FUZZY SEA HORSE OPTIMIZATION ALGORITHM (FSHOA) AND QUINE MCCLUSKEY ENSEMBLE CLASSIFIER (QMEC) FOR HEART DISEASE PREDICTION

V. Manimozhi ,
Research Scholar,
Department of Computer Science,
RVS College of Arts and Science
(Affiliated to Bharathiar University),
Coimbatore,
Tamilnadu, India

Dr. K. Chitra
Assistant Professor,
Department of Computer Science,
RVS College of Arts and Science
(Affiliated to Bharathiar University),
Coimbatore,
Tamilnadu, India

ABSTRACT

Cardiovascular disease is the primary reason for mortality worldwide, responsible for around a third of all deaths. To assist medical professionals in quickly identifying and diagnosing patients, numerous machine learning and data mining techniques are utilized to predict the disease. Many researchers have developed various models to boost the efficiency of these predictions. Feature selection and extraction techniques are utilized to remove unnecessary features from the dataset, thereby reducing computation time and increasing the efficiency of the models. In this paper, fuzzy membership function is introduced to enhance SHOA, which improves the search ability and convergence stability of standard model. It removes the features from the dataset, greatly improves the speed of leader position update, improves the optimization speed, improves the search accuracy, and improves the local optimal solution. SHO algorithm simulates the movement, predation, and reproduction of sea horse for optimal selection of features from the dataset. From the selected features, QMEC model utilizes an ensemble of seven models including Random Forest (RF), k-Nearest Neighbour (kNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). It performs exceptionally well on binary class datasets. Ensemble model is applied to the dataset, utilizing the Quine McCluskey minimum Boolean equation built with an 80:20 train-to-test ratio. Results are measured using the evaluation metrics like precision, recall/sensitivity, specificity, f1-score and accuracy.

Keywords: Fuzzy, Sea Horse, Optimization, Quine, Mccluskey, Ensemble, Classifier, Heart Disease, Prediction, Support Vector Machine

Section 1: Introduction

Traditional machine learning algorithms like SVM and ANN, as well as more contemporary statistical methods like logistic regression, decision trees, and k-nearest neighbours, have been prominent [1]. The use of these methods has predicted heart disease. High dimensionality and nonlinear interactions make medical data complicated, yet current methods, while accurate, are insufficient for managing information [2]. Furthermore, many algorithms struggle with feature selection, a vital step in identifying the most essential cardiac risk variables. Traditional Boolean function simplification methods like the Quine-

McCluskey method achieve logic minimisation [3,4]. Since the risk variables interact in complex ways and heart disease has a large diversity of symptoms, ordinary classifiers often have problems adapting to data which is uncertain [5]. Though fuzzy logic-based techniques manage data uncertainties effectively, they are not very scalable and optimizable when used inside sophisticated frameworks [6]. As a result of their low recall-precision quotient, these methods are usually condemned to make clinical forecasts with very low accuracy. The problem involves high computational requirements and in addition there is no automatic adaptation of schemas so that it can change the parameters dynamically which adds on to the problem even more [7]. Such models which aim to forecast the occurrence of heart diseases are often hampered by this phenomenon leading to failure in most instances.

The FSHOA and Quine-McCluskey Ensemble Classifier are also being employed that so as to assist in heart disease prediction through the integration and optimization of the aforementioned techniques in handling complexity associated with medical data [8]. Since heart disease datasets are complicated, and nonlinear, doing a real time parameter change is impossible for the FSHOA. It is noted that FSHOA was designed with the intention of solving problems under classification by optimising in uncertainty [9]. This holds also for the processing of high-dimensional medical data, where speed of convergence as well as escape from local minima is of the essence. However, it is good for the resolution of uncertainties of the predictions of the algorithm, yet it may cause problems in accepting it by clinicians and hence in using it in practice. For the prediction of cardiac disease, a Quine-McCluskey technique that addresses complexities of Boolean expressions in ensemble classifiers has limitations [10]. In large medical databases it is a challenge to reduce the features and the complicated relationships without losing diagnostic power. Although enhancement of predictive accuracy is the trademark of ensemble classifiers, the challenge comes when FSHOA and Quine-McCluskey have to work together which is not easy [11]. Bringing up the issue of computational cost and overfitting concerning the ensemble has to be addressed without fail. The issue of finding the right degree of complexity which really provides an accurate model is another issue. Even if models are accurate, they are overly complex and lack transparency for practical purposes. Striving towards this equilibrium is, however, not crisis free. Such traits contribute to the instability, reliability and scalability of models for predicting heart diseases presenting such attributes [12].

Though Quine-McCluskey Ensemble Classifier and FSHOA may predict cardiac disease, these algorithms can be overcome in several ways. The problem at hand may be treated in other ways, for example, FSHOA can be optimized by using its adaptive parameters, thus, the program must be able the adapt to the complex of data in real time. Composite systems using FSHOA, particle swarm optimisation, and evolutionary algorithms can improve convergence and prevent local optimums. PCA can be used with the Quine-McCluskey approach to reduce big datasets while keeping critical features [13]. The ensemble's efficiency can improve without sacrificing accuracy or interpretability.

The idea will overcome the problem of using huge, complicated medical datasets with unneeded information to calibrate machine learning models for cardiac disease prediction. Features, convergence, and search are difficult for conventional models. This slows optimisation and gives inaccurate projections. Top priorities include preventing local optimums and boosting processing efficiency and search accuracy. The objective is to combine the FSHOA with the QMEC to overcome these heart disease prediction constraints.

- A FSHOA that efficiently chooses critical qualities from heart disease datasets while improving conventional models' search capacity, convergence stability, and optimisation speed is the goal.
- The QMEC is being integrated to improve binary class dataset heart disease prediction. Seven machine learning models Multilayer Perceptron, Random Forest, Support Vector Machine, and k-Nearest Neighbor will be used.
- Compare the FSHOA and QMEC models' recall/sensitivity, accuracy, specificity, f1-score, and precision to recognised benchmarks.

This section organises the framework of the research study and includes the following: Section II of this work investigates the concepts of fuzzy sea horse optimisation algorithm (FSHOA) and quine MCcluskey system (QMEC) for the prediction of heart diseases. Section III of this dissertation will go into great detail regarding the Fuzzy Sea Horse Optimisation Algorithm (FSHOA). Section IV of the paper includes a thorough evaluation, a comparison to previous methods, and an analysis of the outcomes. A comprehensive analysis of the results that have been assessed is presented in Section V.

Section 2: Literature Survey

Effective management and early identification of cardiovascular diseases (CVDs) require better prediction models. This is because cardiovascular disorders are a global health concern. Recently, researchers have examined various machine learning (ML) strategies to improve heart disease prediction.

Kapila, R et al., [14] using feature selection and principal component analysis, the Quine McCluskey Binary Classifier (QMBC) that was suggested achieves a higher level of accuracy than the approaches that are currently in use. It is comprised of seven different machine learning models for the purpose of predicting cardiac disease.

