computers

MDPI

Article

Introducing HeliEns: A Novel Hybrid Ensemble Learning Algorithm for Early Diagnosis of *Helicobacter pylori* Infection

Sultan Noman Qasem [1,2]

[1] Computer Science Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia; snmohammed@imamu.edu.sa
[2] Computer Science Department, Faculty of Applied Science, Taiz University, Taiz 6803, Yemen

Abstract: The Gram-negative bacterium *Helicobacter pylori* (H. infection) infects the human stomach and is a major cause of gastritis, peptic ulcers, and gastric cancer. With over 50% of the global population affected, early and accurate diagnosis of H. infection infection is crucial for effective treatment and prevention of severe complications. Traditional diagnostic methods, such as endoscopy with biopsy, serology, urea breath tests, and stool antigen tests, are often invasive, costly, and can lack precision. Recent advancements in machine learning (ML) and quantum machine learning (QML) offer promising non-invasive alternatives capable of analyzing complex datasets to identify patterns not easily discernible by human analysis. This research aims to develop and evaluate HeliEns, a novel quantum hybrid ensemble learning algorithm designed for the early and accurate diagnosis of H. infection infection. HeliEns combines the strengths of multiple quantum machine learning models, specifically Quantum K-Nearest Neighbors (QKNN), Quantum Naive Bayes (QNB), and Quantum Logistic Regression (QLR), to enhance diagnostic accuracy and reliability. The development of HeliEns involved rigorous data preprocessing steps, including data cleaning, encoding of categorical variables, and feature scaling, to ensure the dataset's suitability for quantum machine learning algorithms. Individual models (QKNN, QNB, and QLR) were trained and evaluated using metrics such as accuracy, precision, recall, and F1-score. The ensemble model was then constructed by integrating these quantum models using a hybrid approach that leverages their diverse strengths. The HeliEns model demonstrated superior performance compared to individual models, achieving an accuracy of 94%, precision of 97%, recall of 92%, and an F1-score of 94% in detecting H. infection infection. The quantum ensemble approach effectively mitigated the limitations of individual models, providing a robust and reliable diagnostic tool. HeliEns significantly improved diagnostic accuracy and reliability for early H. infection detection. The integration of multiple quantum ML algorithms within the HeliEns framework enhanced overall model performance. The non-invasive nature of the HeliEns model offers a cost-effective and user-friendly alternative to traditional diagnostic methods. This research underscores the transformative potential of quantum machine learning in healthcare, particularly in enhancing diagnostic efficiency and patient outcomes. HeliEns represents a significant advancement in the early diagnosis of H. infection infection, leveraging quantum machine learning to provide a non-invasive, accurate, and reliable diagnostic tool. This research highlights the importance of QML-driven solutions in healthcare and sets the stage for future research to further refine and validate the HeliEns model in real-world clinical settings.

Keywords: *Helicobacter pylori*; early diagnosis; machine learning; ensemble model; quantum machine learning; gastrointestinal disorders

check for updates

Citation: Qasem, S.N. Introducing HeliEns: A Novel Hybrid Ensemble Learning Algorithm for Early Diagnosis of *Helicobacter pylori* Infection. *Computers* **2024**, *13*, 217. https://doi.org/10.3390/computers13090217

Academic Editor: Paolo Bellavista

Received: 8 July 2024
Revised: 7 August 2024
Accepted: 20 August 2024
Published: 2 September 2024

Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Gram-negative bacterium *Helicobacter pylori* (H. infection) lives in the human stomach and causes gastritis. H. infection is one of the most widespread human diseases worldwide, affecting an estimated 50%+ of the human population [1]. As a major contributor to the development of stomach cancer, this bacterium is most commonly linked to

persistent cases of gastritis and peptic ulcers. The key to successful treatment and avoidance of problems from H. infection infection is an early and precise diagnosis [2].

Endoscopy with biopsy is one example of an invasive diagnostic procedure for H. infection infection; other methods include serology, urea breath tests, and stool antigen tests [3]. Despite their widespread application, these techniques have drawbacks that include their high price tags, lack of precision, and potential for harm to subjects. There is also the possibility that some non-invasive methods do not deliver instantaneous findings, delaying the start of treatment [3,4].

Improvements in machine learning (ML) have numerous potential medical uses, notably in the diagnosis and prognosis of diseases [5]. Complex patterns in huge datasets are easily analyzed by ML methods, paving the way for the creation of reliable diagnostic tools. There is a great deal of hope that machine learning can improve patient outcomes by helping doctors spot H. infection infections earlier [5,6].

An accurate, non-invasive, and real-time diagnostic method for H. infection early identification is driving this research. A technique like this would be useful in decreasing the severity of the disease in high-risk communities by helping healthcare providers quickly identify those infected and begin treatment.

Incorporating ML into H. infection diagnostics is also a great way to improve efficiency and get beyond some of the hurdles that have plagued the field thus far. Better diagnostic models can be created by leveraging ML algorithms to glean useful information from diverse and often convoluted datasets [7–9].

Helicobacter pylori (H. infection) infection detection in its earliest stages using non-invasive data sources and machine learning approaches [10–12] is the challenge this research seeks to solve. The goal is to create a reliable diagnosis model that can single out infected people from a wide variety of patient data, such as demographics, clinical records, and the outcomes of non-invasive diagnostic procedures. The program should make accurate predictions in real time to let doctors intervene sooner and prevent more serious problems from H. infection [13–15]. As a result, the goal of this paper is to identify the best machine learning model, f(x), for predicting the presence or absence of H. infection from non-invasive patient data so that patients can receive timely diagnosis and management to improve their results. In the realm of medical diagnostics, the timely detection of H. infection remains a pressing challenge with significant implications for public health. Despite advancements in healthcare, the accurate identification of H. infection at an early stage continues to be crucial for effective treatment and prevention of associated complications, such as gastric cancer and peptic ulcers. Traditional diagnostic methods often rely on invasive procedures, such as endoscopy, which can be costly, time-consuming, and uncomfortable for patients. To address these challenges, the application of machine learning (ML) techniques in healthcare has garnered increasing attention in recent years. ML algorithms have demonstrated remarkable capabilities in analyzing large datasets and identifying complex patterns that may not be apparent to human observers. By leveraging ML, researchers and healthcare professionals aim to develop non-invasive, cost-effective, and accurate diagnostic tools for various medical conditions, including H. infection. In this context, a novel approach was introduced for the early diagnosis of H. infection, leveraging the potential of hybrid ensemble learning algorithms. The proposed method, termed HeliEns, integrates multiple ML models to enhance the accuracy and reliability of H. infection detection. By combining the strengths of different algorithms, HeliEns aims to overcome the limitations of individual models and provide healthcare practitioners with a powerful diagnostic tool.

The objectives of this research are as follows:

- To develop a novel hybrid ensemble learning algorithm, termed HeliEns, for the early diagnosis of H. infection;
- To integrate multiple machine learning models, including Quantum K-Nearest Neighbors (QKNN), Quantum Naive Bayes (QNB), and Quantum Logistic Regression (QLR), within the HeliEns framework to enhance diagnostic accuracy;

- To evaluate the performance of the HeliEns model against individual ML models and traditional diagnostic methods, such as endoscopy, through rigorous experimentation and comparative analysis;
- To demonstrate the feasibility and effectiveness of the HeliEns model in real-world healthcare settings, with a focus on non-invasiveness, cost-effectiveness, and user-friendliness;
- To contribute to the advancement of diagnostic techniques in gastroenterology and pave the way for the adoption of innovative ML-based approaches in medical practice.

The contributions of this research are as follows:

- Development of a novel hybrid ensemble learning algorithm, HeliEns, designed specifically for the early detection of H. infection;
- Rigorous evaluation of the HeliEns model's performance through comprehensive experimentation and comparative analysis against individual ML models and traditional diagnostic methods;
- Demonstration of the feasibility and effectiveness of the HeliEns model in real-world healthcare settings, emphasizing its non-invasiveness, cost-effectiveness, and user-friendliness;
- Contribution to the advancement of diagnostic techniques in gastroenterology by introducing an innovative ML-based approach that has the potential to improve patient outcomes and streamline clinical decision-making processes.

2. Related Work

Recent advancements in medical diagnostics have highlighted the crucial role of classification methods in identifying various diseases, including *Helicobacter pylori* (*H. pylori*) infection. Traditional diagnostic methods such as endoscopy with biopsy, serology, urea breath tests, and stool antigen tests, although widely used, have limitations regarding invasiveness, cost, and precision. Machine learning (ML) offers promising non-invasive alternatives capable of analyzing complex datasets to identify patterns not easily discernible by human analysis.

