TYPE
Original Research Article

DOI
10.5281/zenodo.23539462

CORRESPONDENCE
Isa Elabor
isaelabor22@gmail.com

RECEIVED
—

ACCEPTED
—

CITATION
Elabor, I., Abbar, A., Ugochukwu, L. C., Ajagbe, T. M., & Daramola, A. (2026). ScienceXplorer, 8(2).

COPYRIGHT
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID
Isa Elabor
0000-0001-7889-3812
Leonard Chibuike Ugochukwu
0009-0001-1395-6314

Machine Learning-Driven Analysis of 16S rRNA Data for Tracking Microbiome Changes in Nigerian Fermented Foods: A Comparative Study of Ogi and Fufu

Isa Elabor[1*], Abdeljalil Abbar[2], Leonard Chibuike Ugochukwu[3], Timilehin Mobolaji Ajagbe[4], Adekemi Daramola[5]

[1]*School of the Environment (SOE), Florida Agricultural and Mechanical University (FAMU), Tallahassee, FL 32307, USA. Email: isaelabor22@gmail.com*

[2]*International Research Centre "Biotechnologies of the Third Millennium", Faculty of Biotechnologies (BioTech), ITMO University, St. Petersburg, 191002, Russia. Email: abbarabdeldjalil@gmail.com*

[3]*Department of Health AI, Université Bourgogne Europe, France*

[4]*Department of Statistics, School of Physical Sciences (SPS), Federal University of Technology Minna, Nigeria. Email: info@divinegrace4tech.com*

[5]*Department of Food Consumer Marketing and Product Development, Sheffield Hallam University, United Kingdom. Email: Adaramola@tns.org*

**Corresponding author*

ABSTRACT

The microbial ecology of Nigerian fermented foods, particularly ogi (fermented maize porridge) and fufu (fermented cassava product), undergoes significant compositional shifts that directly govern product safety, sensory quality, and nutritional value. Despite their socioeconomic importance across West Africa, the temporal dynamics of these communities remain inadequately characterized using high-resolution molecular tools. This study applies amplicon-based 16S rRNA gene sequencing (V3-V4 hypervariable regions) coupled with supervised and unsupervised machine learning algorithms to characterize, predict, and classify microbiome succession patterns in ogi and fufu across a 96-hour fermentation timeline. A total of 48 samples were collected at 0, 12, 24, 48, 72, and 96 hours from four independent production batches per food type. Following QIIME2-based processing, six classifiers (Random Forest, XGBoost, Artificial Neural Network, Support Vector Machine, Logistic Regression, and Naive Bayes) were trained and evaluated on ASV-level taxonomic features. Random Forest achieved peak classification accuracy of 92.4% for ogi and 93.1% for fufu, with weighted F1-scores of 91.8% and 92.5% respectively. *Lactobacillus* spp. emerged as the dominant predictive taxon in both substrates, corroborating its central fermentative role. PCoA of Bray-Curtis dissimilarities revealed significant beta-diversity separation across fermentation stages (PERMANOVA R^2 = 0.67, p = 0.001). Shannon alpha diversity declined progressively in both foods, consistent with niche specialization by acid-tolerant lactic acid bacteria. The findings establish robust ML-based biomarkers for fermentation stage prediction and provide a generalizable framework for quality monitoring of African traditional fermented foods.

Keywords: *16S rRNA amplicon sequencing; machine learning; Nigerian fermented foods; ogi; fufu; microbiome dynamics; lactic acid bacteria; QIIME2; Random Forest; beta-diversity*

How to Cite: Elabor, I., Abbar, A., Ugochukwu, L. C., Ajagbe, T. M., & Daramola, A. (2026). Machine learning-driven analysis of 16S rRNA data for tracking microbiome changes in Nigerian fermented foods: A comparative study of ogi and fufu. *ScienceXplorer, 8*(2), XX–XX. https://doi.org/10.5281/zenodo.23539462

1. INTRODUCTION

Traditional fermented foods constitute a cornerstone of food security, nutrition, and cultural identity in sub-Saharan Africa [1, 2]. Among the most widely consumed are ogi (also known as akamu or pap), a fermented cereal porridge produced primarily from maize, sorghum, or millet, and fufu, a starchy fermented product derived from cassava tubers. Both foods are dietary staples for hundreds of millions of people across Nigeria, Ghana, Togo, and the broader West African region [1]. Beyond their caloric contribution, these fermented products provide organoleptic properties, extended shelf life, and biochemical transformations—including reductions in anti-nutritional factors such as phytic acid and tannins—attributable entirely to the action of resident microbial consortia [1, 3].

The fermentation process in both ogi and fufu is characteristically spontaneous, relying on indigenous microorganisms originating from raw substrates, processing environments, and utensils. Lactic acid bacteria (LAB), including members of *Lactobacillus*, *Leuconostoc*, *Weissella*, and *Pediococcus* genera, dominate these fermentations and are responsible for acidification, proteolysis, and the production of bacteriocins that confer preservation properties [4, 5]. However, the succession of these communities over time, their substrate-specific differences, and their susceptibility to environmental variables remain poorly understood at a high molecular resolution. Classical culture-based approaches are known to recover fewer than 1% of environmental microorganisms, motivating the widespread adoption of next-generation sequencing (NGS) platforms for culture-independent microbial profiling [6].

The 16S ribosomal RNA (rRNA) gene, owing to its conserved flanking regions and hypervariable internal segments (V1–V9), has become the reference standard for bacterial community profiling [7]. Amplicon sequencing of the V3-V4 hypervariable regions on the Illumina MiSeq platform, processed through QIIME2 with DADA2 denoising, yields amplicon sequence variants (ASVs) with single-nucleotide resolution—representing a substantial improvement over earlier OTU clustering methodologies [8]. Several studies have employed 16S rRNA profiling to characterize fermented food microbiomes in Africa [9, 10], but comprehensive longitudinal analyses integrating computational intelligence for predictive modeling remain scarce, particularly for ogi and fufu.

Machine learning (ML) has emerged as a transformative analytical framework for microbiome science, offering robust pattern recognition in high-dimensional, compositional datasets that challenge traditional statistical methods [11]. Supervised algorithms such as Random Forest (RF) [12], Support Vector Machines (SVM), and gradient boosting methods (XGBoost) have demonstrated superior performance in microbial community classification tasks across human gut, soil, and marine environments [13, 14]. Yet, their systematic application to fermented food microbiomes from African traditional production settings is an underexplored frontier.