Kapila, R et al., [15] proposed in this study, the categorisation of heart illness is evaluated using Fruit Fly Optimisation, Particle Swarm Optimisation, and Grey Wolf Optimiser. The results show that GWO achieves the highest performance metrics, particularly in terms of precision and accuracy than the other two.

Naser, M. A et al., [16] with the purpose of enhancing early diagnosis and improving public health outcomes, this review paper examines the application of machine learning in the prediction of cardiovascular disease. It focusses on the selection of features, assessment metrics, and trends.

Sarra, R. R et al., [17] proposed a stacked ensemble classifier (EC) that combines decision trees, support vector machines, and multilayer perceptrons with chi-square feature selection is proposed in this research. The classifier achieves an accuracy of 90.8% for the detection of heart disease, thereby improving both efficiency and affordability.

Using only seven ideal features, Jayaparvathy, R. [18] suggested a grey wolf optimization-based feature selection technique (GWO-FSM) that was integrated with a feature-specific extreme gradient boosting classifier. This method achieved an accuracy of 98.8 percent and good performance.

Imran, M et al., [19] using extensive datasets and sophisticated preprocessing procedures, the Ensemble Stacked Neural Network (ESNN) model achieves an accuracy rate of 95% for the early diagnosis of heart illness. This is accomplished by integrating machine learning and deep learning approaches.

Table 1: Summary of related works

Author	Method	Advantage	Disadvantage
Kapila, R et al. [14]	Quine McCluskey Binary Classifier (QMBC) with feature selection and principal component analysis (PCA) using seven machine learning models.	Higher accuracy compared to traditional approaches due to optimized binary classification and model combination.	Computationally intensive due to PCA and multiple model integrations.
Kapila, R et al. [15]	Categorization of heart illness using Fruit Fly Optimisation (FFO), Particle Swarm Optimisation (PSO), and Grey Wolf Optimiser (GWO).	GWO achieves the highest precision and accuracy, improving model performance.	PSO and FFO have lower performance metrics compared to GWO, reducing consistency across models.
Naser, M. A et al. [16]	Review paper focusing on the use of machine learning models for predicting cardiovascular disease, with emphasis on feature selection, assessment metrics, and trends.	Provides a comprehensive analysis of machine learning's role in early cardiovascular disease diagnosis.	Does not propose or test specific new models; focuses on literature review and trends rather than experimental work.
Sarra, R. R et al. [17]	Stacked ensemble classifier (EC) combining decision trees, SVMs, and MLPs with chi-square feature selection for	Achieves 90.8% accuracy with a cost-effective and efficient ensemble classification model.	Lower accuracy compared to more advanced optimization techniques (e.g., GWO or ESNN).

	heart disease detection.		
Jayaparvathy, R. [18]	Grey Wolf Optimization-based Feature Selection Method (GWO-FSM) integrated with feature-specific extreme gradient boosting classifier.	High accuracy of 98.8% with the use of GWO for optimal feature selection and extreme gradient boosting.	Limited to a specific set of seven ideal features, which may not generalize across other datasets.
Imran, M et al. [19]	Ensemble Stacked Neural Network (ESNN) model utilizing extensive datasets and sophisticated preprocessing for early heart illness diagnosis.	Achieves 95% accuracy by combining machine learning and deep learning for robust prediction.	Requires extensive preprocessing and large datasets, making it computationally expensive.

The FSHOA has led the race to enhance cardiac disease diagnostic precision and efficacy by exceeding competitors. All of these advances demonstrate the importance of feature selection and ML algorithms for cardiovascular disease. According to them, diagnostic tools and public health results improve.

Section 3: Proposed Method

Still accounting for almost a third of all deaths worldwide, CVD is by far the biggest killer. Predicting CVD using ML and data mining approaches has become more important because to the increasing need for early and accurate diagnosis. Researchers have created a number of models to improve prediction efficiency to simplify datasets and reduce computing time. These models often use feature selection and extraction approaches. In this proposed method, a fuzzy membership function is introduced to enhance the SHOA, significantly improving its search accuracy and convergence stability. By optimizing the SHO algorithm,

unnecessary features are removed, accelerating leader position updates and improving the algorithm's ability to avoid local optima. Predicting the occurrence of cardiovascular disease using the given variables is done using the QMEC model ensemble. Precision, accuracy, and recall are common metrics used to evaluate the results.

Contribution 1: Enhancement of Feature Selection with Fuzzy SHOA

A fuzzy membership function forms the basis of the suggested solution. The traditional SHOA is its target for improvement. In doing so, it overcomes the typical problems of local optima and sluggish optimisation speed in conventional models, making the algorithm better at searching and achieving convergence stability. By optimizing the SHO algorithm, the system effectively removes redundant features from the dataset, speeding up the leader position update. This improvement improves the predictive model's feature selection by improving the algorithm's search accuracy. This development ensures that the machine learning model takes into account more important information. It sets the groundwork for better cardiovascular disease prediction.