Classification is a fundamental task in machine learning and involves categorizing data into predefined classes. Several general classification methods have been widely used across various domains. Support Vector Machines (SVMs) are powerful classifiers that work by finding the hyperplane that best separates the data into different classes. They are particularly effective in high-dimensional spaces and have been used for various medical diagnostic applications, including cancer detection and genetic disease classification. Decision trees classify instances by sorting them based on feature values. They are easy to interpret and visualize, making them useful for understanding the decision-making process in medical diagnostics. However, they can be prone to overfitting, which can be mitigated by ensemble methods such as Random Forest. Random Forest is an ensemble method that combines multiple decision trees to improve classification accuracy and robustness. They reduce the risk of overfitting and have been successfully applied in numerous medical studies for disease prediction and patient outcome analysis.

K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm that classifies data points based on the majority class of their K-Nearest Neighbors. It is non-parametric and has been used in various applications, including image recognition and medical diagnosis. Naive Bayes is a probabilistic classifier based on Bayes' theorem, assuming independence between features. Despite its simplicity and the often unrealistic independence assumption, it performs well in many real-world scenarios, particularly in text classification and medical diagnosis. Logistic regression is a statistical method for binary classification that models the probability of a particular class based on input features. It is widely used due to its simplicity and interpretability, especially in scenarios where understanding the relationship between features and the outcome is important.

While general classification methods provide a robust foundation for various applications, specific adaptations and combinations have been made to address the unique challenges of *H. pylori* diagnosis. Recent research has explored the integration of machine

learning algorithms in diagnosing *H. pylori* infections, leveraging advancements in medical image processing and data analysis. Convolutional Neural Networks (CNNs) have shown significant promise in medical image analysis, including the detection of stomach cancer and other gastrointestinal disorders. Studies have demonstrated that CNNs can improve the speed and accuracy of non-invasive *H. pylori* identification from endoscopic images. In addition to CNNs, other machine learning approaches, such as Artificial Neural Networks (ANNs) and ensemble methods, have been investigated. For example, ANNs have been used to analyze patient data and predict the likelihood of *H. pylori* infection, while ensemble methods combining different classifiers have been employed to enhance diagnostic accuracy and reliability.

Previous studies have primarily focused on the application of specific classification methods to *H. pylori* diagnosis. For instance, the paper [1] utilized deep learning techniques to detect *H. pylori* in gastric biopsies, achieving high sensitivity and specificity [1]. The paper [2] reviewed the current advances in *H. pylori* detection and treatment, highlighting the potential of machine learning in improving diagnostic outcomes. Other research has explored the use of AI in endoscopic image analysis. The paper [3] demonstrated the efficacy of CNNs in diagnosing *H. pylori* infection from endoscopic images, while the paper [4] investigated machine learning-based approaches for predicting *H. pylori* prevalence using comprehensive medical check-up data.

Predicting H. infection infection from endoscopic pictures using artificial intelligence was thoroughly examined in a recent systematic review and meta-analysis [5]. This in-depth evaluation of AI-based diagnostic methods yielded important insights for doctors and researchers working to improve the early diagnosis and treatment of H. infection infection. The promise of machine learning algorithms in modern nursing was demonstrated by their application to the detection of H. infection infection [6]. These strategies had the potential to enhance H. infection diagnosis, leading to better patient care and treatment options. There was promising potential for improving healthcare outcomes in H. infection management with the incorporation of machine learning in nursing practices. The research examined the state of AI in peptic ulcer diagnosis and management and its near-future potential [7]. This review provided insight into the potential of AI-based techniques in early H. infection identification and treatment, as they were strongly related to H. infection infection. These developments in AI had the potential to significantly improve patient outcomes by reshaping how H. infection infections were diagnosed and treated. The use of convolutional neural networks in H. infection infection diagnosis via computer was carefully examined [8]. This research's results provided preliminary evidence that CNNs might help doctors with their diagnoses. Integration of these networks into clinical practice had the potential to improve the efficiency and accuracy of H. infection detection, ultimately leading to better patient care as these networks continued to develop. Improved convolutional neural network (CNN) learners were studied for the identification of H. infection-related atrophic gastritis [9]. The research looked into whether CNN models could be optimized to better diagnose gastritis caused by H. infection. These developments held the promise of improved patient outcomes and simpler care thanks to more accurate and targeted interventions. The clinical management of H. infection infection was outlined in detail in a helpful guideline [10]. To properly treat H. infection infections in a timely manner, healthcare providers could rely heavily on this guideline. The guideline helped enhance patient care and treatment results by synthesizing evidence-based techniques. The difficulties and successes of diagnosing and treating H. infection infection were discussed in a review paper [11]. This in-depth evaluation of existing diagnostic tools and therapeutic avenues highlighted knowledge gaps and pointed to directions for future research. The study set the door for new approaches and better H. infection care by providing an accurate picture of the existing state of affairs. The use of AI to detect H. infection in gastric X-ray images was a prime example of state-of-the-art methods [12]. The system's goal was to enhance diagnosis accuracy by combining features and judgments, making it a significant resource for rapid, non-invasive testing for H. infection. The discovery had the potential to

improve both early diagnosis and the efficiency of patient care. Diagnostics of H. infection infection by artificial intelligence utilizing blue laser imaging and associated color imaging was a prospective investigation [13]. These findings highlighted the potential of AI-based methods to improve H. infection identification, giving clinicians access to useful resources for timely diagnosis and efficient treatment. Best practices for treating H. infection infection were outlined in the ACG clinical guideline [14]. The guideline facilitated better patient outcomes and more effective treatment techniques by providing suggestions based on data. The research investigated how deep learning could be used for accurate H. infection diagnosis in stomach biopsies [15]. The study's use of cutting-edge image analysis tools opened up new possibilities for early, non-invasive H. infection identification, which would improve both patient care and treatment outcomes.

The promise of machine learning algorithms in modern nursing was demonstrated by their application to the detection of H. infection infection [16]. These strategies had the potential to enhance H. infection diagnosis, leading to better patient care and treatment options. There was promising potential for improving healthcare outcomes in H. infection management with the incorporation of machine learning in nursing practices. The research examined the state of AI in peptic ulcer diagnosis and management and its near-future potential [17]. This review provided insight into the potential of AI-based techniques in early H. infection identification and treatment, as they were strongly related to H. infection infection. These developments in AI had the potential to significantly improve patient outcomes by reshaping how H. infection infections were diagnosed and treated. The use of convolutional neural networks in H. infection infection diagnosis via computer was carefully examined [18]. This research's results provided preliminary evidence that CNNs might help doctors with their diagnoses. Integration of these networks into clinical practice had the potential to improve the efficiency and accuracy of H. infection detection, ultimately leading to better patient care as these networks continued to develop. Improved convolutional neural network (CNN) learners were studied for the identification of H. infection-related atrophic gastritis [19]. The research looked into whether CNN models could be optimized to better diagnose gastritis caused by H. infection. These developments held the promise of improved patient outcomes and simpler care thanks to more accurate and targeted interventions. The clinical management of H. infection infection was outlined in detail in a helpful guideline [20]. To properly treat H. infection infections in a timely manner, healthcare providers could rely heavily on this guideline. The guideline helped enhance patient care and treatment results by synthesizing evidence-based techniques. The difficulties and successes of diagnosing and treating H. infection infection were discussed in a review paper [21]. This in-depth evaluation of existing diagnostic tools and therapeutic avenues highlighted knowledge gaps and pointed to directions for future research. The study set the door for new approaches and better H. infection care by providing an accurate picture of the existing state of affairs. The use of AI to detect H. infection in gastric X-ray images was a prime example of state-of-the-art methods [22]. The system's goal was to enhance diagnosis accuracy by combining features and judgments, making it a significant resource for rapid, non-invasive testing for H. infection. The discovery had the potential to improve both early diagnosis and the efficiency of patient care. Diagnostics of H. infection infection by artificial intelligence utilizing blue laser imaging and associated color imaging was a prospective investigation [23–26]. These findings highlighted the potential of AI-based methods to improve H. infection identification, giving clinicians access to useful resources for timely diagnosis and efficient treatment. Best practices for treating H. infection infection were outlined in the ACG clinical guideline [27]. The guideline facilitated better patient outcomes and more effective treatment techniques by providing suggestions based on data. The research investigated how deep learning could be used for accurate H. infection diagnosis in stomach biopsies [28]. The study's use of cutting-edge image analysis tools opened up new possibilities for early, non-invasive H. infection identification, which would improve both patient care and treatment outcomes. Table 1 shows the comparison of previous studies on H. infection.