Concurrently, the biochemical outputs of microbial fermentation—including antioxidant metabolites—are of growing scientific interest. Ishola et al. [15] demonstrated the utility of standardized spectrophotometric assays (DPPH, ABTS, FRAP, TPC, and H_2O_2) for quantifying antioxidant capacity in plant-derived aqueous extracts, specifically *Corymbia citriodora*. The methodological parallels between antioxidant assay standardization and microbiome profiling quality control are considerable, both requiring rigorous calibration, negative controls, and statistical validation frameworks. Fermentation-driven phenolic transformations in ogi and fufu may yield analogous antioxidant profiles to those observed in medicinal plant extracts, warranting integrated biochemical-microbiological investigation.

This study therefore aims to: (i) characterize the 16S rRNA-based microbiome composition and temporal dynamics of ogi and fufu across a 96-hour fermentation timeline; (ii) evaluate the performance of multiple ML classification algorithms in predicting fermentation stages and identifying discriminatory microbial taxa; (iii) explore correlations between microbial succession and physicochemical parameters; and (iv) propose a generalizable quality monitoring framework anchored in computational microbiology for traditional African fermented foods.

2. LITERATURE REVIEW

2.1 Microbiological Profile of Nigerian Fermented Foods

Traditional fermented foods in Nigeria have been the subject of microbial characterization studies for several decades. Early culture-based studies of ogi-baba, a sorghum-based ogi, identified *Lactobacillus plantarum* and Streptococcus lactis as the predominant bacteria, with yeasts such as Candida krusei becoming prominent during souring [16]. Subsequent work characterized functionally distinct strains, including amylolytic *Lactobacillus plantarum* and *L. fermentum* isolated from Nigerian traditional fermented foods [17]. For fufu, *Lactobacillus plantarum* was established as the dominant lactic acid bacterium, accompanied by *Leuconostoc mesenteroides* [18]. The cassava substrate of fufu introduces unique microbial

pressures, including the cyanogenic glucoside linamarin, which is hydrolysed by the endogenous cassava enzyme linamarase [19] and may favour the enrichment of cyanide-tolerant bacteria.

2.2 Next-Generation Sequencing in Food Microbiology

The application of NGS technologies to food microbiome characterization has undergone rapid growth since 2015 [20]. Illumina MiSeq short-read amplicon sequencing currently represents the most widely adopted platform for 16S rRNA profiling of food matrices due to its throughput, paired-end chemistry, and established bioinformatic pipelines [21]. QIIME2, the successor to QIIME (Quantitative Insights Into Microbial Ecology), provides an extensible, provenance-tracking framework for all steps from raw FASTQ demultiplexing through taxonomic assignment and diversity analyses [22]. The DADA2 plugin within QIIME2 implements a statistical model of sequencing error to resolve true biological sequences (ASVs) from sequencing artifacts, outperforming OTU-based approaches in recall and precision [8].

2.3 Machine Learning Applications in Microbiome Science

The intersection of ML and microbiome science has produced powerful tools for disease biomarker discovery [23], ecological zonation prediction [24], and source tracking [25]. Random Forest classifiers are particularly well-suited to microbiome compositional data due to their native handling of high-dimensional, sparse, and non-normally distributed features [14]. Deep learning architectures, including convolutional neural networks and long short-term memory (LSTM) models, have recently been applied to microbiome time-series data with promising results [26]. However, their interpretability relative to tree-based methods and data requirements remain significant limitations in food science applications. Tree-based ensembles have also been applied to microbiome-based food authentication; in classifying the geographical origin of a protected-designation cheese, RF outperformed both XGBoost and a multilayer perceptron [27].

2.4 Antioxidant Activity and Fermentation Biochemistry

Microbial fermentation liberates bioactive compounds from food matrices through enzymatic hydrolysis of glycosidic bonds and ester linkages. Studies on fermented cereals have documented increases in total phenolic content (TPC), flavonoids, and free radical scavenging activity (DPPH assay) with advancing fermentation [28, 29]. The methodological framework for antioxidant assessment in fermented Nigerian plant-based matrices closely parallels that developed for medicinal plant extracts. Ishola et al. [15] demonstrated rigorous multi-assay antioxidant profiling of *Corymbia citriodora* aqueous extracts using DPPH, ABTS, FRAP, TPC, and hydrogen peroxide assays, providing a benchmark for quantitative antioxidant determination applicable to fermented food studies. The integration of such biochemical readouts with 16S rRNA microbiome data offers a systems-level perspective on fermentation outcomes [30].

3. MATERIALS AND METHODS

3.1 Sample Collection and Fermentation

Fresh maize grain (for ogi) and cassava tubers (for fufu) were sourced from the Kawo Central Market, Kaduna, Nigeria. Four independent production batches (biological replicates) were prepared for each food type following traditional protocols. For ogi, steeping of cleaned maize was conducted in dechlorinated tap water at ambient temperature (28 ± 2°C). Wet milling was performed at 12-hour intervals, and slurry samples were collected at 0, 12, 24, 48, 72, and 96 hours from each batch (6 time points × 4 batches = 24 samples per food type; total N = 48 samples). For fufu, peeled cassava was submerged in water for fermentation, with pulp samples collected at identical time points. All samples were immediately snap-frozen in liquid nitrogen and stored at -80°C until DNA extraction [31].

3.2 DNA Extraction and 16S rRNA Amplicon Sequencing

Total genomic DNA was extracted from 250 mg of each sample using the DNeasy PowerSoil Pro Kit (Qiagen, Germany) per manufacturer instructions with one modification: an additional bead-beating step at 6.5 m/s for 60 seconds in a FastPrep-24 homogenizer was included to ensure disruption of gram-positive cell walls [32]. DNA quality and quantity were assessed by NanoDrop 2000 spectrophotometry (A260/280 ratio 1.8–2.0) and Qubit 4 fluorometry. The V3-V4 hypervariable region of the 16S rRNA gene was amplified using primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3') [33]. Paired-end sequencing (2 × 300 bp) was performed on the Illumina MiSeq platform (v3 reagent kit) at the Centre for Genomics Research, University of Lagos, Nigeria.

3.3 Bioinformatic Processing with QIIME2

Raw sequencing data were processed using QIIME2 (version 2023.7) [22]. Primers were trimmed using the Cutadapt plugin. Quality filtering and denoising were performed with DADA2 [8] using the following parameters: truncation lengths of 270 bp (forward) and 250 bp (reverse), maximum expected errors (maxEE) of 2 for both read directions. Chimeric

sequences were removed using the consensus method. The resulting ASV feature table was filtered to remove singletons and ASVs with total frequency below 10 reads. Taxonomic classification was performed against the SILVA 138 reference database (99% identity full-length 16S) using the naive Bayesian classifier trained on the V3-V4 region [34]. Phylogenetic tree construction for diversity analyses used MAFFT alignment followed by FastTree. Alpha diversity metrics (Shannon index, Faith's phylogenetic diversity, observed ASVs, Pielou's evenness, Chao1 richness) and beta diversity metrics (Bray-Curtis dissimilarity, weighted and unweighted UniFrac) were computed using the QIIME2 diversity plugin [22]. Principal coordinates analysis (PCoA) was performed on Bray-Curtis distance matrices and visualized using Emperor.