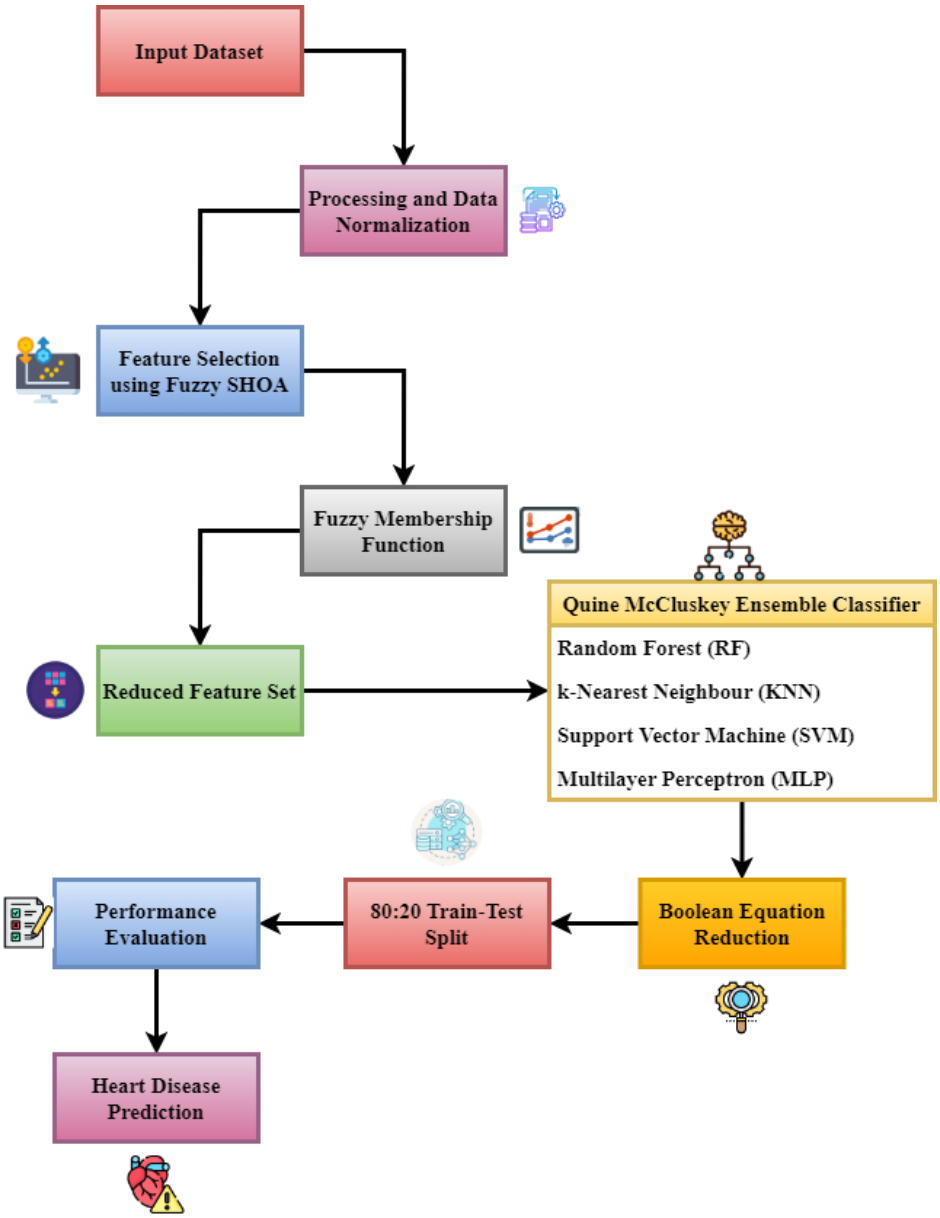

Figure 1: FSHOA and QMEC for Heart Disease Detection

Figure 1 illustrates the proposed methodology for forecasting the incidence of heart disease, including many important phases. The procedure starts with an input dataset that is processed and normalised to ensure data quality. Fuzzy SHO Algorithm is used for feature selection. This method minimises calculation time without sacrificing significant characteristics by optimising the feature set through simulation of sea horse motions. Following this, the SHO's search capability and stability are improved using a Fuzzy Membership Function.

The selected features form a Reduced Feature Set, which is fed into the Quine McCluskey Ensemble Classifier. This classifier ensemble consists of many models, such as Multilayer Perceptron, Random Forest, Support Vector Machine, and k-Nearest Neighbour. For the purpose of becoming efficient, Boolean Equation Reduction is used. Training and testing the model are divided 80:20. Performance metrics are used to assess the accuracy, recall, and precision of the heart disease forecasts.

$$|V(E' - Hj')| = P\big(r - Tn'(M - dw)\big) \quad (1)$$

Equation 1, $V(E' - Hj')$ shows the difference between the expected $M - dw$ and observed values P, while $r - Tn'$ represents the likelihood of a proper classification. The heart disease identification model's prediction performance, this equation shows that feature optimization and ensemble learning are key.

$$|H(Ny' - 2w)| = Mb * V' - 3T(m - Fl') \quad (2)$$

The influence of model complexity $(m - Fl')$, feature interactions $H(Ny' - 2w)$, and data manipulations on predictive performance are shown by $Mb * V' - 3T$, and the efficacy of the hybrid model is given by the equation. Parameter optimization and model dynamics using the ensemble technique, as shown by equation 2.

$$|E(Tr' - sa)| = \big(Ty' - prj(2 - l'gf)\big) \quad (3)$$

The predicted $2 - l'gf$ and actual values are shown by the equation 3, $Ty' - prj$, and the projected true values after correcting for the learned features and their relationships are shown by $E(Tr' - sa)$. This equation shows how important it is to refine model parameters and feature connections.

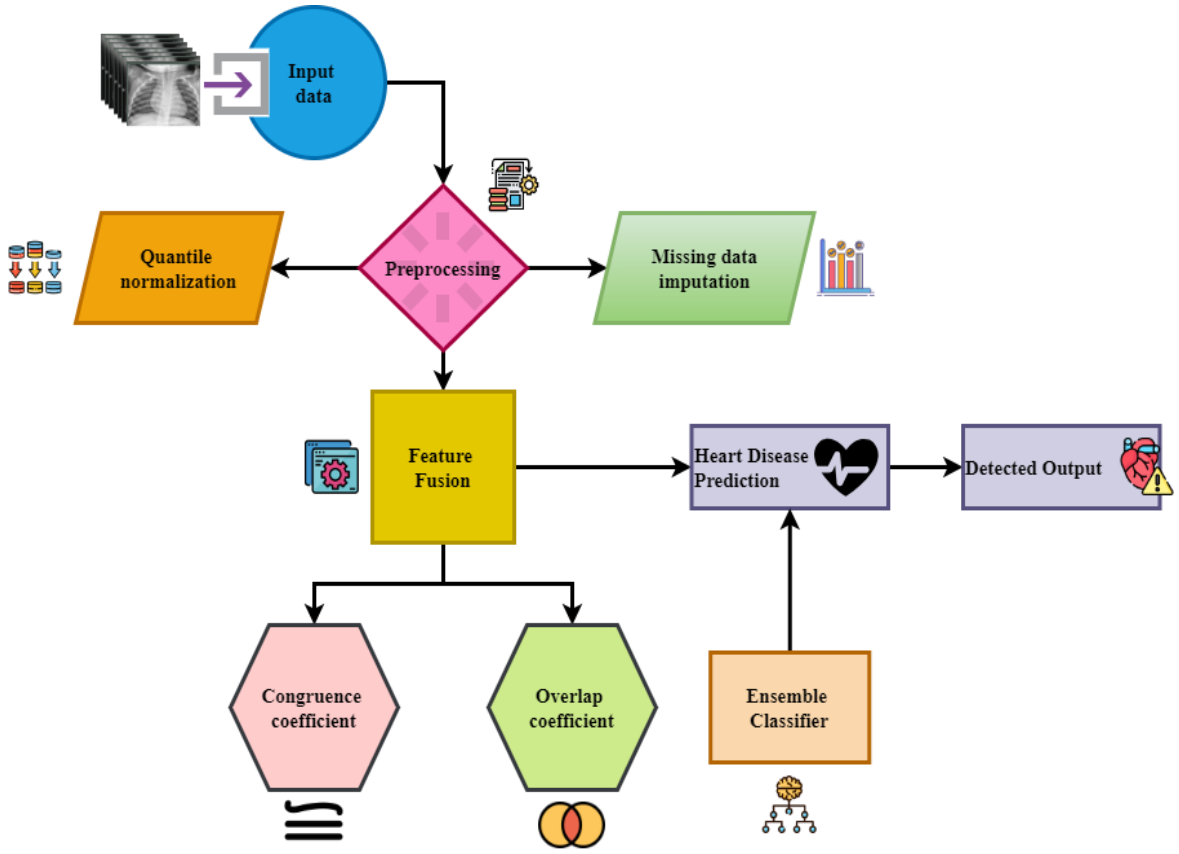

Figure 2: Prediction using Fuzzy Sea Horse Optimization Algorithm

The process flow of the suggested approach for cardiac disease prediction is shown in Figure 2. Quantile normalisation and pre-processing are the first steps in the process that include standardising characteristics in the incoming data. Missing data imputation is used to fill them in when incomplete records are detected. The next stage is feature fusion, which involves merging several features and then utilising fuzzy-enhanced SHOA to remove any characteristics that are unnecessary or redundant. Model correctness may be improved by using the congruence and overlap coefficients to hone the feature selection process. The forecasting of heart disease incidence relies on a collection of machine learning techniques. The system utilises the input data to generate many outputs, including a prediction of the likelihood of heart disease. This strategy improves prediction accuracy and optimises feature selection, ensuring efficient performance in medical diagnostics.

