Original Research Article

Enhancing Microorganism Classification Using Multinomial Logit Regression, k-Nearest Neighbor, and a Hybrid Approach

Alicia J. Lam

Temple City High School, 9501 Lemon Ave, Temple City, CA 91780, United States

ABSTRACT

The aim of this study is to address the current challenges in microorganism identification and classification–particularly those that impede timely and accurate medical diagnoses and treatments for patients. Recognizing the constraints of traditional bacterial classification processes, this study explores the potential of machine learning in streamlining microorganism identification using their morphological features. For this research, three modeling techniques were tested: multinomial logistic regression (MLR), k-nearest neighbor (k-NN), and a novel hybrid model integrating the two. Using a dataset sourced from Kaggle–a Google website that serves as a platform for members of the scientific community to publish their datasets–and individually benchmarked in their abilities to accurately distinguish between ten distinct microorganism species *(Spirogyra, Volvox, Pithophora, Ulothrix, Diatom, Fungi, Yeast, Rhizopus, Penicillium, Aspergillus* sp., and *Protozoa)* based on twenty-four numerical features detailing their geometry and structure. The experiment's findings revealed that while the k-NN model outperformed the multinomial logistic model, the hybrid approach yielded the highest degree of accuracy in its classification of the ten microorganism species. In comparison to conventional cultivation techniques employed in clinical microbiology, machine learning forgoes the lengthy processing time associated with it, as it has the capability to accurately identify pathogens based on existing data. As a result, patients–especially those in urgent cases–are able to receive rapid diagnosis and necessary treatment before the bacterial culture is fully grown.

Keywords: Microorganism classification; Machine learning; Morphological features; Multinomial logistic regression; k-Nearest Neighbor; Hybrid model; Rapid diagnosis

INTRODUCTION

In 2023, of the 34.4 billion defined daily doses (DDDs) of antibiotics distributed per 1,000 people annually, an estimated 30-50% of it was inappropriately prescribed (1, 2). This large proportion of antibiotics misuse can be largely attributed to the delays in pathogen identification processes, which often force healthcare providers to prescribe broad-spectrum antibiotics as a first-line of treatment or as a precautionary measure. Not only can limitations in rapidly and accurately determining when antibiotics are necessary contribute to the growing threat of antibiotic resistance, it can also be life-threatening, leading to fatal bacterial infections such as sepsis, pneumonia, and anthrax. Traditional bacterial cultivation and identification techniques implemented in healthcare settings today are lengthy and often impractical for fast-progressing infections

Corresponding author: Alicia J. Lam, E-mail: lamalicia336@gmail.com.

Accepted November 11, 2025
https://doi.org/10.70251/HYJR2348.36485493

where immediate diagnosis and treatment are critical. The current "gold standard" process applied in hospitals and healthcare clinics for classifying infectious pathogens involves first cultivating bacteria from a patient's samples, then isolating individual microbes to grow pure colonies, before identifying these pure colonies using biochemical tests, microscopy, or other molecular techniques (3). After they have been successfully identified, the bacteria cultures are exposed to various antibiotics to determine which drug will be most effective in treating the infection. However, both the identification process and antibiotic susceptibility testing require substantial amounts of bacterial growth from cultivation. While the incubation period for most pathogens ranges from 24 to 48 hours, some bacterial species, such as the *Bartonella* species, can require more than 45 days to incubate (4). This extensive processing time is infeasible for life-threatening infections such as sepsis–a condition triggered by bacterial infections that often leads to organ failure and even death. According to the Sepsis Alliance Institute (5), there is a 4-9% increase in morbidity and mortality rates with each hour treatment is delayed. However, with rapid identification, diagnosis and treatment, as many as 80% of septic shock patients can be saved.

These issues emphasize the urgent need for a more efficient microorganism identification process that can provide results before culture confirmation. Despite the advancement of technology and its widespread incorporation in various scientific and health-related fields, there has been a lack of research exploring the use of machine learning in making the microorganism classification process more efficient. While a few literature reviews discussing the application of machine learning in microorganism classification were published in the recent decade, such as one by Kotwal *et al.* (6) and another by Rani *et al.* (7), overall contributions to and experimental studies in this field have been limited.