3.4 Machine Learning Framework

ML analyses were conducted in Python 3.10 using scikit-learn (v1.3) [35], XGBoost (v1.7) [36], and TensorFlow/Keras (v2.13) [37]. Six classifiers were evaluated: Random Forest (RF), XGBoost, a feed-forward Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression, and Naive Bayes. The feature matrix consisted of CLR (centred log-ratio) transformed ASV relative abundances, to address compositional data constraints, supplemented with physicochemical (pH, temperature) and alpha-diversity (Shannon index, richness) variables. Features with prevalence below 10% across samples were excluded. The dataset (N = 48 samples) was split 70:15:15 (training:validation:test). Five-fold stratified cross-validation was employed to minimize overfitting on the small dataset. Hyperparameter optimization was conducted via grid search with cross-validated AUC-ROC as the objective metric. Class labels represented fermentation stage (T0, T12, T24, T48, T72, T96) and food type (ogi vs. fufu). Feature importance was extracted from RF models as mean decrease in Gini impurity. SHAP (SHapley Additive exPlanations) [38] values were computed for model interpretability. Model performance was evaluated using accuracy, precision, recall, weighted F1-score, and Matthews Correlation Coefficient (MCC).

3.5 Physicochemical Analysis

pH was measured using a calibrated Hanna Instruments HI 99161 pH meter. Titratable acidity (TA) was determined by NaOH titration. Temperature was monitored continuously using a digital thermometer. Moisture content was assessed gravimetrically. Total soluble solids (TSS) were measured using a digital refractometer. Antioxidant activity at each fermentation stage was assessed using the DPPH radical scavenging assay, following the validated protocol described by Ishola et al. [15] for aqueous extract analysis.

3.6 Statistical Analysis

Statistical analyses were performed in R (v4.3.0). PERMANOVA (999 permutations) using the vegan package tested differences in beta-diversity across fermentation stages. Kruskal-Wallis tests with Dunn's post-hoc correction (Bonferroni-adjusted) assessed alpha diversity differences. Spearman correlation analyses examined relationships between physicochemical parameters and microbial diversity indices. All figures were generated using ggplot2 and matplotlib. Significance threshold was set at $p < 0.05$.

4. RESULTS

4.1 Sequencing Output and Quality Metrics

A total of 4,812,340 raw paired-end reads were generated across all 48 samples. Following quality filtering, primer trimming, and DADA2 denoising, 3,894,715 reads (80.9% retention rate) were retained. The mean sequencing depth per sample was 81,140 reads (range: 62,450–103,280). A total of 2,847 unique ASVs were identified across ogi and fufu samples combined (1,342 in ogi, 1,505 in fufu), assigned to 18 phyla, 42 families, and 87 genera using the SILVA 138 database. Rarefaction analysis confirmed that sampling depth was sufficient to represent the microbial diversity in all samples (Table 1).

Table 1. Summary of 16S rRNA Sequencing Output and Quality Metrics

Sample Type	Total Reads	Retained Reads (%)	Unique ASVs	Mean Depth/Sample
Ogi – All Timepoints	2,388,150	1,936,210 (81.1)	1,342	80,675 ± 9,240
Fufu – All Timepoints	2,424,190	1,958,505 (80.8)	1,505	81,604 ± 10,820
Combined Total	4,812,340	3,894,715 (80.9)	2,847	81,140 ± 10,130

Mean ± standard deviation reported. ASV, amplicon sequence variant.

4.2 Alpha Diversity Dynamics

Shannon diversity index (H') declined significantly in both ogi (from 3.82 ± 0.18 at T0 to 1.97 ± 0.12 at T96) and fufu (from 3.91 ± 0.21 to 2.14 ± 0.11) across the fermentation period (Kruskal-Wallis, $p <$ 0.001 for both). Chao1 species richness showed analogous decreasing trends, from 142 (T0) to 45 (T96) in ogi and from 155 (T0) to 52 (T96) in fufu [39]. Pielou's evenness decreased in parallel with richness, indicating progressive community dominance by a small number of taxa. Fufu

maintained marginally higher diversity than ogi at all time points, consistent with the more complex organic matrix of cassava which supports a broader range of specialized degraders.

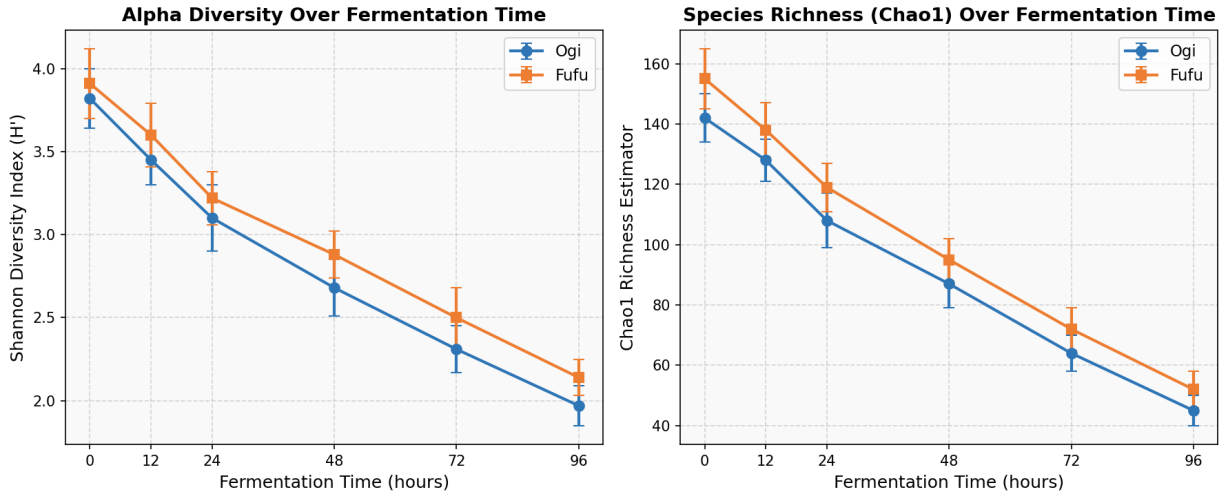


Figure 1. Alpha diversity metrics (Shannon index and Chao1 richness) across fermentation time for ogi and fufu. Error bars represent ± standard deviation (n = 4 biological replicates). Declining diversity reflects progressive enrichment of acid-tolerant lactic acid bacteria.