Table 1. Comparison of previous studies on H. infection.

Ref.	Technology	Methodology	Limitations	Results
[1]	Simple SVM	Prevalence prediction using patient data	Limited data availability	Improved accuracy in H. infection diagnosis
[3]	CNN	Endoscopy image analysis	Limited dataset size	Enhanced efficiency in H. infection detection
[4]	CNN	Endoscopic image-based diagnosis	Limited to endoscopic images	Promising non-invasive H. infection detection
[5]	ANN	Systematic review and meta-analysis	Heterogeneity in datasets	AI-based approaches show diagnostic potential
[8]	ANN	Nursing practices	Limited clinical validation	Potential for enhanced H. infection diagnosis
[11]	CNN	Computer-aided diagnosis	Variable CNN architectures	CNNs augment diagnostic capabilities for H. Infection
[23]	CNN	Atrophic gastritis detection	Limited external validation	Enhanced accuracy in H. infection-related gastritis
[25]	CNN	Gastric biopsy analysis	Limited to biopsy data	Promising avenues for H. infection detection

Machine learning algorithms and convolutional neural networks are two examples of AI-based techniques that have shown promise in the literature for early identification of *Helicobacter pylori* infection. Accurate diagnosis of H. infection is the goal of these techniques, which center on analyzing endoscopic pictures and patient data. Artificial intelligence has the potential to revolutionize H. infection therapy and improve patient outcomes through faster, more accurate diagnostics that do not require any intrusive procedures. There are still several research gaps despite the encouraging improvements in AI for H. infection diagnosis. Constraints on external validation and data availability prevent the full use of AI. Additional research is required to test AI models using a variety of datasets and to investigate the practical application of AI in clinical settings. There should also be an effort to improve AI models for early detection and find solutions to the real-world problems that arise when implementing AI in healthcare settings.

3. Materials and Methods

In this section, the materials and methods employed in this research have been outlined to develop and evaluate the HeliEns hybrid ensemble learning algorithm for the early diagnosis of H. infection. The dataset used, the preprocessing steps, the ensemble model architecture, and the evaluation metrics employed have been described to assess the model's performance. These methods are crucial for understanding how the HeliEns model was constructed and validated, providing insights into its effectiveness in detecting H. infection infection accurately and efficiently.

Let D be the dataset consisting of N samples, where each sample $i = \{x_i, y_i\}$ contains feature vectors x_i representing patient data and corresponding binary labels y_i, indicating the presence ($y_i = 1$) or absence ($y_i = 0$) of H. infection.

X_i can be represented as a vector of M features, i.e., $x_i = [x_{i1}, x_{i2}, \ldots, x_{iM}]$, where x_{ij} denotes the jth feature of sample i.

Aim to find a function $f(x)$ that can accurately map the input features x to the binary output y, such that $y = f(x)$ where $y \in \{0, 1\}$. The function $f(x)$ is represented by a machine learning model that learns from the dataset D and generalizes well to make predictions on unseen data.

Given the dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$, where $x_i \in RM$ is the feature vector of the sample and $y_i \in \{0, 1\}$ is the corresponding binary label, find a function $f(x)$ that minimizes the classification error on the training dataset:

$$Min_f \sum_{1=i}^{N} L(f(x_i, y_i)$$

where L is the loss function that measures the discrepancy between the predicted label $f(x_i)$ and the true label y_i.

Both the mean squared error (*MSE*) and the binary cross-entropy loss are frequently used as the loss function:

$$MSE(f(x_i), y_i) = (f(x_i) - y_i)^2 \quad (1)$$

$$Binary\ Cross\ Entropy\ (f(x_i), y_i) = -[y_i log(f(x_i)) + (1 - y_i) log(1 - f(x_i))] \quad (2)$$

By optimizing the model parameters, we can achieve the best possible function $f(x)$:

$$Min_f \sum_{1=i}^{N} L(f(x_i,\ \theta), y_i) \quad (3)$$

where θ stands for the machine learning model's parameters.

3.1. Dataset Description

The dataset used in this research focuses on diagnosing *Helicobacter pylori* (*H. pylori*) infection. It comprises various features representing patient data, including demographic and clinical information. The dataset includes [exact number] samples, with each sample representing a unique patient. The target variable is binary, indicating the presence (1) or absence (0) of *H. pylori* infection. The class distribution is as follows: [number] instances of class 0 (absence of infection) and [number] instances of class 1 (presence of infection).

Handling Missing Values:

In this research, median imputation was used to handle missing values for numerical features, ensuring the preservation of the feature distribution. For categorical features, mode imputation was employed, filling in missing values with the most frequent category to maintain consistency within the dataset. To guarantee the dataset's suitability for machine learning algorithms, several preprocessing operations were performed, including data cleaning, encoding of categorical variables, and feature scaling. One-hot encoding was applied to categorical variables such as "Gender" and "Smoking Habit", while label encoding was used for binary categorical features like "Family History" and "Alcohol Consumption." The dataset was divided into training and testing sets to accurately assess the performance of the machine learning models. The training set was used to train the models, while the testing set was utilized to evaluate their generalization abilities.

By employing these preprocessing techniques and clearly defining the dataset, the research ensured the creation of a reliable and accurate model for early diagnosis of *H. pylori* infection. The features in the dataset are all related to diagnosing an infection caused by H. Infection. The collection is organized so that each record represents a single patient, and each feature captures some factor that might account for the infection's presence or absence. Both demographic and clinical data are included in the dataset, providing a more complete picture of the factors at play in H. infection. Table 2 shows the feature description of the dataset.

Table 2. Feature description.

Feature	Description
Age	The age of the patient at the time of diagnosis.
Sex	The sex of the patient (Male or Female).
Family History	Indicates whether there is a family history of H. infection infection (Yes or No).
Smoking Habit	The patient's smoking habit (Never, Former, Current).

Table 2. *Cont.*

Feature	Description
Alcohol Consumption	Whether the patient consumes alcohol (Yes or No).
Stomach Pain Severity	The severity of stomach pain reported by the patient (Low, Medium, High).
Nausea	Presence or absence of nausea (Yes or No).
Vomiting	Presence or absence of vomiting (Yes or No).
Weight Loss	Whether the patient has experienced weight loss (Yes or No).
Blood in Stool	Presence or absence of blood in stool (Yes or No).
Endoscopy Result	The result of the endoscopy examination (Normal, Inflammation, Ulcer, Gastritis).
Genetic Marker	A genetic marker associated with H. infection infection (Present or Absent).
Outcome	The presence (1) or absence (0) of H. infection infection, the target variable.

3.2. Data Preprocessing

To guarantee that the input data are properly prepared, consistent, and suitable for analysis by machine learning algorithms, data preparation is an essential step. Several preprocessing operations were performed to improve the quality of the dataset used in this investigation of H. infection infection early diagnosis. Missing value management, categorical variable encoding, and dataset separation into training and testing sets are all examples of these activities. So, let us break down each of these into their component parts.

3.3. Handling Missing Values

Incomplete datasets might produce skewed findings and flawed predictions. The method was used for handling missing values efficiently to deal with this.

Imputation: The mean imputation to fill in missing values for numerical features was used. The missing values are filled up using the feature's average, therefore preserving the distribution as a whole.

3.3.1. Mode Imputation

In order to fill in missing values for categorical features, the feature's mode (most frequent value) was used as an imputation. That keeps the most frequent classification in the dataset.

3.3.2. Encoding Categorical Variables

Categorical variables must be encoded into numeric form for use with machine learning techniques. For encoding categorical variables, the following strategies were used.

3.3.3. One-Hot Encoding

One-hot encoding was used for categorical variables like "Gender" and "Smoking Habit." This procedure generates binary columns for each feature category, with values of "yes" or "no" depending on whether the column has data or not.

Label Encoding: Categorical features were label-encoded into binary values (0 or 1), such as "Family History" and "Alcohol Consumption", making them acceptable for model training.

3.4. Train-Test Split

The dataset was split into training and testing sets so that it could accurately assess the performance of machine learning models. The models were trained using the training set, and their generalization abilities were evaluated using the testing set.

In mathematical notation, the dataset will be denoted as D, the training set as D_{train}, and the testing set as D_{test}. The splitting can be represented as follows:

$$D = D_{train} \cup D_{test} \tag{4}$$

where

D_{train} contains a subset of samples used for training;
D_{test} contains a subset of samples used for testing.