To bridge the significant gap in research, this study aims to address and explore the potential of well-established machine learning methods–multinomial logistic regression and k-nearest neighbor–in classifying microorganism species. Additionally, a hybrid model that embeds the output of a k-NN classifier directly into the multinomial logistic regression model as an additional predictor variable was created to capitalize on the two models' distinct advantages while also mitigating their weaknesses. The three machine learning models were tested using an extensive dataset sourced from Kaggle, which detailed 24 numeric features describing the geometry and structure of 10 distinct microorganism species.

METHODS AND MATERIALS

The research organizes and analyzes the dataset through the employment of three machine learning approaches: multinomial logistic regression, k-nearest neighbor, and a hybrid approach integrating the two aforementioned models. All machine learning algorithms were run in Python 3.9 using the scikit-learn 1.3.0 library to build and test the models. Pandas was used to refine and organize the data before model training. The models' performance metrics were exported and visualized using the Chart feature in Google Sheets. The dataset was divided into training and testing subsets using an 80:20 ratio. Before training the models, the dataset was normalized using z-score standardization to ensure all numeric descriptors were on a comparable scale.

Data Description

The Microbes Dataset used for this research includes 30,526 independent observations. Each observation corresponds to a microorganism from one of ten species examined that was analyzed and identified through image analysis. From those images, numerical descriptors detailing the specific geometry and structural features of ten types of microorganisms were recorded. The numerical descriptors include measurements for twenty-six features: area, perimeter, convex area, solidity, eccentricity, equivalent diameter, centroid coordinates, radii, etc. The dataset was selected for its extensive array of morphological and quantitative descriptors, which make it well suited for training machine learning models to distinguish microbial species and identify pathogenic microorganisms. The dataset, published in CSV format, is available to the public on Kaggle's platform. It has a monthly average of 21,100 views and 2,600 downloads, indicating its current credibility and relevance to the scientific community. However, since the experiment is only using the dataset to train and compare the performance of machine learning models, the data's absolute accuracy and representativeness are not completely necessary to achieve the research's objectives. Instead, the most important aspect is that the dataset has a large sample size and captures the diverse morphological features of different microorganism species. Detailed information on the data is described in Table 1.

Table 1. List of Variables and the Associated Definition and Descriptive Statistics

Variables	Definition	Descriptive Statistics
Solidity	How compact the microorganism is; helps distinguish between smooth/ compact organisms and irregular/branched ones $Solidity = \frac{Area}{Convex\ Hull\ Area}$	Mean: 9.677744 StDev: 4.063437 Range: 23.0
Eccentricity	How elongated the microorganism is (0 = round/circle, close to 1 = very elongated)	Mean: 19.466921 StDev: 3.479828 Range: 23.0
EquivDiameter	Diameter of a circle with the same area as the organism $EquivDiameter = \sqrt{\frac{4 \bullet Area}{\pi}}$	Mean: 3.63348 StDev: 2.210851 Range: 23.0
Extrema	The extreme points of the microorganism (top, bottom, left, right, etc.)	Mean: 11.871832 StDev: 6.045135 Range: 23.0
Filled_Area	The area after filling in any internal holes in the microorganism's body	Mean: 0.420022 StDev: 0.875091 Range: 23.0
Extent	Ratio of the object's area to the bounding box area $Extent = \frac{Area}{Bounding\ Box\ Area}$	Mean: 5.840625 StDev: 3.250999 Range: 23.0
Orientation	The angle (in degrees) between the x-axis and the major axis of the fitted ellipse to the object	Mean: 11.751004 StDev: 6.57319 Range: 23.0
Euler_Number	A topological descriptor referring to the connectivity and holes inside the microorganism's shape *Euler Number = Number of connected objects - Number of holes*	Mean: 22.380901 StDev: 0.962906 Range: 23.0
1. BoundingBox1 **2. BoundingBox2** **3. BoundingBox3** **4. BoundingBox4** **1. ConvexHull1** **2. ConvexHull2**	Four numerical descriptors representing position, size, and shape of the bounding box	1. Mean: 10.919027 StDev: 6.093280 Range: 23.0 2. Mean: 10.399429 StDev: 5.797144 Range: 23.0 3. Mean: 2.085481 StDev: 2.166312 Range: 23.0 4. Mean: 2.640499 StDev: 2.488448 Range: 23.0 1. Mean: 11.113760 StDev: 6.033357 Range: 23.0 2. Mean: 11.113760 StDev: 6.033357 Range: 23.0

Continued Table 1. List of Variables and the Associated Definition and Descriptive Statistics