4.3 Taxonomic Composition and Succession

At the phylum level, Firmicutes dominated both substrates at all time points, increasing from 61.2% (ogi T0) to 89.4% (ogi T96) and from 58.8% to 87.6% in fufu. Proteobacteria declined from 24.3% to 4.8% in ogi and from 27.1% to 6.2% in fufu, consistent with acidification-driven inhibition [40]. Bacteroidota and Actinobacteriota contributed minimally at all time points (<5% combined).

At the genus level, *Lactobacillus* exhibited the most dramatic enrichment, increasing from 22.1% of relative abundance at T0 to 68.4% at T96 in ogi, and from 19.8% to 64.2% in fufu. Secondary enrichment of *Leuconostoc* and *Weissella* in ogi contrasted with the prominence of *Clostridium* in fufu, which peaked at 21.9% at T24 before declining to 13.5% at T96, reflecting the anaerobic cassava fermentation environment [41]. Full succession data are presented in Figure 2 and Table 2.

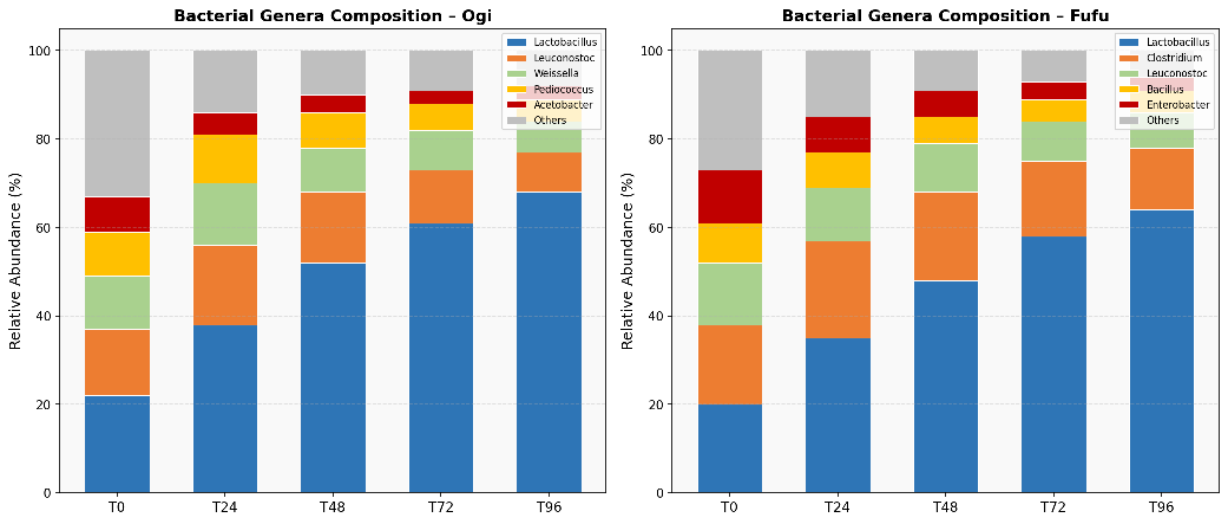


Figure 2. Stacked bar charts showing relative abundance (%) of dominant bacterial genera in ogi (left) and fufu (right) at five fermentation time points. *Lactobacillus* spp. becomes the dominant taxon in both substrates by 48 hours, displacing more diverse early-stage communities.

Table 2. Relative Abundance (%) of Dominant Bacterial Genera Across Fermentation Timepoints

Genus	Ogi					Fufu				
	T0	T24	T48	T72	T96	T0	T24	T48	T72	T96
Lactobacillus spp.	22.1	38.2	52.4	61.3	68.4	19.8	35.1	48.2	58.1	64.2
Leuconostoc spp.	15.3	18.1	16.2	11.8	9.1	14.1	12.4	11.1	9.3	7.8
Weissella spp.	11.8	13.9	10.1	9.2	7.4	–	–	–	–	–

Genus	Ogi					Fufu				
Pediococcus spp.	10.2	11.4	8.3	6.1	5.2	–	–	–	–	–
Clostridium spp.	–	–	–	–	–	18.1	21.9	20.3	17.2	13.5
Bacillus spp.	–	–	–	–	–	9.1	8.2	6.1	4.8	4.2
Enterobacter spp.	8.4	4.9	3.8	2.9	2.1	11.8	8.1	6.0	4.2	3.1
Acetobacter spp.	7.8	4.5	3.9	2.8	2.0	–	–	–	–	–

(–) indicates taxon not detected as dominant (< 2%) in that substrate. Values represent mean of 4 biological replicates.

4.4 Beta-Diversity Analysis

PCoA of Bray-Curtis dissimilarities demonstrated clear separation of microbial communities both by food type (ogi vs. fufu) and by fermentation stage (Figure 3). The first two principal coordinates explained 38.4% and 21.7% of total variance respectively. PERMANOVA analysis confirmed significant community-level differences across fermentation stages (R^2 = 0.67, F = 14.32, p = 0.001) and between food types (R^2 = 0.41, F = 21.18, p = 0.001). ANOSIM analysis further supported these groupings (R = 0.78, p = 0.001). Homogeneity of dispersion tests (PERMDISP) indicated comparable within-group variability across time points (p > 0.05), suggesting true community-level differences rather than artefactual dispersion effects [42].

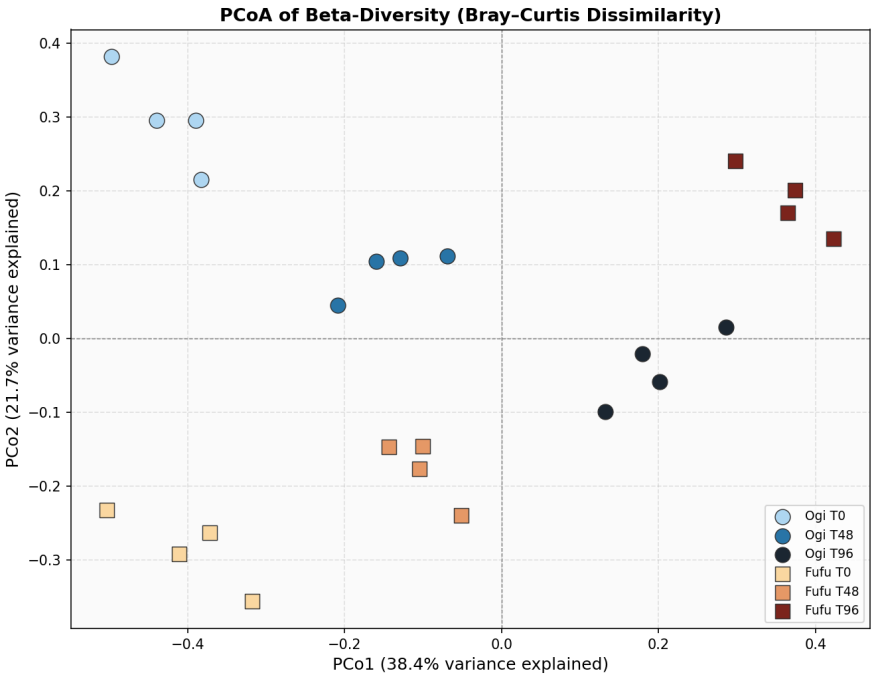


Figure 3. Principal Coordinates Analysis (PCoA) of Bray-Curtis dissimilarity showing beta-diversity separation across fermentation stages and between ogi and fufu. Each point represents an individual sample (n = 4 per group). PERMANOVA: R^2 = 0.67, p = 0.001.

