in Prediction of Potential Cancer Problematic Cases In Nigeria," *ARRUS Journal of Mathematics and Applied Science*, vol. 1, no. 2, pp. 110–120, 2021, https://doi.org/10.35877/mathscience614.

[89] S. N. Okofu *et al.*, "Pilot Study on Consumer Preference, Intentions and Trust on Purchasing-Pattern for Online Virtual Shops," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, pp. 804–811, 2024, https://doi.org/10.14569/IJACSA.2024.0150780.

[90] A. A. Ojugo *et al.*, "Forging a User-Trust Memetic Modular Neural Network Card Fraud Detection Ensemble: A Pilot Study," *Journal of Computing Theories and Applications*, vol. 1, no. 2, pp. 50–60, 2023, https://doi.org/10.33633/jcta.v1i2.9259.

[91] S. Ju *et al*., "Optimal county-level crop yield prediction using MODIS-based variables and weather data: A comparative study on machine learning models," *Agricultural and Forest Meteorology*, vol. 307, p. 108530, 2021, https://doi.org/10.1016/j.agrformet.2021.108530.

[92] Y. Abakarim, M. Lahby, and A. Attioui, "An efficient real time model for credit card fraud detection based on deep learning," in *ACM International Conference Proceeding Series*, pp. 1–7, 2018, https://doi.org/10.1145/3289402.3289530.

[93] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients," *Decision Analytics Journal*, vol. 7, p. 100242, 2023, https://doi.org/10.1016/j.dajour.2023.100242.

[94] C. L. Kumar *et al.*, "Metaparameter optimized hybrid deep learning model for next generation cybersecurity in software defined networking environment," *Scientific Reports*, vol. 15, no. 1, 2025, https://doi.org/10.1038/s41598-025-96153-w.

[95] M. D. Okpor *et al*., "Pilot study on enhanced detection of cues over malicious sites using data balancing on the random forest ensemble," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 2, pp. 109-123, 2024, https://doi.org/10.62411/faith.2024-14.

[96] K. Muhamada, D. R. I. M. Setiadi, U. Sudibyo, B. Widjajanto, and A. A. Ojugo, "Exploring Machine Learning and Deep Learning Techniques for Occluded Face Recognition: A Comprehensive Survey and Comparative Analysis," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 2, pp. 160–173, 2024, https://doi.org/10.62411/faith.2024-30.

[97] S. Pavithra and K. Venkata Vikas, "Detecting Unbalanced Network Traffic Intrusions With Deep Learning," in *IEEE Access*, vol. 12, pp. 74096-74107, 2024, https://doi.org/10.1109/ACCESS.2024.3405187.

[98] D. Sun, M. Liu, M. Li, Z. Shi, P. Liu, and X. Wang, "DeepMIT: A Novel Malicious Insider Threat Detection Framework based on Recurrent Neural Network," in *Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021*, May 2021, pp. 335–341, 2021, https://doi.org/10.1109/CSCWD49262.2021.9437887.

[99] A. A. Ojugo, P. O. Ejeh, O. C. Christopher, A. O. Eboka, and F. U. Emordi, "Improved distribution and food safety for beef processing and management using a blockchain-tracer support framework," *International Journal of Informatics and Communication Technology*, vol. 12, no. 3, pp. 205–213, 2023, https://doi.org/10.11591/ijict.v12i3.pp205-213.

[100] R. E. Yoro *et al.*, "Adaptive DDoS detection mode in software-defined SIP-VoIP using transfer learning with boosted meta-learner," *Plos One*, vol. 20, no. 6, p. e0326571, 2025, https://doi.org/10.1371/journal.pone.0326571.

[101] L. K. Y. Loh *et al*., "An ensembling architecture incorporating machine learning models and genetic algorithm optimization for forex trading," *FinTech*, vol. 1, no. 2, pp. 100-124, 2022, https://doi.org/10.3390/fintech1020008.

[102] L. Lakshmi, M. P. Reddy, C. Santhaiah, and U. J. Reddy, "Smart Phishing Detection in Web Pages using Supervised Deep Learning Classification and Optimization Technique ADAM," *Wireless Personal Communications*, vol. 118, no. 4, pp. 3549–3564, 2021, https://doi.org/10.1007/s11277-021-08196-7.

[103] D. R. I. M. Setiadi, S. Widiono, A. N. Safriandono, and S. Budi, "Phishing Website Detection Using Bidirectional Gated Recurrent Unit Model and Feature Selection," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 2, pp. 75–83, 2024, https://doi.org/10.62411/faith.2024-15.

[104] Y. A. Alsariera, A. V. Elijah, and A. O. Balogun, "Phishing Website Detection: Forest by Penalizing Attributes Algorithm and Its Enhanced Variations," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10459–10470, 2020, https://doi.org/10.1007/s13369-020-04802-1.

[105] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. C. Odiakaose, F. U. Emordi, and A. A. Ojugo, "DeLClustE: Protecting Users from Credit-Card Fraud Transaction via the Deep-Learning Cluster Ensemble," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 94–100, 2023, https://doi.org/10.14569/IJACSA.2023.0140610.