Variables	Definition	Descriptive Statistics
3. ConvexHull3 **4. ConvexHull4**	Four numerical descriptors representing position, size, and shape of the bounding box	3. Mean: 11.046482 StDev: 6.089508 Range: 23.0 4. Mean: 11.021988 StDev: 6.089467 Range: 23.0
1. Centroid1 **2. Centroid2**		1. Mean: 11.752783 StDev: 6.029756 Range: 23.0 2. Mean: 11.554286 StDev: 5.700637 Range: 23.0
1. Major_Axis_Length **2. Minor_Axis_Length**	Describes the microorganism's longest and shortest dimensions	1. Mean: 1.605159 StDev: 1.662437 Minimum: 0.0 Maximum: 23.0 2. Mean: 1.014179 StDev: 1.561427 Minimum: 0.0 Maximum: 23.0
Perimeter	The total length of the boundary of the microorganism's shape	Mean: 0.829416 StDev: 1.152165 Minimum: 0.0 Maximum: 23.0
Convex_Area	The area of the convex hull surrounding the microorganism	Mean: 0.254596 StDev: 0.971035 Minimum: 0.0 Maximum: 23.0
Area	The number of pixels inside the microorganism's boundary	Mean: 0.802780 StDev: 1.170430 Minimum: 0.0 Maximum: 23.0
Radii	Distances measured from the centroid of the microorganism to its boundary	Mean: 5.214598 StDev: 2.805199 Minimum: 0.0 Maximum: 23.0
1. Ulothrix **2. Volvox** **3. Protozoa** **4. Aspergillus sp** **5. Yeast** **6. Rhizopus** **7. Diatom** **8. Pithophora** **9. Penicillium** **10. Spirogyra**	Type of microorganism	1. Frequency: 7420 2. Frequency: 4320 3. Frequency: 3888 4. Frequency: 3888 5. Frequency: 3600 6. Frequency: 2552 7. Frequency: 1818 8. Frequency: 1350 9. Frequency: 1080 10. Frequency: 611

Multinomial Logit Regression

Multinomial logistic regression (MLR) is a modeling technique capable of predicting an observation's membership across three or more outcome categories. It works by calculating the log-odds relative to a reference group to estimate the probability that an observation belongs to each class. The model assumes a linear relationship between the predictor variables and the outcome categories (8, 9). In this study, multinomial logistic regression is evaluated on how accurately it is able to predict the respective categories or species of microorganisms based on the given numerical descriptors of their morphological features.

The probability (P_{mn}) that an observation (m) belongs to a specific species (n) out of M possible categories can be calculated using the following equation:

$$P_{mn} = \frac{e^{X'_m \beta_n}}{\sum_{k=1}^{M} e^{X'_m \beta_n}} \quad (1)$$

Where P_{mn} is the probability that observation m belongs to category nn X_m represents the vector of explanatory (independent) variables for observation m, β_n is the coefficient vector for category n, M is the total number of categories; and e symbolizes the exponential function.

k-Nearest Neighbor

The k-nearest neighbor algorithm (k-NN) is a non-parametric classification method in which the category of a new sample or observation is computed using the majority label among its k neighbors in the feature space (10). k-NN assumes that observations that are close in their feature values likely belong to the same category. The Euclidean distance between data points is calculated to determine which samples are most similar to each other in their feature values. In this study, the k-NN algorithm is tested to evaluate how accurately it can identify a microorganism species by comparing each sample's morphological and numerical features with those of its nearest neighbors. The k-Nearest Neighbor classifier was set to 5, as it consistently yielded the most accurate results in comparison to k values ranging from 3 to 9, which were also tested. The predicted class (ŷm) of observation (m) based on the majority of its *k* nearest neighbors can be calculated using the following expression.

$$\hat{y}_m = \arg\max_{n \in N} \sum_{i=1}^{k} I(y_i = n) \quad (2)$$

Where ŷm is the predicted class of observation m, n is the possible category of species that sample could belong to, N is the total number of possible microorganism species, k is the number of nearest neighbors considered, yi is the true class of the i^{th} nearest neighbor, and arg max is the function that selects the class n with the highest total count among the k nearest neighbors. For this experiment, k=5.

Hybrid Approach

Although multinomial logistic regression and k-nearest neighbor models can perform effectively well individually, both can be improved to achieve greater accuracy in microorganism classification. To elaborate, the assumption of a linear relationship between predictor variables and outcome categories in MLR may potentially hinder its ability to capture complex, nonlinear patterns sometimes present in microorganisms' morphological and numerical characteristics.