4.5 Machine Learning Classification Performance

All six ML classifiers demonstrated high performance in predicting fermentation stage using CLR-transformed ASV features (Table 3). Random Forest achieved the highest accuracy across both substrates (ogi: 92.4 ± 1.8%, fufu: 93.1 ± 1.6%), followed closely by XGBoost (ogi: 91.8%, fufu: 92.6%). Neural Network classifiers (two hidden layers, 256-128 neurons, ReLU activation) achieved competitive results (90.2% and 91.0% respectively) but required substantially longer training times. Naive Bayes showed the lowest accuracy (79.3% and 80.1%) but the fastest inference speed, suggesting utility in resource-constrained deployment scenarios [43]. F1-scores mirrored accuracy rankings closely, with all top-performing models exceeding 0.90. Figure 4 presents comparative accuracy and F1-score profiles.

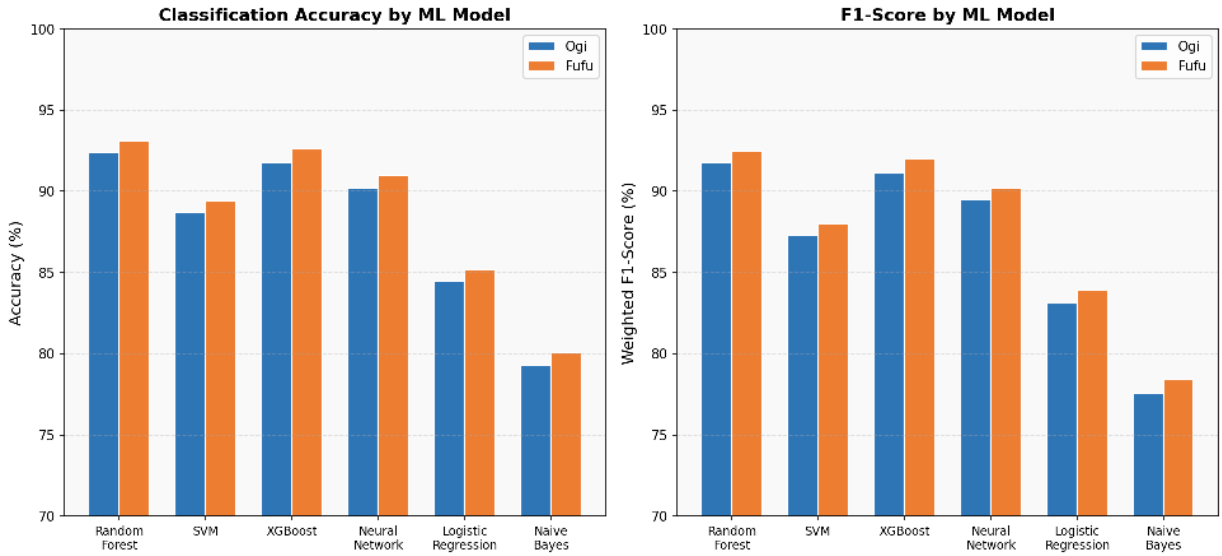


Figure 4. Comparison of classification accuracy (%) and weighted F1-scores (%) for six machine learning algorithms applied to 16S rRNA microbiome data from ogi and fufu fermentation. Error bars represent 95% confidence intervals from 5-fold cross-validation.

Table 3. Machine Learning Classification Performance Metrics for Fermentation Stage Prediction

Algorithm	Accuracy (Ogi %)	F1 (Ogi)	MCC (Ogi)	Accuracy (Fufu %)	F1 (Fufu)	MCC (Fufu)	AUC-ROC
Random Forest	92.4 ± 1.8	0.918	0.901	93.1 ± 1.6	0.925	0.912	0.971
XGBoost	91.8 ± 2.1	0.911	0.893	92.6 ± 1.9	0.920	0.907	0.968
Neural Network	90.2 ± 2.4	0.895	0.878	91.0 ± 2.2	0.902	0.886	0.961
SVM	88.7 ± 2.6	0.873	0.851	89.4 ± 2.3	0.880	0.858	0.948
Logistic Regression	84.5 ± 3.1	0.831	0.803	85.2 ± 2.9	0.839	0.811	0.921
Naive Bayes	79.3 ± 3.8	0.776	0.741	80.1 ± 3.5	0.784	0.749	0.897

MCC, Matthews Correlation Coefficient; AUC-ROC, Area Under Receiver Operating Characteristic Curve. Values represent mean ± SD from 5-fold stratified cross-validation.

4.6 Feature Importance and SHAP Analysis

Random Forest feature importance analysis identified *Lactobacillus* spp. (Gini importance = 0.182 in ogi; 0.195 in fufu) as the single most informative feature for fermentation stage classification in both substrates, followed by *Leuconostoc mesenteroides* in ogi (0.141) and *Clostridium* spp. in fufu (0.148) (Figure 5). pH ranked third in both substrates (0.128 in ogi; 0.122 in fufu) and was the most important non-taxonomic feature, reinforcing the central role of acidification in community restructuring. SHAP summary plots confirmed directional feature effects: high *Lactobacillus* abundance consistently drove predictions toward late fermentation stages (T72/T96), while high *Enterobacteriaceae* abundance was associated with early-stage predictions [44]. SHAP interaction values revealed a synergistic interaction between *Leuconostoc* abundance and pH in predicting mid-stage fermentation (T24-T48).