$$B - (< \partial - \forall'(Tx - y'z) >) = \left(V_f(m' - nr)\right) \quad (4)$$

The baseline performance measure is denoted by equation 4, B, and the departure from predicted outcomes owing to $V_f(m' - nr)$ in feature mappings and model predictions is

denoted by $\partial - \forall'(Tx - y'z)$. The impact of feature variance and model modification suggests that improving model performance relies heavily on optimizing feature selection.

$$H(t' - rq) = P'(qw - kn'') * M\big(T(n - v's)\big) \quad (5)$$

Equation 5, $H(t' - rq)$ shows the output for health prediction $P'(qw - kn'')$ denotes the probabilistic evaluation of feature contributions and their interactions, and $M\big(T(n - v's)\big)$ shows optimization parameters modulate these interactions. The significance of features and optimization dynamics must be balanced, as this equation shows.

Algorithm 1: Fuzzy-Enhanced SHOA and QMEC Ensemble Model

Step 1	**Preprocessing the Dataset** *Load dataset D* *Normalize features in F* *Split dataset into training (T_r) and testing (T_t) sets (80:20)*
Step 2	**Feature Selection Using Fuzzy-Enhanced SHOA** *Initialize sea horse population P with random feature subsets* *Initialize fuzzy membership values for each feature* *Set iteration i = 1* *while i ≤ N:* *for each sea horse in P:* *Compute fitness score f(F) using model accuracy* *if sea horse is near the leader:* *Adjust position slightly* *else:* *Move towards the leader dynamically* *if fitness score improves:* *Keep new feature subset* *else:* *Apply fuzzy update to adjust leader position* *if sea horse has best fitness:* *Clone top-performing sea horses* *else:*

	Generate new feature subsets *i = i + 1*
Step 3	**Classification Using QMEC Ensemble Model** *Train ensemble model QMEC using optimized feature set F'* *Apply Random Forest, kNN, SVM, and MLP classifiers* *for each classifier in QMEC:* *Train model on T_r* *Predict on T_t* *Compute evaluation metrics (Accuracy, Precision, Recall, F1-score)* *if Accuracy > Threshold:* *Print "Model is optimal for cardiovascular disease prediction"* *else:* *Print "Re-evaluate feature selection or model parameters"*

The algorithm 1 enhances cardiovascular disease prediction using a Fuzzy-Enhanced SHOA and QMEC Ensemble Model. First, the dataset is preprocessed and split into an 80:20 train-test ratio. SHOA optimizes feature selection by simulating sea horse movement, predation, and reproduction, with fuzzy logic improving search accuracy and convergence. The selected features are used in the QMEC ensemble model, which integrates classifiers like Random Forest, kNN, SVM, and MLP. Models are trained and evaluated using accuracy, precision, recall, and F1-score. If accuracy surpasses a threshold, predictions are deemed optimal; otherwise, feature selection and model parameters are re-evaluated for improvement.

Contribution 2: Integration of QMEC Ensemble Model for Prediction

As part of the method's implementation, the fuzzy-enhanced SHOA is used to extract the most important characteristics from a dataset including information on cardiovascular illness. Once features are selected, the ensemble-based QMEC model, which incorporates seven classifiers like Random Forest, k-Nearest Neighbors, Support Vector Machine, and Multilayer Perceptron, is applied. To improve the ensemble model's classification process, the Quine McCluskey minimal Boolean equation approach is used. A well-validated model

requires an 80:20 split of the dataset between training and testing. Applying a portion of the available data, the algorithm successfully predicts the incidence of cardiovascular disease.