Additionally, while the non-parametric nature of the k-NN algorithm enables it to model and capture nonlinear relationships, its predictive power is highly sensitive to the quality and content of existing data. Since k-NN determines the classification for new samples according to their resemblance to known samples, its accuracy is compromised when working with noisy datasets and underrepresented microorganism classes. Furthermore, since k-NN does not produce explicit coefficients indicating which features had the greatest impact on microorganism classification, it lacks interpretability.

To mitigate these limitations, the study created a hybrid modeling approach that combines the two methods–the output from the k-NN algorithm is incorporated into the MLR model as an additional predictor variable–allowing the hybrid model to exploit the flexibility of k-NN and the interpretability of MLR while avoiding their shortcomings.

RESULTS

The results of the study present the outcomes of using three machine-learning-based approaches–multinomial logistic regression, k-nearest neighbor, and a hybrid model combining both methods–to classify ten microorganism species based on twenty-six types of quantitative descriptors of their morphology. Four classification metrics were used to assess the performance of all three models: accuracy, precision, recall, and F1-score. Accuracy is the percentage of correctly classified samples compared to the total; precision and recall reflect how well the model is able to identify true positives compared to false detections and missed cases. The F1-score is the trade-off between precision and recall. While not used to directly evaluate the machine learning models, important metrics included in the following data tables include support–the number of samples labeled as a specific microorganism species; macro average–the mean of the precision, recall, or F1-score across all classes independent of the number of support; and weighted average–the mean of the precision, recall, or F1-score across all classes dependent on the number of support. The models' performance was examined and compared, and the findings are summarized below.

Multinomial Logistic Regression (MLR)

Table 2. Classification Metrics for Multinomial Logistic Regression Across Ten Microorganism Species

Microorganism	Precision	Recall	F1-Score	Support
1. *Aspergillus*	0.39	0.40	0.39	778
2. *Diatom*	0.38	0.24	0.29	364
3. *Penicillium*	0.00	0.00	0.00	216
4. *Pithophora*	0.63	0.56	0.59	270
5. *Protozoa*	0.46	0.43	0.44	778
6. *Rhizopus*	0.62	0.73	0.67	510
7. *Spirogyra*	0.00	0.00	0.00	122
8. *Ulothrix*	0.47	0.63	0.54	1484
9. *Volvox*	0.50	0.51	0.51	864
10. *Yeast*	0.55	0.48	0.52	720
Accuracy			0.49	6106
Macro avg.	0.40	0.40	0.39	6106
Weighted avg.	0.46	0.49	0.47	6106

As shown in Table 2, the multinomial logistic regression (MLR) model achieved an overall accuracy of 0.49 on a scale ranging from 0 to 1, where a value closer to 1 indicates a more precise classification. Based on this scale, MLR's accuracy score reflects a moderate ability to classify microorganisms based on their morphological features. As shown in Figure 1, among the ten microorganism species, *Rhizopus* and *Pithophora* had the highest F1-scores of 0.67 and 0.59, respectively, suggesting that their structural characteristics were likely more distinct and therefore easier for the model to recognize and distinguish. Conversely, *Penicillium* and *Spirogyra* had the lowest F1-scores of 0.00, indicating frequent false detections and missed positives due to increased overlap in their morphological features with other microorganism categories. MLR's weighted average F1-score of 0.47–which is also ranked on a scale from 0 to 1, where values closer to 1 represent optimal classification–indicates an uneven performance in accurately classifying the microorganism types. As shown in Figure 1, while MLR performed effectively in identifying some clear differences between the microorganisms, it had difficulty classifying those with similar or overlapping morphological features.

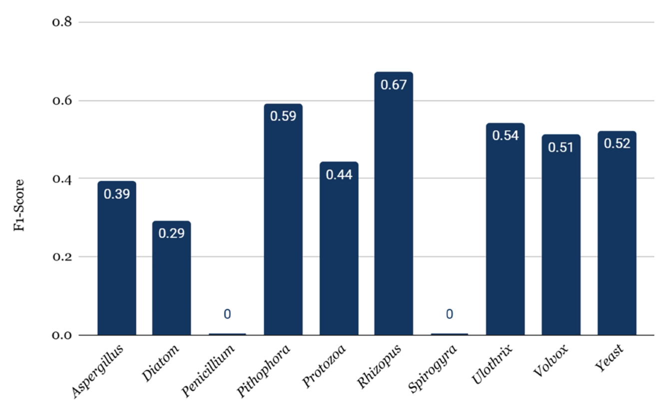

Figure 1. Multinomial Logit Model F1-score Distribution Across Ten Microorganism Categories.