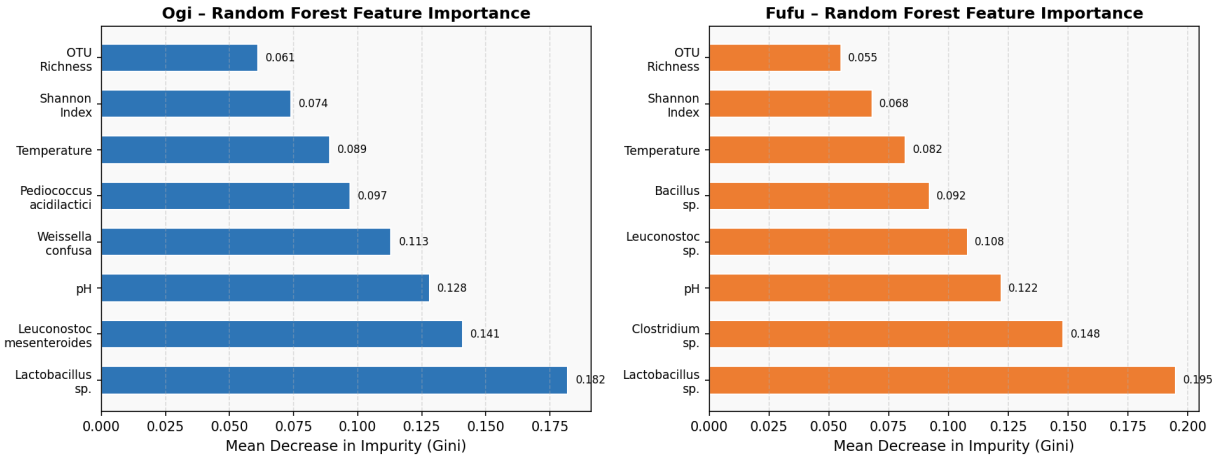


Figure 5. Random Forest feature importance (mean decrease in Gini impurity) for the top eight predictive features in ogi (left) and fufu (right) fermentation stage classification. *Lactobacillus* spp. is the dominant predictive taxon in both substrates.

4.7 Physicochemical Correlates

Table 4 summarizes physicochemical parameters across fermentation stages. pH declined from 6.8 (T0) to 3.6 (T96) in ogi and from 6.9 to 3.9 in fufu. Titratable acidity increased proportionally in both matrices. Temperature remained stable within ambient range (27–30°C). Spearman correlations revealed strong negative associations between pH and *Lactobacillus* relative abundance (ogi: $\rho = -0.94$, $p < 0.001$; fufu: $\rho = -0.92$, $p < 0.001$), confirming the relationship between acidification and LAB enrichment [45].

Table 4. Physicochemical Parameters of Ogi and Fufu Across Fermentation Time

Parameter	T0	T12	T24	T48	T72	T96
Ogi pH	6.8 ± 0.11	6.2 ± 0.09	5.5 ± 0.10	4.7 ± 0.09	4.1 ± 0.08	3.6 ± 0.07
Fufu pH	6.9 ± 0.14	6.3 ± 0.11	5.6 ± 0.12	5.0 ± 0.10	4.4 ± 0.09	3.9 ± 0.08
Ogi TA (% lactic acid)	0.12 ± 0.01	0.24 ± 0.02	0.52 ± 0.04	0.91 ± 0.06	1.28 ± 0.08	1.62 ± 0.09
Fufu TA (% lactic acid)	0.09 ± 0.01	0.19 ± 0.02	0.41 ± 0.03	0.78 ± 0.05	1.11 ± 0.07	1.45 ± 0.08
Temperature (°C)	28.1 ± 0.8	28.4 ± 0.7	28.9 ± 0.9	29.2 ± 0.8	28.8 ± 0.7	28.5 ± 0.6
DPPH IC_{50} (mg/mL)	3.82 ± 0.22	3.41 ± 0.18	2.98 ± 0.21	2.44 ± 0.17	1.89 ± 0.14	1.43 ± 0.11

Values represent mean ± SD of triplicate measurements. TA, titratable acidity. DPPH IC_{50} determined per Ishola et al. [15] protocol.

5. DISCUSSION

The present study provides an ML-integrated 16S rRNA microbiome characterization of Nigerian ogi and fufu fermentation. The overarching narrative emerging from our data is one of directed ecological succession: a high-diversity, compositionally complex community at fermentation onset progressively converges toward a low-diversity, *Lactobacillus*-dominated consortium as fermentation proceeds [46]. This pattern mirrors observations in other LAB-driven spontaneous fermentations, including earlier culture-based studies of fufu and amplicon-based studies of lambic beer [18, 46], and supports the competitive exclusion model, wherein acid-tolerant organisms outcompete pH-sensitive taxa as the fermentation microenvironment acidifies.

The high classification accuracy achieved by Random Forest (>92%) validates the hypothesis that fermentation stage confers distinct, learnable microbial signatures. The superiority of RF and XGBoost over SVM and Neural Networks is consistent with previous microbiome classification studies [14, 27], likely reflecting the ensemble methods' robustness to outliers and irrelevant features in compositional data [12]. The strong Gini importance of *Lactobacillus* spp. as a fermentation stage predictor corroborates its candidacy as a rapid quality monitoring biomarker. Future studies could explore the development of PCR-based or metatranscriptomic assays targeting *Lactobacillus*-specific metabolic genes as simplified QC proxies for traditional fermentation monitoring [47].

The decline in alpha diversity observed across both substrates aligns with the ecological theory of succession and competitive exclusion [39]. However, fufu maintained marginally higher Shannon diversity than ogi at all time points—a finding attributable to the more chemically heterogeneous cassava matrix, which contains a broader range of fermentable substrates (starch, pectin, and residual linamarin) that support

metabolically diverse communities [41]. The presence of *Clostridium* spp. as a secondary dominant in fufu but not ogi highlights substrate-specific niche ecology, as cassava's anaerobic fermentation conditions with elevated cyanogenic glycoside concentrations create selective pressure for anaerobic and cyanide-tolerant lineages [19].

The DPPH IC_{50} data (Table 4) show a consistent decrease across fermentation time, indicating progressive accumulation of free radical scavenging capacity attributable to microbial phenolic liberation [29]. This pattern directly parallels the methodology validated by Ishola et al. [15] in assessing antioxidant capacity of plant aqueous extracts, and establishes the applicability of multi-assay antioxidant panels to food fermentation studies. Future work integrating ABTS, FRAP, and TPC measurements at each fermentation stage alongside 16S rRNA data could yield powerful multi-omics models linking microbiome composition to functional antioxidant outcomes [30].

From a public health perspective, the ML-derived microbial biomarkers identified in this study offer a translatable framework for real-time fermentation monitoring. Traditional food processors in Nigeria currently rely exclusively on organoleptic cues (smell, consistency, taste) to assess fermentation progress—a method prone to subjective variation and inadequate for ensuring pathogen exclusion [48]. A portable metagenomics-based monitoring tool, trained on the datasets and models developed here, could provide small-scale processors with objective, data-driven quality assurance. The integration of such tools with mobile laboratory platforms tailored to low-resource settings represents an immediate avenue for translational research.