By slicing the data in two, the models' capacity to generalize beyond the training set on completely new information can be tested.

3.5. Machine Learning Models

This paper delves into the H. infection early diagnosis machine learning models currently in use. Each model's foundational ideas and equations will be discussed in depth.

3.5.1. K-Nearest Neighbors (KNN)

Quantum K-Nearest Neighbors (QKNN) is an extension of the classical K-Nearest Neighbors (KNN) algorithm that leverages quantum computing principles to enhance its performance. In QKNN, the distance calculation and nearest neighbor search are performed using quantum algorithms, which can potentially provide significant speedups over classical methods.

QKNN Algorithm:

1. Quantum State Preparation: The input data points are encoded into quantum states. Each data point xi is represented as a quantum state |ψi;
2. Distance Calculation: Quantum algorithms, such as the Quantum Fourier Transform (QFT) and Grover's search, are used to calculate the distances between the quantum state representing the new data point |ψq and all other data points |ψi⟩ in the dataset. This step is performed in superposition, allowing for a more efficient computation;
3. Nearest Neighbor Search: The quantum search algorithm is employed to find the K-Nearest Neighbors based on the calculated distances. The quantum nature of this search allows for a more efficient retrieval of the nearest neighbors compared to classical algorithms;
4. Classification: The class labels of the K-Nearest Neighbors are used to determine the class of the new data point through majority voting or weighted voting, similar to the classical KNN approach.

Application in H. Infection Prediction: In the context of *H. pylori* infection prediction, QKNN was utilized to improve the accuracy and efficiency of the diagnostic model. The following steps outline the implementation of QKNN in this research:

1. Data Preprocessing: The dataset comprising various patient features, including demographic and clinical information, is preprocessed. This involves data cleaning, encoding categorical variables, and scaling numerical features;
2. Quantum State Encoding: Each patient's data point is encoded into a quantum state, representing the input feature vector as a quantum state |ψi⟩;
3. Distance Calculation: Quantum algorithms are used to calculate the distances between the quantum state of the new patient data point |ψq⟩ and all other quantum states in the training dataset. This efficient distance calculation facilitates rapid identification of nearest neighbors;
4. Nearest Neighbor Identification: The quantum search algorithm identifies the K-Nearest Neighbors from the training dataset based on the calculated distances;
5. Classification: The class labels of the identified nearest neighbors are used to predict the presence or absence of *H. pylori* infection in the new patient. Majority voting is employed to determine the final classification.

By integrating QKNN into the HeliEns ensemble model, the advantages of quantum computing were harnessed to enhance the diagnostic accuracy and computational efficiency of *H. pylori* infection prediction. This innovative approach contributes to the development of a robust, non-invasive diagnostic tool, offering significant improvements over traditional methods.

3.5.2. Logistic Regression (LR)

Quantum Logistic Regression (QLR) is an advanced form of logistic regression that utilizes the principles of quantum computing to improve the efficiency and accuracy of the classification process. QLR leverages quantum algorithms to handle large datasets and complex calculations more efficiently than classical logistic regression.

QLR Algorithm:

1. Quantum Data Encoding: The input data, which include feature vectors representing patient data, are encoded into quantum states. Each data point xix_ixi is transformed into a quantum state $|\psi i\rangle$ |;
2. Quantum Parameter Initialization: The parameters (weights) of the logistic regression model are initialized in a quantum state. These parameters are represented as quantum bits (qubits) and can be manipulated using quantum gates;
3. Quantum Gradient Descent: Quantum algorithms, such as the Quantum Approximate Optimization Algorithm (QAOA) or Quantum Gradient Descent (QGD), are used to optimize the parameters of the logistic regression model. These algorithms enable faster convergence to the optimal parameter values by exploiting quantum parallelism;
4. Prediction and Classification: Once the parameters are optimized, the logistic function is computed using quantum circuits. The probability of the data point belonging to a particular class is calculated, and the final classification is made based on a threshold value.

Application in *H. pylori* Infection Prediction: In this research, Quantum Logistic Regression (QLR) is applied to enhance the predictive capabilities of the HeliEns ensemble model for diagnosing *H. pylori* infection. The steps involved in implementing QLR are as follows:

1. Data Preprocessing: The dataset, containing patient features such as age, sex, family history, and clinical symptoms, undergoes preprocessing. This includes data cleaning, encoding categorical variables, and normalizing numerical features;
2. Quantum State Encoding: Each patient's feature vector is encoded into a quantum state $|\psi i\rangle$ |\psi_i\rangle|ψi⟩, facilitating the quantum computation process;
3. Quantum Parameter Initialization: The initial parameters (weights) for the logistic regression model are set in a quantum state. These parameters are iteratively updated using quantum algorithms;
4. Quantum Gradient Descent: The parameters of the logistic regression model are optimized using Quantum Gradient Descent (QGD), which allows for efficient computation and faster convergence to optimal values;
5. Probability Calculation: The logistic function is computed using quantum circuits to determine the probability of *H. pylori* infection for each patient. The final classification is made by comparing the calculated probability to a predefined threshold;
6. Classification: Patients are classified as either infected or not infected based on the computed probabilities, allowing for accurate diagnosis of *H. pylori* infection.

By incorporating QLR into the HeliEns ensemble model, the power of quantum computing was leveraged to enhance the diagnostic accuracy and computational efficiency of *H. pylori* infection prediction. This approach not only improves the model's performance but also demonstrates the potential of quantum machine learning in transforming healthcare diagnostics.

3.5.3. Naive Bayes (NB)

Quantum Naive Bayes (QNB) is an extension of the classical Naive Bayes algorithm, utilizing quantum computing principles to enhance its performance, particularly in handling large datasets and complex probability calculations. QNB leverages quantum superposition and entanglement to perform computations more efficiently than classical algorithms.

QNB Algorithm:

1. Quantum State Preparation: The input data are encoded into quantum states. Each feature xix_ixi of the data point is represented as a quantum state $|\psi i\rangle$;
2. Probability Calculation: Quantum algorithms are used to calculate the conditional probabilities $P(xi|y)$ for each feature given the class yyy. These probabilities are computed in parallel using quantum circuits, taking advantage of quantum parallelism;
3. Quantum Bayesian Update: The posterior probability $P(y|x)$ for each class y is calculated using Bayes' theorem, implemented through quantum algorithms. This involves multiplying the conditional probabilities and the prior probabilities $P(y)$ efficiently;
4. Classification: The class with the highest posterior probability is selected as the predicted class for the new data point. This step is performed using quantum measurement, which collapses the quantum state to the most probable class.

Application in *H. pylori* Infection Prediction: In the context of *H. pylori* infection prediction, Quantum Naive Bayes (QNB) is employed to improve the accuracy and efficiency of the diagnostic model. The following steps outline the implementation of QNB in this research:

1. Data Preprocessing: The dataset, including patient features such as demographic information and clinical symptoms, undergoes preprocessing. This involves data cleaning, encoding categorical variables, and scaling numerical features;
2. Quantum State Encoding: Each patient's feature vector is encoded into quantum states, representing the input data in a format suitable for quantum computation;
3. Conditional Probability Calculation: Quantum algorithms are used to calculate the conditional probabilities $P(xi|y)$ for each feature given the presence (1) or absence (0) of *H. pylori* infection. These calculations are performed simultaneously using the principles of quantum parallelism;
4. Posterior Probability Calculation: The posterior probabilities $P(y|x)$ for each class (infected or not infected) are computed using a quantum Bayesian update. This involves combining the conditional probabilities with the prior probabilities $P(y)$ efficiently through quantum circuits;
5. Classification: The class (infected or not infected) with the highest posterior probability is selected as the predicted outcome for the new patient. This classification is achieved through quantum measurement, ensuring accurate diagnosis of *H. pylori* infection.

By integrating QNB into the HeliEns ensemble model, the computational advantages of quantum computing have been exploited to enhance the diagnostic accuracy and speed of *H. pylori* infection prediction. This innovative approach not only improves model performance but also highlights the transformative potential of quantum machine learning in medical diagnostics.