The MLR model assumes a linear relationship between predictor variables and output classes. However, in reality, the relationship between microorganism morphology and their species identity is more complex and non-linear–their morphological features do not separate into distinct and independent groups. Hence, while MLR has the capability to accurately identify microorganism species with distinct geometric features–such as *Rhizopus*, its ability to correctly classify species with overlapping morphological features–such as *Spirogyra* and *Penicillium*–was limited. Conversely, k-NN and the hybrid model are able to consider nonlinear patterns between microorganisms' morphological features and their species identity, allowing them to be more accurate in their species classification.

k-Nearest Neighbor (k-NN)

The k-nearest neighbor model classified organisms by comparing each sample's morphological and numerical features to those of its five closest neighbors (k=5) in the dataset. As shown in Table 3, the model was highly successful, producing an overall accuracy of 0.84, with one class, *Volvox*, achieving the maximum recall score of 1.00, meaning it was correctly identified by k-NN each time. The k-NN model's overall accuracy of 0.84 indicates a strong ability to accurately classify the majority of the 30,526 samples. As shown in Figure 2, microorganism species *Volvox* and *Yeast* obtained the highest F1-scores of 0.96 and 0.95, respectively, suggesting that their structural characteristics are highly recognizable by the model. Additionally, the microorganism species *Pithophora*, which MLR struggled to classify (Table 2), presented a significantly higher F1-score of 0.68 using k-NN. k-NN's lowest F1-score was 0.03, compared to MLR's 0.00. k-NN's overall higher macro average (0.72), weighted average (0.83), and accuracy score (0.84) classify k-NN as a highly accurate modeling technique for bacterial identification.

The k-NN model's classification of microorganism species based on their morphological characteristics was more accurate compared to MLR's performance, as it identifies species based on similar patterns between local samples. Its use of similarity-based comparisons means k-NN has the ability to distinguish species with similar, overlapping, or irregular morphological features–something MLR could not achieve.

Table 3. Classification Report for k-Nearest Neighbor Across Ten Microorganism Species

Microorganism	Precision	Recall	F1-Score	Support
1. Aspergillus	0.74	0.81	0.77	778
2. Diatom	0.69	0.76	0.72	364
3. Penicillium	0.86	0.41	0.55	216
4. Pithophora	0.81	0.58	0.68	270
5. Protozoa	0.83	0.86	0.84	778
6. Rhizopus	0.78	0.83	0.80	510
7. Spirogyra	0.13	0.02	0.03	122
8. Ulothrix	0.87	0.91	0.89	1484
9. Volvox	0.93	1.00	0.96	864
10. Yeast	0.95	0.94	0.95	720
Accuracy			0.84	6106
Macro avg.	0.76	0.71	0.72	6106
Weighted avg.	0.83	0.84	0.83	6106

Hybrid Modeling Approach (k-NN & MLR)

The hybrid model combines k-NN's local similarity-based predictions with MLR's broader statistical classification approach. As shown in Table 4, the model achieved an overall accuracy of 0.94, which is significantly higher than both MLR's 0.49 (Table 2) and k-NN's 0.84 (Table 3) scores, demonstrating that combining both modeling techniques yields the most precise classifications. As illustrated in Figure 3, eight out of ten microorganism classes achieved an F1-score of 0.90 or higher, with three classifications obtaining a maximum recall score of 1.00 (Table 4). The classification of microorganisms that both MLR and k-NN struggled to distinguish showed substantial

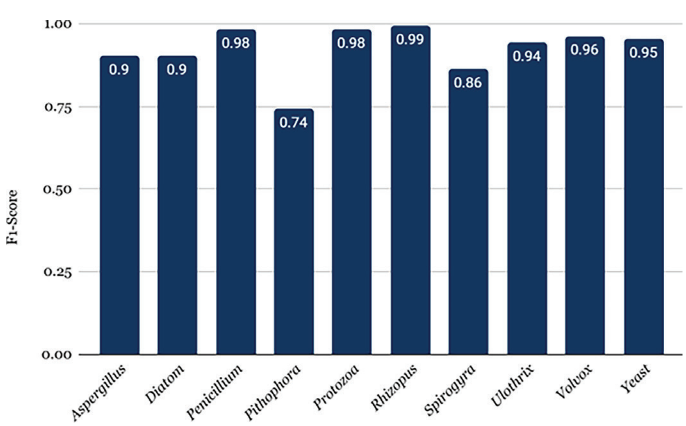

Figure 2. k-Nearest Neighbor F1-score Distribution Across Ten Microorganism Categories.