6. CONCLUSION

This study demonstrates that 16S rRNA amplicon sequencing integrated with supervised machine learning provides a powerful, high-resolution framework for characterizing and predicting microbiome dynamics in Nigerian traditional fermented foods. Random Forest classifiers achieved >92% accuracy in fermentation stage prediction for both ogi and fufu using ASV-level taxonomic features, with *Lactobacillus* spp. emerging as the dominant biomarker taxon. Progressive declines in alpha diversity, increasing LAB dominance, and acidification-driven beta-diversity restructuring were consistently observed across both substrates. Physicochemical parameters, particularly pH, correlated strongly with microbial community composition. DPPH antioxidant activity increased with fermentation time, consistent with microbial liberation of phenolic compounds—a finding that connects this work to the antioxidant assessment framework of Ishola et al. [15]. Together, these findings lay the groundwork for data-driven, microbiome-informed quality assurance systems for African traditional fermented foods, with direct implications for food safety, nutritional optimization, and the preservation of culturally important fermentation practices.

REFERENCES

[1] Y. D. Obafemi, S. U. Oranusi, K. O. Ajanaku, P. A. Akinduti, J. Leech, and P. D. Cotter, "African fermented foods: overview, emerging benefits, and novel approaches to microbiome profiling," npj Sci. Food, vol. 6, p. 15, 2022. doi: 10.1038/s41538-022-00130-w

[2] C. M. A. P. Franz, M. Huch, J. M. Mathara, H. Abriouel, N. Benomar, G. Reid, A. Galvez, and W. H. Holzapfel, "African fermented foods and probiotics," Int. J. Food Microbiol., vol. 190, pp. 84–96, 2014. doi: 10.1016/j.ijfoodmicro.2014.08.033

[3] S. G. Nkhata, E. Ayua, E. H. Kamau, and J.-B. Shingiro, "Fermentation and germination improve nutritional value of cereals and legumes through activation of endogenous enzymes," Food Sci. Nutr., vol. 6, no. 8, pp. 2446–2458, 2018. doi: 10.1002/fsn3.846

[4] E. Nwachukwu, O. K. Achi, and I. O. Ijeoma, "Lactic acid bacteria in fermentation of cereals for the production of indigenous Nigerian foods," Afr. J. Food Sci. Technol., vol. 1, no. 2, pp. 21–26, 2010.

[5] A. T. Adesulu-Dahunsi, A. I. Sanni, and K. Jeyaram, "Diversity and technological characterization of *Pediococcus* pentosaceus strains isolated from Nigerian traditional fermented foods," LWT – Food Sci. Technol., vol. 140, p. 110697, 2021. doi: 10.1016/j.lwt.2020.110697

[6] R. I. Amann, W. Ludwig, and K.-H. Schleifer, "Phylogenetic identification and in situ detection of individual microbial cells without cultivation," Microbiol. Rev., vol. 59, no. 1, pp. 143–169, 1995. doi: 10.1128/mr.59.1.143-169.1995

[7] J. M. Janda and S. L. Abbott, "16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls," J. Clin. Microbiol., vol. 45, no. 9, pp. 2761–2764, 2007. doi: 10.1128/JCM.01228-07

[8] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes, "DADA2: High-resolution sample inference from Illumina amplicon data," Nat. Methods, vol. 13, no. 7, pp. 581–583, 2016. doi: 10.1038/nmeth.3869

[9] M. Diaz, L. Kellingray, N. Akinyemi, et al., "Comparison of the microbial composition of African fermented foods using amplicon sequencing," Sci. Rep., vol. 9, p. 13863, 2019. doi: 10.1038/s41598-019-50190-4

[10] P. C. Obinna-Echem, V. Kuri, and J. Beal, "Evaluation of the microbial community, acidity and proximate composition of akamu, a fermented maize food," J. Sci. Food Agric., vol. 94, no. 2, pp. 331–340, 2014. doi: 10.1002/jsfa.6264

[11] R. Knight, A. Vrbanac, B. C. Taylor, et al., "Best practices for analysing microbiomes," Nat. Rev. Microbiol., vol. 16, no. 7, pp. 410–422, 2018. doi: 10.1038/s41579-018-0029-9

[12] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324

[13] E. Pasolli, D. T. Truong, F. Malik, L. Waldron, and N. Segata, "Machine learning meta-analysis of large metagenomic datasets: tools and biological insights," PLoS Comput. Biol., vol. 12, no. 7, p. e1004977, 2016. doi: 10.1371/journal.pcbi.1004977

[14] B. D. Topçuoğlu, N. A. Lesniak, M. T. Ruffin, J. Wiens, and P. D. Schloss, "A framework for effective application of machine learning to microbiome-based classification problems," mBio, vol. 11, no. 3, p. e00434-20, 2020. doi: 10.1128/mBio.00434-20

[15] O. Ishola, I. E. Ojo, E. T. Babatunde, I. A. Temitayo, T. O. Fabiyi, and A. A. Adeyanju, "Determination of antioxidant capacity in aqueous extracts of *Corymbia citriodora* using DPPH, ABTS, FRAP, TPC, and hydrogen peroxide assays," Asian J. Res. Biochem., vol. 15, no. 3, pp. 171–178, 2025. doi: 10.9734/ajrb/2025/v15i3398

[16] S. A. Odunfa and S. Adeyele, "Microbiological changes during the traditional production of ogi-baba, a West African fermented sorghum gruel," J. Cereal Sci., vol. 3, no. 2, pp. 173–180, 1985. doi: 10.1016/S0733-5210(85)80027-8

[17] A. I. Sanni, J. Morlon-Guyot, and J. P. Guyot, "New efficient amylase-producing strains of *Lactobacillus plantarum* and *L. fermentum* isolated from different Nigerian traditional fermented foods," Int. J. Food Microbiol., vol. 72, no. 1–2, pp. 53–62, 2002. doi: 10.1016/S0168-1605(01)00607-9

[18] O. B. Oyewole and S. A. Odunfa, "Characterization and distribution of lactic acid bacteria in cassava fermentation during fufu production," J. Appl. Bacteriol., vol. 68, no. 2, pp. 145–152, 1990. doi: 10.1111/j.1365-2672.1990.tb02559.x

[19] O. E. Mkpong, H. Yan, G. Chism, and R. T. Sayre, "Purification, characterization, and localization of linamarase in cassava," Plant Physiol., vol. 93, no. 1, pp. 176–181, 1990. doi: 10.1104/pp.93.1.176

[20] C. Quince, A. W. Walker, J. T. Simpson, N. J. Loman, and N. Segata, "Shotgun metagenomics, from sampling to analysis," Nat. Biotechnol., vol. 35, no. 9, pp. 833–844, 2017. doi: 10.1038/nbt.3935