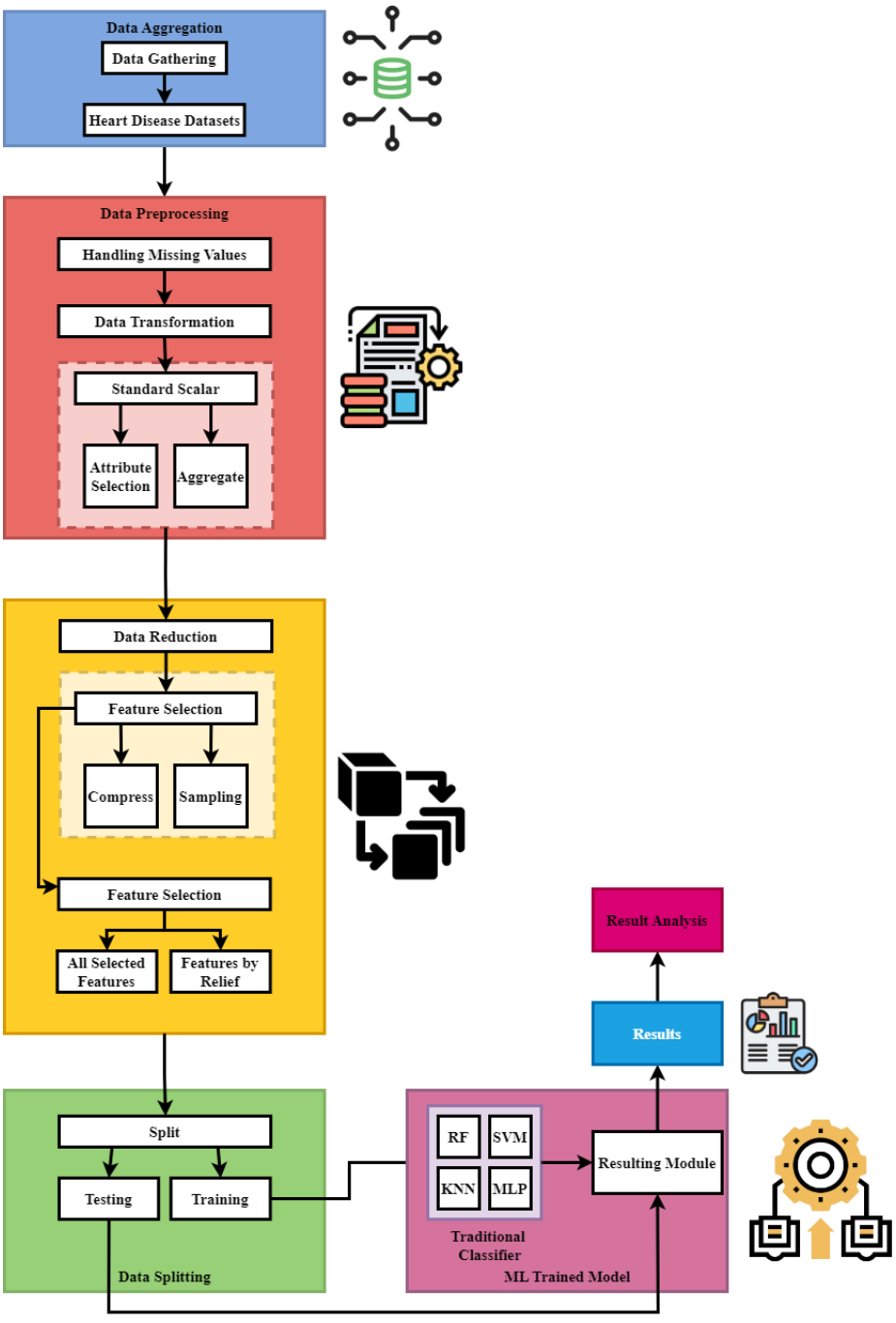

Figure 3: Training and testing the heart disease using classifiers

The suggested approach to predicting cardiac illness using ML is shown in Figure 3. The first step is to collect datasets related to heart disorders. The second step is to aggregate the

data. Data preparation for analysis include aggregation, modifications (such standard scale), attribute selection, and filling in missing data.

Feature selection approaches are used in data reduction, using compression and sampling techniques. Features are optimised by the Relief algorithm to determine the most important ones.

The split phase divides the data into training and testing sets. The traditional classifiers such as Random Forest (RF), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) are used to train models on the selected features.

The result module integrates model outputs. It is used throughout the analysis phase to calculate performance metrics like accuracy, precision, recall, specificity, and F1-score. This method prioritises efficient feature selection using ensemble learning to improve the prediction of heart disease.

$$B\big(Re(n' - tr)\big) = Xz' - M\big(Re(q - bj'')\big) \quad (6)$$

The baseline performance is shown by the equation 6, $B\big(Re(n' - tr)\big)$, which is calculated by regressing $Re(q - bj'')$ the predicted values against the training data. The adjusted outputs, which include model improvements and feature interactions, are represented by $Xz' - M$. This equation emphasizes the need to consistently evaluate and guarantee precise forecasts of cardiac disease.

$$M * (V - |N(\forall - \partial' A)|) = B(Xz' - mw'') \quad (7)$$

Based on the output variance V and model improvements $N(\forall - \partial' A)$, the baseline adjustments are denoted by $B(Xz' - mw'')$, and the adjusted measure of performance after faults is represented by the equation 7, M. Model performance must be balanced with residual reliability and precision of heart disease diagnosis, as shown in this equation.

$$M * Z(x - cfd'') = S(x' - ewp'') * Lop'' \quad (8)$$

To account for crucial feature deviations M, the updated output is represented by the equation Lop'', and the performance metrics are $S(x' - ewp'')$ based on projected values about actual results, scaled by $Z(x - cfd'')$, which indicates the model's optimization level.

Equation 8 highlights the significance of improving feature contributions to guarantee accurate and reliable heart disease categorization.

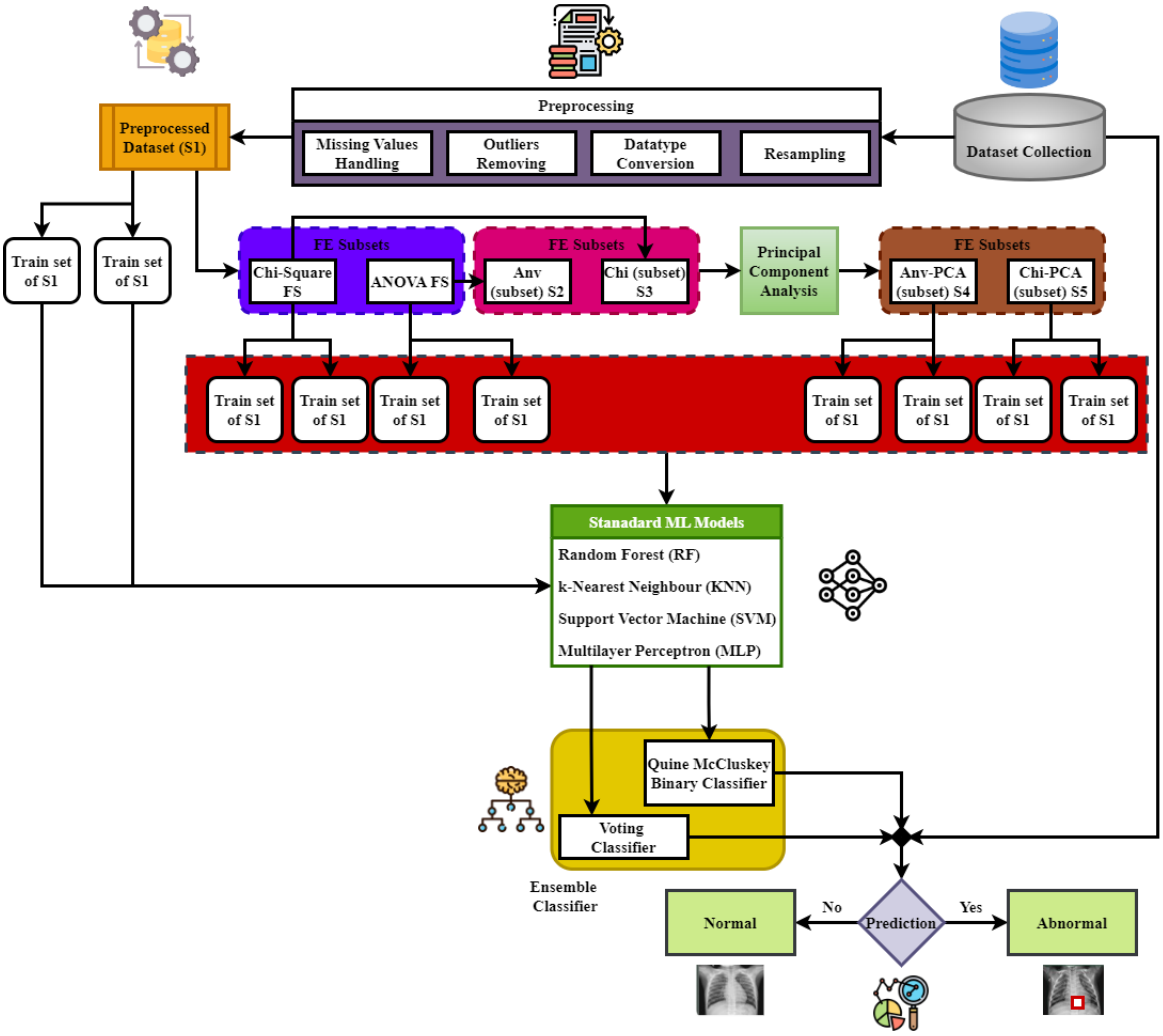

Figure 4: Cardiovascular Disease Prognosis Using QMEC: A Flow Diagram

Figure 4 shows the preprocessed dataset is the starting point of the suggested technique. Pre-processing steps include dataset resampling, outlier elimination, format transformation, and missing value imputation. Many statistical methods are used to produce unique subsets of characteristics. Forecasting cardiac disease using these subgroups is important as it improves predictive accuracy and reduces computational burden.