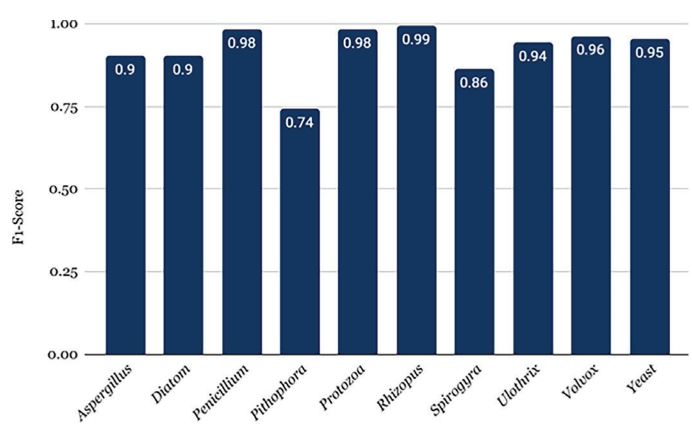

Figure 3. Hybrid Model F1-score Distribution Across Ten Microorganism Categories.

improvement using the hybrid model–*Spirogyra*, for example, reached an F1-score of 0.86. As shown in Figure 4, which compares the overall accuracy and weighted average F1-scores of all three machine-learning-based approaches, there is a significant increase from MLR and k-NN to the hybrid model, illustrating the model's ability to address most of the misclassification issues faced by the other two models. The macro and weighted average F1-scores of 0.92 and 0.94 (Table 4) confirm that the hybrid model, by effectively incorporating k-NN's flexibility and MLR's interpretability, produced the most accurate classification among the three machine-learning-based approaches.

Table 4. Classification Report for the Hybrid Model Across Ten Microorganism Species

Microorganism	Precision	Recall	F1-Score	Support
1. Aspergillus	0.92	0.87	0.90	778
2. Diatom	0.83	0.98	0.90	364
3. Penicillium	0.95	1.00	0.98	216
4. Pithophora	0.88	0.64	0.74	270
5. Protozoa	0.97	1.00	0.98	778
6. Rhizopus	0.98	1.00	0.99	510
7. Spirogyra	0.98	0.77	0.86	122
8. Ulothrix	0.95	0.92	0.94	1484
9. Volvox	0.93	1.00	0.96	864
10. Yeast	0.94	0.95	0.95	720
Accuracy			0.94	6106
Macro avg.	0.93	0.91	0.92	6106
Weighted avg.	0.94	0.94	0.94	6106

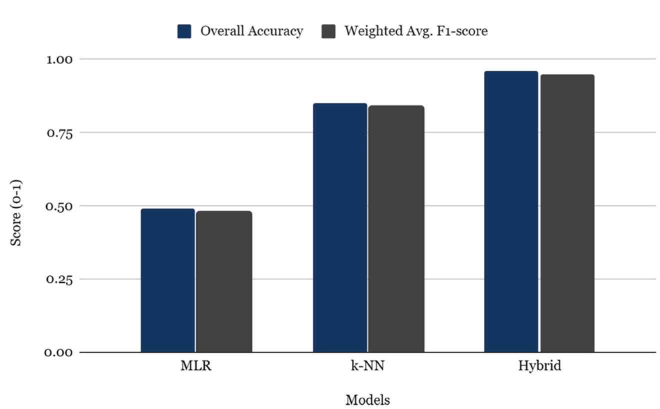

Figure 4. Comparison of Overall Accuracy and Weighted Average F1-score Across Three Classification Models.