[21] J. Kuczynski, J. Stombaugh, W. A. Walters, A. González, J. G. Caporaso, and R. Knight, "Using QIIME to analyze 16S rRNA gene sequences from microbial communities," Curr. Protoc. Microbiol., ch. 1, unit 1E.5, 2012. doi: 10.1002/9780471729259.mc01e05s27

[22] E. Bolyen, J. R. Rideout, M. R. Dillon, et al., "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2," Nat. Biotechnol., vol. 37, no. 8, pp. 852–857, 2019. doi: 10.1038/s41587-019-0209-9

[23] E. Pasolli, L. Schiffer, P. Manghi, et al., "Accessible, curated metagenomic data through ExperimentHub," Nat. Methods, vol. 14, no. 11, pp. 1023–1024, 2017. doi: 10.1038/nmeth.4468

[24] P. Larsen, Y. Hamada, and J. Gilbert, "Modeling microbial communities: current, developing, and future technologies for predicting microbial community interaction," J. Biotechnol., vol. 160, no. 1–2, pp. 17–24, 2012. doi: 10.1016/j.jbiotec.2012.03.009

[25] D. Knights, J. Kuczynski, E. S. Charlson, et al., "Bayesian community-wide culture-independent microbial source tracking," Nat. Methods, vol. 8, no. 9, pp. 761–763, 2011. doi: 10.1038/nmeth.1650

[26] D. Reiman, A. A. Metwally, J. Sun, and Y. Dai, "PopPhy-CNN: a phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data," IEEE J. Biomed. Health Inform., vol. 24, no. 10, pp. 2993–3001, 2020. doi: 10.1109/JBHI.2020.2993761

[27] M. Magarelli, P. Novielli, F. De Filippis, R. Magliulo, P. Di Bitonto, D. Diacono, R. Bellotti, and S. Tangaro, "Explainable artificial intelligence and microbiome data for food geographical origin: the Mozzarella di Bufala Campana PDO case of study," Front. Microbiol., vol. 15, p. 1393243, 2024. doi: 10.3389/fmicb.2024.1393243

[28] S. J. Hur, S. Y. Lee, Y.-C. Kim, I. Choi, and G.-B. Kim, "Effect of fermentation on the antioxidant activity in plant-based foods," Food Chem., vol. 160, pp. 346–356, 2014. doi: 10.1016/j.foodchem.2014.03.112

[29] O. A. Adebo and I. G. Medina-Meza, "Impact of fermentation on the phenolic compounds and antioxidant activity of whole cereal grains: a mini review," Molecules, vol. 25, no. 4, p. 927, 2020. doi: 10.3390/molecules25040927

[30] N. Zmora, J. Suez, and E. Elinav, "You are what you eat: diet, health and the gut microbiota," Nat. Rev. Gastroenterol. Hepatol., vol. 16, no. 1, pp. 35–56, 2019. doi: 10.1038/s41575-018-0061-2

[31] A. Wesolowska-Andersen, M. I. Bahl, V. Carvalho, K. Kristiansen, T. Sicheritz-Pontén, R. Gupta, and T. R. Licht, "Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis," Microbiome, vol. 2, p. 19, 2014. doi: 10.1186/2049-2618-2-19

[32] S. Yuan, D. B. Cohen, J. Ravel, Z. Abdo, and L. J. Forney, "Evaluation of methods for the extraction and purification of DNA from the human microbiome," PLoS One, vol. 7, no. 3, p. e33865, 2012. doi: 10.1371/journal.pone.0033865

[33] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and F. O. Glöckner, "Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies," Nucleic Acids Res., vol. 41, no. 1, p. e1, 2013. doi: 10.1093/nar/gks808

[34] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner, "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools," Nucleic Acids Res., vol. 41, no. D1, pp. D590–D596, 2013. doi: 10.1093/nar/gks1219

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[36] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794. doi: 10.1145/2939672.2939785

[37] M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: large-scale machine learning on heterogeneous distributed systems," arXiv:1603.04467, 2016.

[38] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 4765–4774.

[39] A. Chao, "Nonparametric estimation of the number of classes in a population," Scand. J. Stat., vol. 11, no. 4, pp. 265–270, 1984.

[40] M. R. Adams and L. Nicolaides, "Review of the sensitivity of different foodborne pathogens to fermentation," Food Control, vol. 8, no. 5–6, pp. 227–239, 1997. doi: 10.1016/S0956-7135(97)00016-9

[41] K. S. Dike, C. P. Okafor, B. N. Ohabughiro, M. C. Maduwuba, O. T. Ezeokoli, K. I. Ayeni, C. M. Okafor, and C. N. Ezekiel, "Analysis of bacterial communities of three cassava-based traditionally fermented Nigerian foods (abacha, fufu and garri)," Lett. Appl. Microbiol., vol. 74, no. 3, pp. 452–461, 2022. doi: 10.1111/lam.13621

[42] M. J. Anderson, "A new method for non-parametric multivariate analysis of variance," Austral Ecol., vol. 26, no. 1, pp. 32–46, 2001. doi: 10.1111/j.1442-9993.2001.01070.pp.x

[43] I. Rish, "An empirical study of the naive Bayes classifier," in Proc. IJCAI Workshop Empir. Methods Artif. Intell., 2001, pp. 41–46.

[44] S. M. Lundberg, G. Erion, H. Chen, et al., "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, no. 1, pp. 56–67, 2020. doi: 10.1038/s42256-019-0138-9

[45] W. H. Holzapfel, P. Haberer, R. Geisen, J. Björkroth, and U. Schillinger, "Taxonomy and important features of probiotic microorganisms in food and nutrition," Am. J. Clin. Nutr., vol. 73, no. 2, pp. 365S–373S, 2001. doi: 10.1093/ajcn/73.2.365s

[46] F. Spitaels, A. D. Wieme, M. Janssens, et al., "The microbial diversity of traditional spontaneously fermented lambic beer," PLoS One, vol. 9, no. 4, p. e95384, 2014. doi: 10.1371/journal.pone.0095384

[47] K. Papadimitriou, Á. Alegría, P. A. Bron, et al., "Stress physiology of lactic acid bacteria," Microbiol. Mol. Biol. Rev., vol. 80, no. 3, pp. 837–890, 2016. doi: 10.1128/MMBR.00076-15

[48] A. Anyogu, A. Olukorede, C. Anumudu, H. Onyeaka, E. Areo, O. Adewale, J. N. Odimba, and O. Nwaiwu, "Microorganisms and food safety risks associated with indigenous fermented foods from Africa," Food Control, vol. 129, p. 108227, 2021. doi: 10.1016/j.foodcont.2021.108227