The standard procedure for training ML models applies each feature subset to several training sets of the preprocessed data. The proposed method incorporates a voting classifier into the Quine McCluskey binary classifier to improve the accuracy of predictions. The

method forecasts heart illness. Accuracy, precision, and recall are some of the metrics used to measure performance by producing a normal or negative categorisation.

$$T \leq M\big(\delta - \varepsilon\alpha(\gamma + 4\Delta)\big) = \tau'(\sigma\rho + \mu'') \quad (9)$$

Emphasizes optimizing Adjustments performed τ' based on feature T interactions and optimization parameters are quantified by $M\big(\delta - \varepsilon\alpha(\gamma + 4\Delta)\big)$, whereas the desired performance measure is represented by equation 9, $\sigma\rho + \mu''$. This equation emphasises the need of optimising model parameters and improving feature selection.

$$V(n' - rt)| \geq K\big(v(n' - wq)\big) * M[v(n2 - Tm)] \quad (10)$$

The variation in the results of prediction is represented by the equation $V(n' - rt)|$, a scaling factor determined by the interaction is denoted by $K\big(v(n' - wq)\big)$, and the effect of the model's modifications on the predictions is reflected by $M[v(n2 - Tm)]$. The importance of the predictive model maintaining a certain accuracy level is shown by equation 10.

Contribution 3: Performance Evaluation Using Evaluation Metrics

A number of well-established evaluation standards, such as accuracy, F1-score, recall (sensitivity), and precision, are used to assess the suggested technique. The Quine McCluskey minimum Boolean equation is used to compute the optimal classification boundaries for the ensemble model. By testing the model on the 20% validation dataset, the evaluation process quantifies how well the proposed method predicts cardiovascular disease.

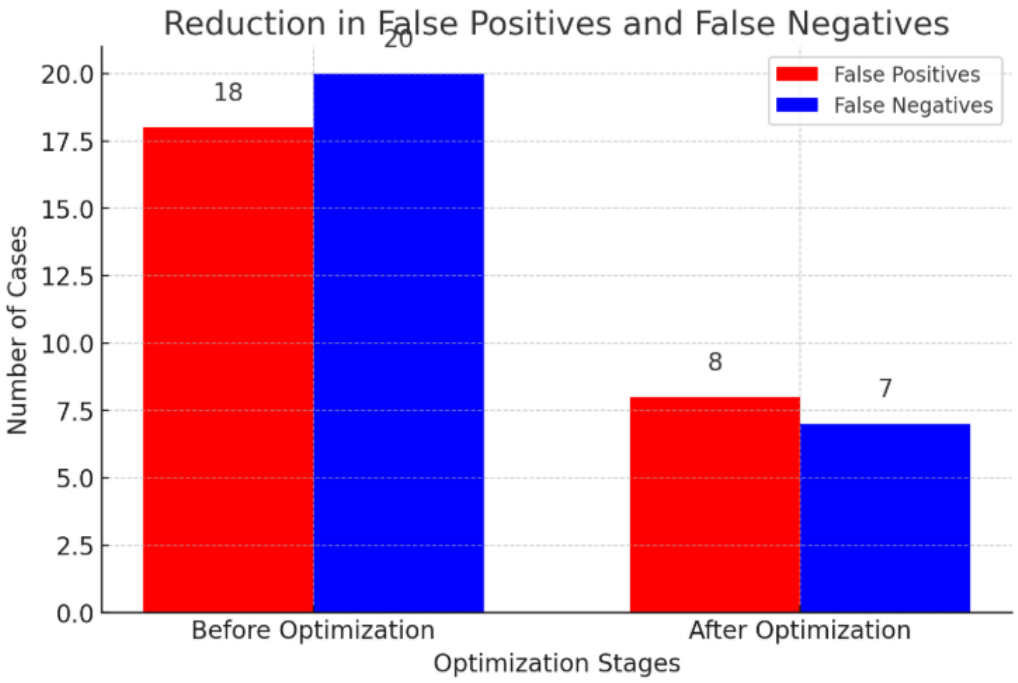

Figure 5: Analysis of false positives and false negatives

Through precise equations and evaluation metrics, the fuzzy-enhanced SHOA and QMEC ensemble model demonstrate improved performance in reducing false positives and false negatives, ensuring reliable predictions is explained in figure 5. Results show that the suggested approach does improve the accuracy of cardiovascular disease prediction.

$$Sin\,T(Y' - pr) = |V(n2 - cv'')| * Tz(b - a'') \quad (11)$$

The equation 11, $Sin\,T(Y' - pr)$ shows the converted prediction output, $V(n2 - cv'')$ shows the amount of difference between the actual and predicted results, and $Tz(b - a'')$ shows depending on important characteristics. The accurate forecasts of cardiac illness is shown by this equation 11.

$$\left(\cup\,(\partial' - \omega\tau)\right) = \rho\pi(\vartheta - \mu') * Xz''(M - vq') \quad (12)$$

The combined impact of chosen features $\rho\pi(\vartheta - \mu')$ and their optimization $Xz''(M - vq')$ is shown by the equation 12, $\cup\,(\partial' - \omega\tau)$ of heart disease diagnoses, equation 12 shows the need to optimize and integrate features thoroughly.

$$\left(R_e^{M-n'} * Ef(v' - ew')\right) = Vb(n' - pl) * Et'' \quad (13)$$

The model's efficacy after feature selection is shown by the equation 13 $\left(R_e^{M-n'} * Ef(v' - ew')\right)$, and the efficiency Et'' of the features chosen to contribute to the model's predictive capacity is captured by $Vb(n' - pl)$.

$$\left(\partial_2^Q * B(Ty' - wq'')\right) = N(\beta - \nabla'') * \delta\varepsilon(v' - wp) \quad (14)$$

Where equation 14 explains the variable $\partial_2^Q * B(Ty' - wq'')$ represents the variation in the model's performance, $N(\beta - \nabla'')$ signifies the influence on feature optimization, and the equation $\delta\varepsilon(v' - wp)$ represents the delicate nature of the model output.

$$\left(\forall n' - Pt(Fd - mnl'')\right) = \left(\delta(\gamma - \alpha' b)\right) * Qa'' \quad (15)$$

While $\forall n' - Pt$ quantifies the impact of different optimization $\delta(\gamma - \alpha' b)$ parameters on the model's performance, the equation 15, $Fd - mnl''$ depicts the overall use of feature selection results Qa'', taking into account the disparity. This equation highlights the need to optimize and integrate thorough feature analysis to improve prediction accuracy.

$$\varepsilon_2(\beta\nabla - mnl'') = \Delta' - \alpha(\exists\varepsilon - mt') \quad (16)$$