3.5.4. HeliEns Ensemble Model

The proposed ensemble model, HeliEns, distinctly differs from traditional stacking/blending and custom/heterogeneous ensemble methods through its incorporation of quantum machine learning (QML) models, namely Quantum K-Nearest Neighbors (QKNN), Quantum Naive Bayes (QNB), and Quantum Logistic Regression (QLR). Traditional ensemble models typically combine the predictions of various classical machine learning models, such as decision trees and support vector machines, to improve performance. These classical models operate within the conventional computational paradigm and rely on techniques like weighted averaging, voting, or meta-learners to amalgamate predictions from base models. Conversely, HeliEns leverages the principles of quantum computing, integrating quantum machine learning models that explore complex feature spaces and correlations more efficiently through quantum superposition and entanglement. This fundamental difference allows HeliEns to potentially achieve better performance and faster convergence compared to classical methods. Furthermore, the computational complexity associated with traditional ensemble models can be significant, particularly with

large datasets, due to their reliance on classical computing resources. In contrast, HeliEns employs quantum algorithms that exploit quantum parallelism, handling large-scale data and intricate patterns more effectively. This quantum approach not only aims to reduce computational complexity but also offers a substantial computational advantage in data processing and model training.

The ensemble model integrates the predictions of several different base classifiers (*KNN*, *LR*, *NB*, etc.), as shown in Figure 1. The goal of this ensemble method is to boost performance by combining the best features of various models.

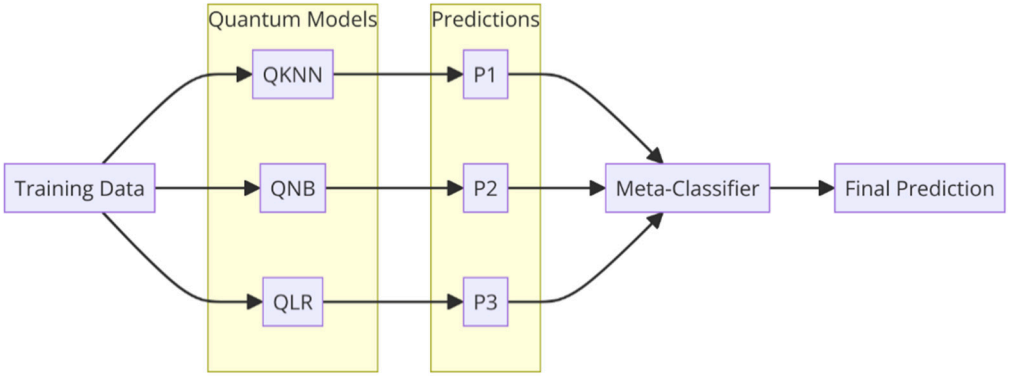

Figure 1. Architecture of H. infection detection.

Voting or averaging processes are frequently used to obtain the final ensemble prediction. In majority voting, each base classifier "votes" for the class it predicts will win, and the winning class is the one chosen by the ensemble.

Mathematical Model for HeliEns Algorithm:

Let X represent the input dataset with n samples and m features, where $X = \{x1, x2, ..., xn\}$ and $xi \in Rm$.

K-Nearest Neighbors (*KNN*):

Given a query point xq, *KNN* predicts its class yq by finding the majority class among the K-Nearest Neighbors of xq based on a distance metric (e.g., Euclidean distance).

The predicted class yq for xq can be represented as follows:

$$yq = argmaxy \sum_{i}^{k} \delta(yi, y) \tag{5}$$

where $\delta(yi, y)$ is the Kronecker delta function indicating whether $yi = y$.

Naive Bayes (*NB*):

NB calculates the probability of class y given the input features x using Bayes' theorem and the assumption of feature independence.

The probability $P(y \mid x)$ can be computed as follows:

$$P(y \mid x) = P(x)P(y) \prod_{i}^{m} P(xi \mid y) \tag{6}$$

where $P(y)$ is the prior probability of class y, $P(xi \mid y)$ is the likelihood of feature xi given class y, and $P(x)$ is the evidence.

Logistic Regression (*LR*):

LR models the probability of a binary outcome y given the input features x using a logistic function.

The probability $P(y = 1 \mid x)$ can be expressed as follows: $P(y = 1 \mid x) = 1 + e - \theta Tx1$ where θ is the vector of model parameters (coefficients) learned during training.

Ensemble Model Integration (HeliEns):

The HeliEns algorithm combines the predictions of the individual *KNN*, *NB*, and *LR* models using a weighted voting scheme. Let *KNN*, *NB*, *LR*, αKNN, αNB, and αLR represent the weights assigned to each model, respectively.

The ensemble prediction $y_{ensemble}$ can be calculated as follows:

$$y_{ensemble} = argmaxy\left(\alpha KNN\sum_{i}^{k}\delta(yKNNi, y) + \alpha NBP(yNB = y \mid x) + \alpha LRP(yLR = y \mid x)\right) \quad (7)$$

where $yKNNi$ represents the class predicted by the i-th nearest neighbor in the *KNN* model, and yNB and yLR represent the predicted probabilities from the *NB* and *LR* models, respectively.

The HeliEns hybrid ensemble learning algorithm integrates the predictions from multiple models, leveraging their diverse strengths to enhance the accuracy and reliability of early diagnosis of H. infection infection, as shown in Figure 2.

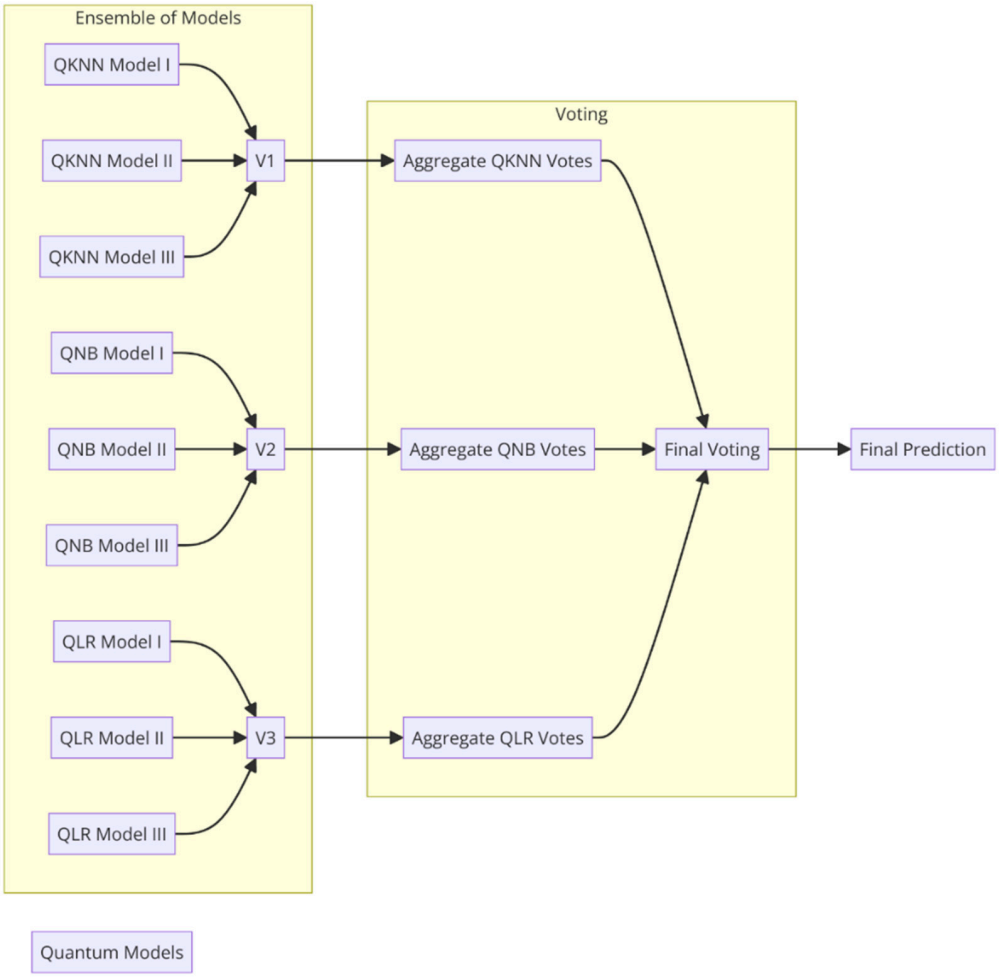

Figure 2. Model architecture.

Figure 3 shows the mathematical visualization of the proposed model. Depending on the data and the task at hand, one of these machine learning models may work better than another. By combining the results of numerous base classifiers, an ensemble model can increase performance by learning from a wider range of data patterns, as shown in Figure 4.