The hybrid model achieved the highest accuracy in classifying microorganism species based on their morphological features as it effectively combines the strengths of both k-NN and MLR, while mitigating their limitations. MLR has the ability to consider all the samples at once to find general statistical relationships between species and their features, preventing the hybrid model from mistaking small local variations for new species. MLR also provides interpretable coefficients that precisely indicate which features are most important for classifying microorganisms, providing the hybrid model with a strong statistical foundation for decision-making. The hybrid model's use of k-NN improves local classification, as this approach is effective in identifying nonlinear similarities between microorganism samples. As shown in Table 4 and Figures 3–4, by integrating these strengths, the hybrid model achieved the highest overall accuracy and F1-scores across nearly all microorganism species.

CONCLUSIONS

This study employed machine-learning models to identify a more efficient microorganism identification and classification method. The research's findings reflect the potential of using machine learning to accelerate microorganism identification. For the study, three models were tested–multinomial logistic regression (MLR), k-nearest neighbor (k-NN), and a hybrid model that incorporates both. Among the three different models, the hybrid approach displayed the highest degree of accuracy and the most balanced performance overall. These results demonstrate that combining traditional modeling techniques to exploit their strengths can improve the accuracy of how microorganisms are classified through their geometric and structural features, effectively quickening the process without compromising reliability. If used alongside current cultivation techniques employed in microbiology clinics, machine learning can help healthcare providers identify potential infectious pathogens without bacterial cultivation.

Not only will this streamline treatment for patients–especially those in urgent situations–but it will also reduce the misuse of antibiotics.

While this study yielded promising results on the application of machine learning in healthcare, there were also certain limitations. Firstly, the dataset used to train the machine-learning models, while extensive, was imbalanced, with certain microorganism types being represented more frequently than others. Secondly, only one *k* value and one logistic model were tested. Future research on this topic should use a more balanced dataset and explore the incorporation of additional parameters and different advanced machine-learning models to obtain more comprehensive results. Additionally, expanding the dataset to include more microorganism species and different numerical descriptors of morphology would make the experiment more accurate. Overall, this study successfully showed that combining machine-learning techniques can streamline microorganism identification and classification without compromising reliability.

ACKNOWLEDGEMENT

I sincerely thank my research mentor for their invaluable guidance throughout this research. Their expertise, feedback, and support were crucial to the success and completion of this study. I also extend my gratitude to the anonymous reviewers and editors for their constructive feedback and input, and to Kaggle, which provided the dataset used in this research.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

REFERENCES

1. Nelson R, Bran M. Antibiotic prescribing and resistance: A review of global trends. *J Glob Antimicrob Resist.* 2022; 31: 45–53. https://doi.org/10.1016/j.jgar.2022.02.004.
2. Rani R, Kumar V, Singh M. Machine learning and deep learning-based computational approaches in automatic microorganisms image recognition: Methodologies, challenges, and developments. *Arch Comput Methods Eng.* 2021; 28 (8): 5733–5761. Available from: https://doi.org/10.1007/s11831-021-09639-x.
3. Murray PR, Rosenthal KS, Pfaller MA. *Medical Microbiology.* 9th ed. Philadelphia (PA): Elsevier; 2021. ISBN 9780323674508.
4. Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. *Pathog Dis.* 2015; 73 (4): ftv033. https://doi.org/10.1093/femspd/ftv033.
5. Sepsis Alliance Institute. What is sepsis? *Sepsis Alliance.* [date unknown]. Available from: https://www.sepsis.org/sepsis-basics/what-is-sepsis/ (accessed on 2025-07-22).
6. Kotwal S, Kshirsagar A, Ghosh P, Kale S. Automated bacterial classifications using machine learning: A review of methods and datasets (1998–2020). *Front Microbiol.* 2021; 12: 735480. https://doi.org/10.3389/fmicb.2021.735480.
7. Rani R, Kumar V, Singh M. Machine learning and deep learning-based computational approaches in automatic microorganisms image recognition: Methodologies, challenges, and developments. *Arch Comput Methods Eng.* 2021; 28 (8): 5733–5761. https://doi.org/10.1007/s11831-021-09639-x.
8. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013. ISBN 9780470582473. https://doi.org/10.1002/9781118548387
9. Menard S. Applied Logistic Regression Analysis. 2nd ed. Thousand Oaks (CA): Sage Publications; 2002. ISBN 9780761922087. https://doi.org/10.4135/9781412983433
10. Cover TM, Hart PE. Nearest neighbor pattern classification. *IEEE Trans Inf Theory.* 1967; 13 (1): 21–27. https://doi.org/10.1109/TIT.1967.1053964