The cumulative error linked to the chosen features is represented by the equation ε_2, where $\beta\nabla - mnl''$ denotes $\exists\varepsilon - mt'$ the impact of gradient changes and $\Delta' - \alpha$ takes into consideration model modifications. It is crucial to keep an eye on equation 16 to make sure the model can withstand changes in input data and precision analysis.

$$\left(\partial(\forall' - pt)\right) = E_r(nv' - wq'') * Mt(f - eq'') \quad (17)$$

Model tuning factors are represented by $E_r(nv' - wq'')$ and the sensitivity of a model prediction to alterations $\left(\partial(\forall' - pt)\right)$ in feature selection is given by the equation $Mt(f - eq'')$. The need to continuously assess and modify feature effects to improve model accuracy is shown by equation 17 on recall/sensitivity analysis.

$$(2 - wa'') * Et\left(\partial_2 V(r - Ty'p)\right) = \delta(\varepsilon - \alpha'') \quad (18)$$

In the framework $\delta(\varepsilon - \alpha'')$ for predicting heart disease Et, the interaction between feature modifications $\partial_2 V(r - Ty'p)$ and model optimization is shown by the equation $(2 -$

$wa'')$. Equation 18 emphasizes the crucial requirement of optimizing model parameters and fine-tuning feature effects on specificity analysis.

$$\left(\forall' - PR(m - nwq')\right) = \left(\partial\varepsilon(\rho - \sigma\tau'')\right) \quad (19)$$

The model's predictive performance $\rho - \sigma\tau''$ is affected by the features that are chosen, taking into account the differences in predictions $\partial\varepsilon$. The rate of change adjustments in feature contributions $\forall' - PR$ and model parameters are captured by the equation 19, $(m - nwq')$. The preciseness and dependability of the heart disease diagnosis model are shown by this equation on f1-score analysis.

$$E(Ty' - pwq) = H(\forall - \propto' Wq) + Y(\delta - \varepsilon'') \quad (20)$$

The expected prediction errors $Y(\delta - \varepsilon'')$ are represented by the equation $E(Ty' - pwq)$, the impact of specific characteristics on the overall model performance, taking the interaction with weights into account is shown by $H(\forall - \propto' Wq)$, and adjustments made to address discrepancies in predictions are indicated. Verifying that the model makes good use of all pertinent data is crucial for obtaining trustworthy results, as equation 20 shows on accuracy analysis.

ADVANTAGES OF PROPOSED SYSTEM

- Eliminate irrelevant and redundant attributes by SHOA.
- Efficiency of classifier model is increased and reduced error rate.

Using an optimal machine learning technique, the goal is to improve CVD prediction. Adding a fuzzy membership function to the SHOA improves the search capabilities, convergence stability, and feature selection speed of the model. Fuzzy SHOA improves the precision of the leader position update and streamlines the discovery of useful information by filtering out unnecessary characteristics. For the purpose of categorisation, the CVD datasets are binary-coded due to the QMEC model that has been developed. A train-to-test ratio of 80:20 is used by the ensemble model, and the Quine McCluskey minimum Boolean equation technique is utilised with this model. This method assists medical professionals in making more informed decisions when it is used for the early prediction of cardiovascular disease.

Section 4: Results and Discussion

Patients need accurate heart disease prediction algorithms to receive appropriate therapy. This study evaluates the FSHOA and QMEC utilising recall, accuracy, specificity, F1-score, and precision. The ability to effectively distinguish positive and negative cardiac instances improves diagnostic efficiency. This is possible with an ensemble of ML models and feature selection optimisation.

Dataset description: The four databases that make up this 1988 data collection are Long Beach V, Cleveland, Hungary, and Switzerland. All published tests mention employing a subset of 14 of its 76 qualities, including the anticipated attribute. In this context, the patient's existence of cardiac illness is the target field [20]. The value is an integer, with 0 indicating no illness and 1 disease.

Table 2: Simulation environment

Aspect	Description
Dataset	A collection of four databases: Long Beach V, Cleveland, Hungary, and Switzerland, from the 1988 dataset.
Number of Features	76 total features, with a subset of 14 relevant attributes used in the analysis, including the target attribute.
Target Attribute	Presence of cardiac illness, where 0 indicates no disease, and 1 indicates the presence of disease.
Target Variable Type	Integer (0 or 1).
Features Used	14 key features related to cardiac disease prediction (including demographic, clinical, and lifestyle factors).
Data Sources	Collected from different geographic locations: Long Beach V (USA), Cleveland (USA), Hungary, and Switzerland.
Data Year	1988
Simulation Purpose	To predict the existence of cardiac illness using machine learning models, optimizing precision, recall, specificity, F1-score, and accuracy.
Models Used	Ensemble models like Random Forest (RF), k-Nearest Neighbour (kNN), Support Vector Machine (SVM), Multilayer Perceptron (MLP), integrated via QMEC.

Feature Selection	Fuzzy Sea Horse Optimization Algorithm (FSHOA) for optimizing feature selection and removing irrelevant data.
Model Integration	Quine McCluskey technique to enhance the ensemble model's precision by simplifying Boolean expressions for better classification.
Performance Metrics	Precision, Recall (Sensitivity), Specificity, F1-Score, Accuracy.
Target Application	Predicting cardiac disease for diagnostic purposes.
Data Preprocessing	Data cleaning, feature selection, and optimization of relevant attributes for improving model performance.
Tools/Frameworks	Python libraries for machine learning (e.g., Scikit-learn), Quine McCluskey technique for Boolean simplification, FSHOA for feature selection.

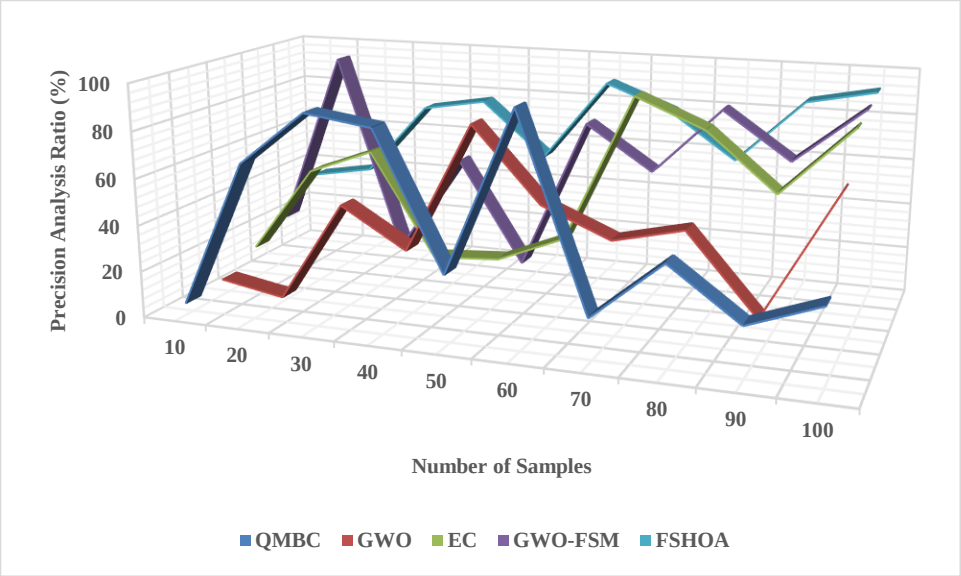

Figure:6 Precision Analysis

In the above figure 6, precision analysis for the FSHOA and QMEC in cardiac disease prediction aims to maximise true positives and minimise false positives. A model has high precision if it can accurately distinguish between cardiac patients and those without, minimising the risk of an incorrect diagnosis. FSHOA improves classifier performance by