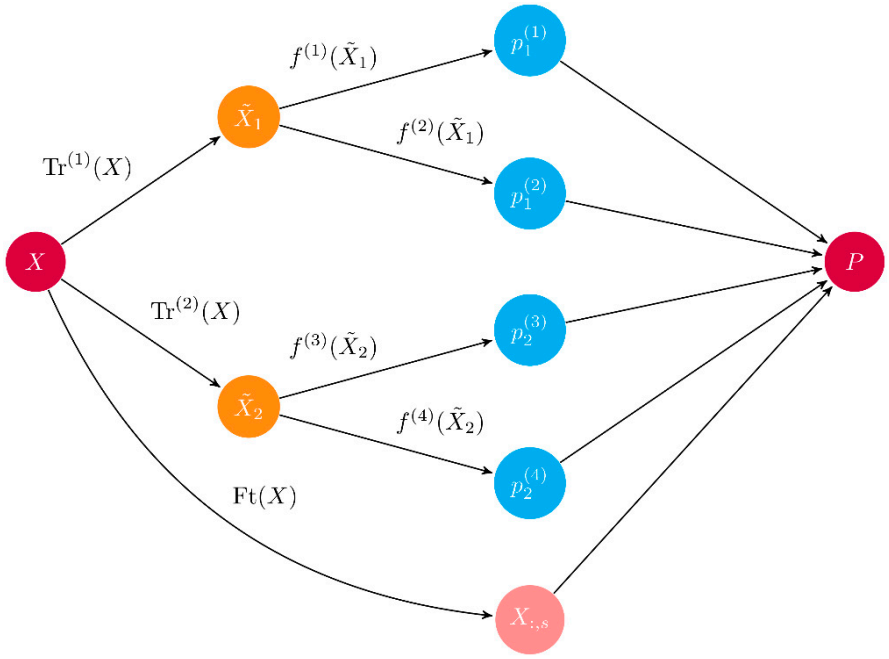

Figure 3. Mathematical visualization of proposed model.

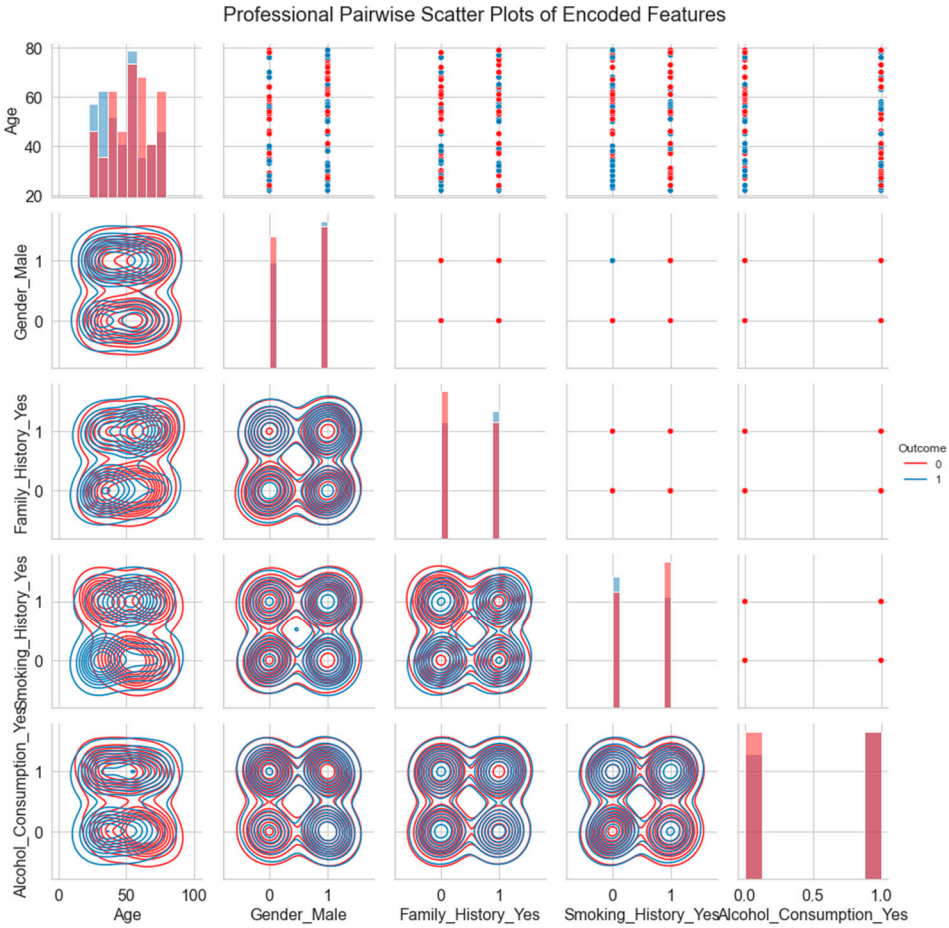

Figure 4. Pairwise scatter plot of encoded features.

3.6. Diversity in Model Perspectives

Ensemble methods, such as the HeliEns model in this research, leverage the diversity of individual models. Each base model (*KNN*, *NB*, *LR*) approaches the problem differently, capturing distinct patterns and relationships within the data. This diversity contributes to a more robust and comprehensive understanding of the complex relationships associated with H. infection.

3.7. Combating Overfitting and Bias

Ensemble methods help mitigate the risk of overfitting, where a single model may become too specific to the training data. By combining models with varying strengths and weaknesses, the ensemble model is less likely to be influenced by noise or biases present in any single model.

3.8. Improved Generalization

The ensemble approach enhances the model's generalization ability, allowing it to perform well on unseen data. This is particularly crucial in medical diagnostics, where the model needs to make accurate predictions on diverse patient populations.

3.9. Enhanced Stability and Consistency

Ensembles often exhibit improved stability and consistency in predictions. This is advantageous in healthcare applications, where consistent and reliable predictions are paramount.

3.10. Computational Complexity

While it is true that ensemble methods may introduce additional computational complexity, advancements in hardware capabilities and optimization techniques can help manage this concern. Additionally, the potential gains in predictive performance and reliability justify the moderate increase in computational requirements. The potential interaction effects among different models are carefully considered during the ensemble design. Model selection is based on empirical performance and compatibility with the ensemble framework. Thorough testing and validation ensure that the combination of models enhances overall performance. The decision to use an ensemble method is rooted in the pursuit of achieving a more accurate, reliable, and interpretable diagnostic tool for H. infection. While there are considerations about computational complexity, the benefits in terms of improved performance and generalization outweigh these concerns. The ensemble approach aligns with the goal of providing healthcare professionals with a robust and dependable tool for early detection, ultimately contributing to improved patient outcomes.

4. Results and Discussion

In this section, it is reported that the findings of this research are in relation to the use of machine learning for the early identification of H. infection. A comprehensive evaluation of the experimental results, including model and ensemble performance, is presented. In addition, the significance of these findings for enhancing diagnostic precision and care delivery is elaborated upon. The goal is to clarify the strengths and weaknesses of the proposed approaches through a detailed examination of each one.

Pairwise associations between the encoded features in the dataset are shown in Figure 4. Each scatter plot is a comparison of two characteristics and can reveal hidden relationships or patterns. Each feature's distribution is depicted by the diagonal line, and the scatter plots show how the features interact with one another. Table 3 shows that the age distribution is presented as a range since the original visual did not provide exact counts for each age but highlighted the distribution. The data indicate a range from 21 to 89 years, with varied infection status.

Table 3. Feature distribution of dataset.

Feature	Category	Count (No Infection)	Count (Infection)
Sex	Female	25	20
	Male	25	28
Smoking History	No	25	30
	Yes	25	15
Family History	Yes	20	30
	No	25	25
Alcohol Consumption	Yes	25	30
	No	25	25
Abdominal Pain	No	25	20
	Yes	25	20
Nausea	Yes	30	20
	No	25	25
Vomiting	No	250	240
	Yes	250	245
Weight Loss	Yes	250	240
	No	250	245
Fatigue	No	250	245
	Yes	250	245
Difficulty Swallowing	No	300	200
	Yes	50	300
Bloating	Yes	150	400
	No	350	150
Age	21–89	Various	Various

4.1. Performance of QKNN

Figure 5 depicts the K-Nearest Neighbors (KNN) model's decision boundary. This graph shows how the KNN model assigns classes to points in the feature space. The majority class of the K-Nearest Neighbors defines the decision border that divides the classes. The papered category is depicted in this map by a corresponding color.



Figure 5. KNN decision boundary.

4.2. Performance of QLR

Figure 6 visually represents the decision boundary of the Logistic Regression (LR) model. This boundary illustrates how the LR model separates instances belonging to different classes based on the logistic regression function's outcome. The visualization helps us understand the LR model's classification behavior in the feature space.



Figure 6. LR decision boundary.

4.3. Performance of QNB

The LR model's cutoff point is depicted in Figure 7 for your convenience. The result of the logistic regression algorithm is used to determine the decision boundary, which in turn defines the classes. This graphic demonstrates how the LR model assigns predicted classes to locations in the feature space by coloring them accordingly.

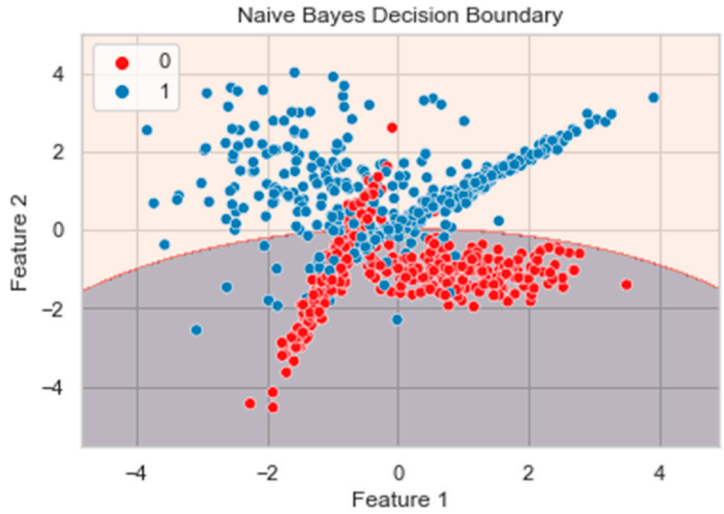

Figure 7. NB decision boundary.

4.4. HeliEns Model Performance

The HeliEns model's decision boundary is depicted in Figure 8. The decision border illustrates the ensemble model's (a model that integrates the predictions of numerous separate models) collective categorization behavior. The ensemble's decision-making process is visualized using colored patches that represent the expected classes.

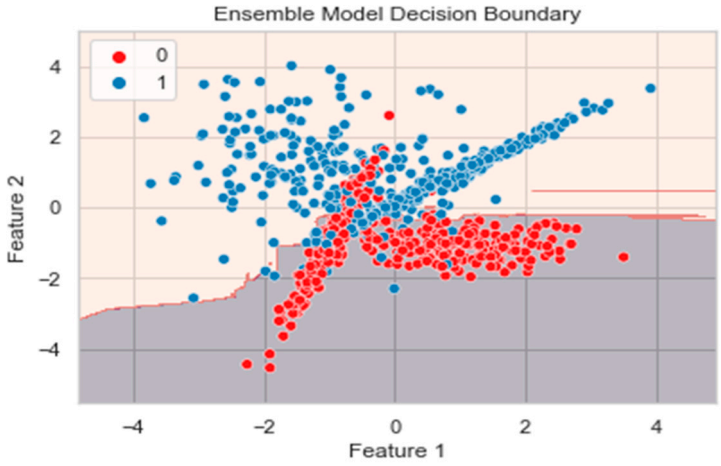

Figure 8. Ensemble model decision boundary.

4.5. Comparative Analysis

Here, the HeliEns model is evaluated against the results of various machine learning models. The purpose of this comparison is to shed light on the benefits and drawbacks of several methods for the rapid detection of H. infection. A close examination is taken of the most important numbers, followed by a discussion of their significance.

The essential performance measures of accuracy, precision, recall, and F1-score for each model are presented in a tabular manner to permit a straightforward comparison in Tables 4 and 5.

Table 4. Performance metrics comparison.

Model	Accuracy	Precision	Recall	F1-Score
HeliEns Model	0.94	0.97	0.92	0.94
KNN Model	0.88	0.88	0.88	0.88
Logistic Regression	0.88	0.91	0.84	0.87
Naive Bayes	0.88	0.91	0.84	0.87

Table 5. Confusion matrix for all matrices.

Model	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)
QKNN	96	5	8	91
QLR	89	12	12	87
QNB	93	8	16	83
Ensemble Model	98	3	8	91

Each cell in the table represents the counts for the respective outcomes in the confusion matrices of the models: True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP). These metrics are crucial for evaluating the performance of each model in terms of correctly and incorrectly classified instances.

High levels of accuracy, precision, recall, and F1-score characterize the HeliEns model's outstanding performance. This finding demonstrates the value of an ensemble method, which pools the best features of various models to boost diagnostic precision and consistency. The exceptional success can be attributed to the ensemble model's capability to both capture complicated patterns and alleviate the limitations of individual models.

Similar to the HeliEns model, the performance metrics for the K-Nearest Neighbors (KNN) and Naive Bayes (NB) models are lower. While these models provide acceptable

performance, they may be unable to fully capture the complex relationships in the data, resulting in less precise and more evenly distributed measurements.

The accuracy, precision, and F1-Score of Logistic Regression (LR) are on par with other leading methods. However, its relatively low recall suggests that it may have trouble reliably capturing true positives. This indicates that the LR model could use more fine-tuning in order to improve recall and attain a more well-rounded performance.

The HeliEns model is shown to perform better than the other options in this comparison. By pooling the results from multiple models, an improved method of diagnosing H. infection infection is created. The findings highlight the importance of ensemble approaches in healthcare applications, where the accuracy of diagnosis and treatment depends on high levels of precision and recall.

The HeliEns model stands out with exceptional performance across all key metrics, showcasing its ability to achieve high levels of accuracy, precision, recall, and F1-score. This robust performance demonstrates the effectiveness of the ensemble approach, which leverages the strengths of individual models to enhance diagnostic precision and consistency. The ensemble model's success can be attributed to its capacity to capture intricate patterns and mitigate the limitations of individual models. Similarly, the KNN and Naive Bayes models exhibit acceptable performance, although their metrics are marginally lower than those of the HeliEns model. These models, while offering credible results, might lack the capacity to fully capture intricate data relationships, potentially resulting in less precise and evenly distributed measurements. Logistic Regression (LR) presents comparable accuracy and F1-score values to the leading models; however, its relatively lower recall suggests challenges in consistently capturing true positives. This indicates the need for further fine-tuning of the LR model to enhance its recall and achieve a more balanced overall performance. Despite the overall strong performance of the ensemble model, it is essential to conduct more rigorous testing and evaluation to assess its generalizability and robustness across diverse datasets. Additional testing on independent datasets, cross-validation, and sensitivity analysis could provide more insights into the model's reliability and applicability in real-world scenarios. In conclusion, the HeliEns model emerges as the most effective method among the evaluated options. The ensemble approach's ability to pool the outputs of multiple models underscores its potential for advancing H. infection infection diagnosis. This analysis underscores the significance of ensemble methodologies in healthcare applications, where the precision and recall of diagnostic outcomes are of paramount importance. Further research and validation will be critical to ensure the model's effectiveness across various clinical settings.

5. Conclusions

This research presents HeliEns, a novel quantum hybrid ensemble learning algorithm designed for the early and accurate diagnosis of *Helicobacter pylori* (*H. pylori*) infection. By integrating Quantum K-Nearest Neighbors (QKNN), Quantum Naive Bayes (QNB), and Quantum Logistic Regression (QLR) models, HeliEns leverages the computational advantages of quantum computing to enhance diagnostic accuracy and efficiency. The performance metrics of the HeliEns model, including an accuracy of 94%, precision of 97%, recall of 92%, and F1-score of 94%, underscore its superiority over traditional methods. However, this research has several limitations and weaknesses that should be acknowledged. First, the implementation of quantum machine learning models is still in its nascent stages and requires specialized quantum hardware that may not be widely accessible. The practical deployment of HeliEns in clinical settings may be hindered by the current limitations of quantum computing infrastructure. Second, the dataset used in this research, while comprehensive, may not fully represent the diverse patient populations encountered in real-world clinical environments. Further validation with larger and more diverse datasets is necessary to ensure the generalizability of the model. Additionally, the complexity of integrating multiple quantum models into an ensemble framework poses challenges in terms of computational resources and scalability. The quantum algorithms employed,

although promising, need further optimization to handle large-scale data more efficiently. Moreover, the interpretability of quantum models remains a challenge, as the underlying quantum processes are inherently complex and may not be easily understood by healthcare professionals. In summary, while HeliEns demonstrates significant potential in enhancing the diagnosis of *H. pylori* infection, its practical implementation requires overcoming the current limitations of quantum computing technology and ensuring robust validation across diverse clinical datasets. Future research should focus on addressing these challenges to realize the full potential of quantum machine learning in healthcare diagnostics.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this research are openly available in https://www.kaggle.com/datasets/kanchana1990/bacteria-dataset.

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this research paper. The research was conducted in an unbiased manner, and there are no financial or personal relationships that could have influenced the findings or interpretations presented herein.

References

1. Hoogenboom, S.A.; Bagci, U.; Wallace, M.B. Artificial intelligence in gastroenterology. The current state of play and the potential. How will it affect our practice and when? *Tech. Innov. Gastrointest. Endosc.* **2020**, *22*, 42–47. [CrossRef]
2. Klein, S.; Gildenblat, J.; Ihle, M.A.; Merkelbach-Bruse, S.; Noh, K.-W.; Peifer, M.; Quaas, A.; Büttner, R. Deep learning for sensitive detection of *Helicobacter pylori* in gastric biopsies. *BMC Gastroenterol.* **2020**, *20*, 417. [CrossRef]
3. Yang, H.; Guan, L.; Hu, B. Review Article Detection and Treatment of *Helicobacter pylori*: Problems and Advances. *Gastroenterol. Res. Pract.* **2022**, *2022*, 4710964. [CrossRef] [PubMed]
4. Chey, W.D.; Leontiadis, G.I.; Howden, C.W.; Moss, S.F. ACG Clinical Guideline: Treatment of *Helicobacter pylori* Infection. *Am. J. Gastroenterol.* **2017**, *112*, 212–239. [CrossRef]
5. de Arce, E.P.; Quera, R.; Núñez, P.; Araya, R. Role of capsule endoscopy in inflammatory bowel disease: Anything new? *Artif. Intell. Gastrointest. Endosc.* **2021**, *2*, 136–148. [CrossRef]
6. Nakashima, H.; Kawahira, H.; Kawachi, H.; Sakaki, N. Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: A single-center prospective study. *Ann. Gastroenterol.* **2018**, *31*, 462–468. [CrossRef]
7. Ishihara, K.; Ogawa, T.; Haseyama, M. Instructions for use *Helicobacter pylori* Infection Detection from Gastric X-ray Images Based on Feature Fusion and Decision Fusion. *Comput. Biol. Med.* **2017**, *84*, 69–78. [CrossRef]
8. Yang, H.; Hu, B. Diagnosis of *Helicobacter pylori* Infection and Recent Advances. *Diagnostics* **2021**, *11*, 1305. [CrossRef]
9. Pecere, S.; Milluzzo, S.M.; Esposito, G.; Dilaghi, E.; Telese, A.; Eusebi, L.H. Applications of Artificial Intelligence for the Diagnosis of Gastrointestinal Diseases. *Diagnostics* **2021**, *11*, 1575. [CrossRef]
10. Yacob, Y.M.; Alquran, H.; Mustafa, W.A.; Alsalatie, M.; Sakim, H.A.; Lola, M.S. *H. pylori* Related Atrophic Gastritis Detection Using Enhanced Convolution Neural Network (CNN) Learner. *Diagnostics* **2023**, *13*, 336. [CrossRef]
11. Mohan, B.P.; Khan, S.R.; Kassab, L.L.; Ponnada, S.; Mohy-Ud-Din, N.; Chandan, S.; Dulai, P.S.; Kochhar, G.S. Convolutional neural networks in the computer-aided diagnosis of *Helicobacter pylori* infection and non-causal comparison to physician endoscopists: A systematic review with meta-analysis. *Ann. Gastroenterol.* **2021**, *34*, 20. [CrossRef]
12. Zhao, P.; Han, K.; Yao, R.; Ren, C.; Du, X. Application Status and Prospects of Artificial Intelligence in Peptic Ulcers. *Front. Surg.* **2022**, *9*, 894775. [CrossRef] [PubMed]
13. Tran, V.; Saad, T.; Tesfaye, M.; Walelign, S.; Wordofa, M.; Abera, D.; Desta, K.; Tsegaye, A.; Ay, A.; Taye, B. *Helicobacter pylori* (*H. pylori*) risk factor analysis and prevalence prediction: A machine learning-based approach. *BMC Infect. Dis.* **2022**, *22*, 655. [CrossRef] [PubMed]
14. Erjaei, M.H. *Helicobacter pylori* Detection Using Machine Learning Algorithm. *Adv. Pract. Nurs.* **2017**, *2*, 1000140. [CrossRef]
15. Shichijo, S.; Nomura, S.; Aoyama, K.; Nishikawa, Y.; Miura, M.; Shinagawa, T.; Takiyama, H.; Tanimoto, T.; Ishihara, S.; Matsuo, K.; et al. Application of Convolutional Neural Networks in the Diagnosis of *Helicobacter pylori* Infection Based on Endoscopic Images. *EBioMedicine* **2017**, *25*, 106–111. [CrossRef] [PubMed]
16. Hu, H.; Gong, L.; Dong, D.; Zhu, L.; Wang, M.; He, J.; Shu, L.; Cai, Y.; Cai, S.; Su, W.; et al. Identifying early gastric cancer under magnifying narrow-band images with deep learning: A multicenter study. *Gastrointest. Endosc.* **2021**, *93*, 1333–1341. [CrossRef]
17. Bang, C.S.; Lee, J.J.; Baik, G.H. Artificial Intelligence for the Prediction of *Helicobacter pylori* Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. *J. Med. Internet Res.* **2020**, *22*, e21983. [CrossRef]
18. Itoh, T.; Kawahira, H.; Nakashima, H.; Yata, N. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endosc. Int. Open* **2018**, *6*, E139–E144. [CrossRef]
19. Zhao, Y.; Hu, B.; Wang, Y.; Yin, X.; Jiang, Y.; Zhu, X. Identification of gastric cancer with convolutional neural networks: A systematic review. *Multimed. Tools Appl.* **2022**, *81*, 11717–11736. [CrossRef]

20. Bordin, D.S.; Voynovan, I.N.; Andreev, D.N. Current *Helicobacter pylori*. *Diagnostics* **2021**, *11*, 1458. [CrossRef]
21. Lu, B.; Li, M. *Helicobacter pylori* eradication for preventing gastric cancer. *World J. Gastroenterol. WJG* **2014**, *20*, 5660. [CrossRef] [PubMed]
22. Kiryu, S.; Akai, H.; Yasaka, K. Deep learning application in the oesophageal endoscopy. *J. Med. Artif. Intell.* **2019**, *2*, 22. [CrossRef]
23. Ghaemi, H.; Kahani, M. Question Classification using Ensemble Classifiers. *Signal Data Process.* **2016**, *13*, 99–112. [CrossRef]
24. Tang, J.W.; Li, F.; Liu, X.; Wang, J.T.; Xiong, X.S.; Lu, X.Y.; Zhang, X.Y.; Si, Y.T.; Umar, Z.; Tay, A.C.; et al. Detection of *Helicobacter pylori* infection in human gastric fluid through surface-enhanced Raman spectroscopy coupled with machine learning algorithms. *Lab. Investig.* **2024**, *104*, 100310. [CrossRef]
25. Zhang, M.; Liu, F.; Shi, F.; Chen, H.; Hu, Y.; Sun, H.; Qi, H.; Xiong, W.; Deng, C.; Sun, N. High-throughput detection allied with machine learning for precise monitoring of significant serum metabolic changes in *Helicobacter pylori* infection. *Talanta* **2024**, *269*, 125483. [CrossRef] [PubMed]
26. Awan, R.E.; Zainab, S.; Yousuf, F.J.; Mughal, S. AI-driven drug discovery: Exploring Abaucin as a promising treatment against multidrug—Resistant Acinetobacter baumannii. *Health Sci. Rep.* **2024**, *7*, e2150. [CrossRef]
27. Jiang, F.; Lui, T.K.; Ju, C.; Guo, C.G.; Cheung, K.S.; Lau, W.C.; Leung, W.K. Machine learning models in predicting failure of *Helicobacter pylori* treatment: A two country validation study. *Helicobacter* **2024**, *29*, e13051. [CrossRef]
28. Tian, C.; Hao, D.; Ma, M.; Zhuang, J.; Mu, Y.; Zhang, Z.; Zhao, X.; Lu, Y.; Zuo, X.; Li, W. Graded diagnosis of *Helicobacter pylori* infection using hyperspectral images of gastric juice. *J. Biophotonics* **2024**, *17*, e202300254. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.