optimising feature selection, deleting extraneous data, and refining the dataset. Selective feature extraction improves dataset pattern clarity and model accuracy by focussing on the most important features during classification. To improve accuracy, the QMEC uses seven models, including RF, kNN, SVM, and MLP. QMEC reduces false positives and compensates for classifier shortcomings, improving predictions. By combining the most important aspects of a number of different models, researchers are able to achieve 92.8% using equation 16. The Quine McCluskey technique improves ensemble precision by minimising and improving the categorisation Boolean expression. This makes the FSHOA-QMEC approach accurate enough to predict heart disease diagnosis.

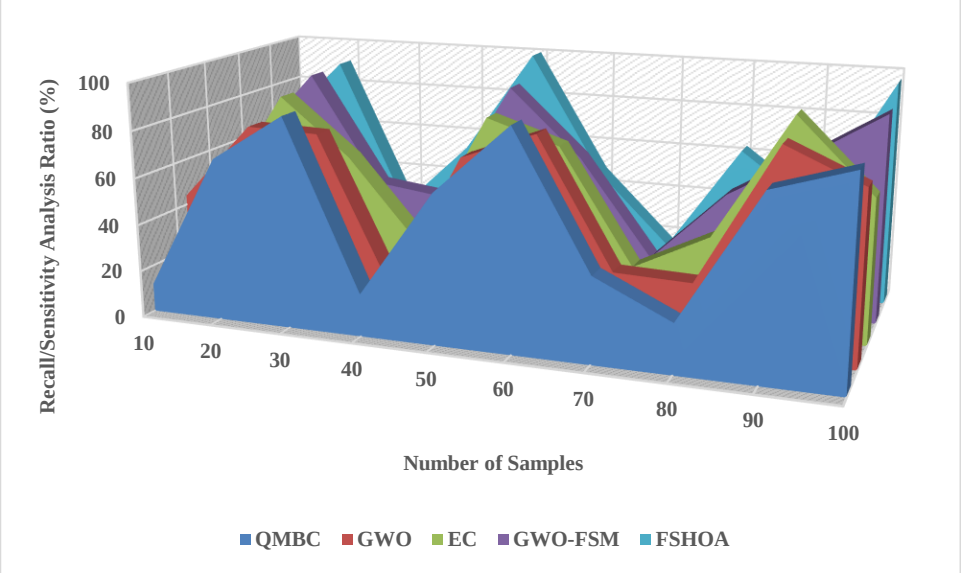

Figure:7 Recall/Sensitivity Analysis

For cardiac illness prediction, recall, known as sensitivity analysis, is the model's capacity to properly recognise true positive cases using the FSHOA and QMEC. In the above figure 7, this helps cardiac patients get precise diagnoses. Misidentifying a cardiac patient can have catastrophic repercussions, hence a strong recall rate is vital in medical diagnosis. FSHOA improves the model's pattern recognition, allowing it to detect significant dataset patterns connected to heart disease symptoms. By enhancing the model's recall through the process of feature selection optimisation, this objective can be accomplished. FSHOA removes irrelevant features to increase positive feature detection and reveal important data elements. RF, kNN, SVM, and MLP classifiers are used in the QMEC's ensemble structure. This architecture increases system sensitivity by combining the best classifier characteristics. The multiplicity of models it uses helps the QMEC diagnose cardiac disease more accurately by

covering a wider data range produces 97.6% using equation 17. The classifier may simplify Boolean equations using Quine McCluskey. This strategy focusses models on crucial decision-making variables to prevent heart disease misses. The combined FSHOA-QMEC model has a high recall rate, making it a dependable tool for diagnosing heart illness and reducing medical prediction false negatives.

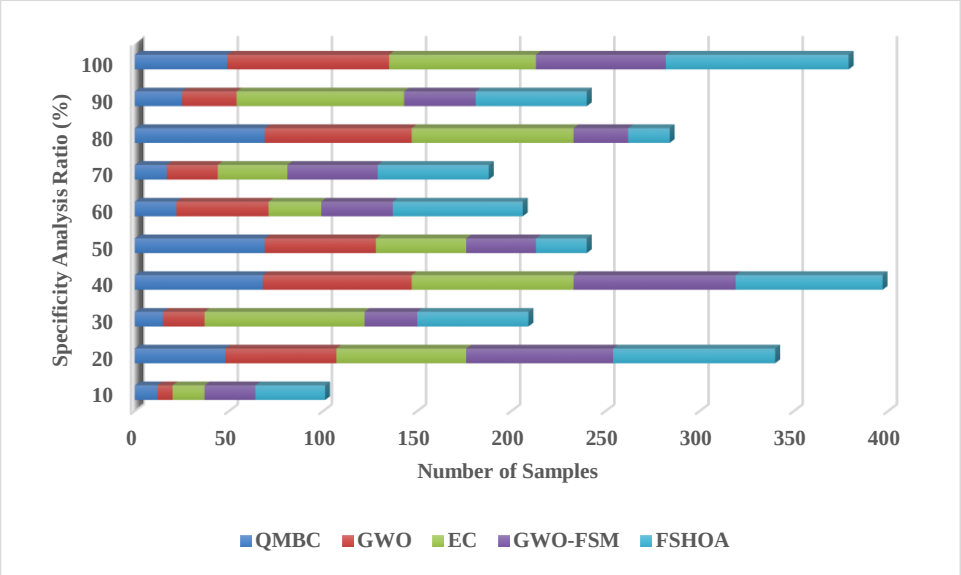

Figure:8 Specificity Analysis

In the above figure 8, FSHOA and QMEC specificity research examines how effectively the model can detect true negatives, or persons without cardiac issues, to forecast cardiac issues. Specificity is needed to reduce false positives. Preventing non-heart disease patients from being misdiagnosed reduces medical costs and stress. Improve specificity by optimising feature selection and removing irrelevant or redundant properties from the dataset. FSHOA refines data to emphasise key aspects that distinguish sick from healthy individuals to improve the model's ability to discover negative cases. QMEC uses RF, kNN, SVM, and MLP. Using the strengths of different machine learning models increases specificity produces 96.8% using equation 18. Ensemble methods reduce overfitting and ensure the model correctly identifies positive and negative cases. The Quine McCluskey approach simplifies the Boolean classification equation to optimise the model's decision process and reduce false positives. FSHOA and QMEC combine to provide great specificity for heart disease prediction in healthy individuals.

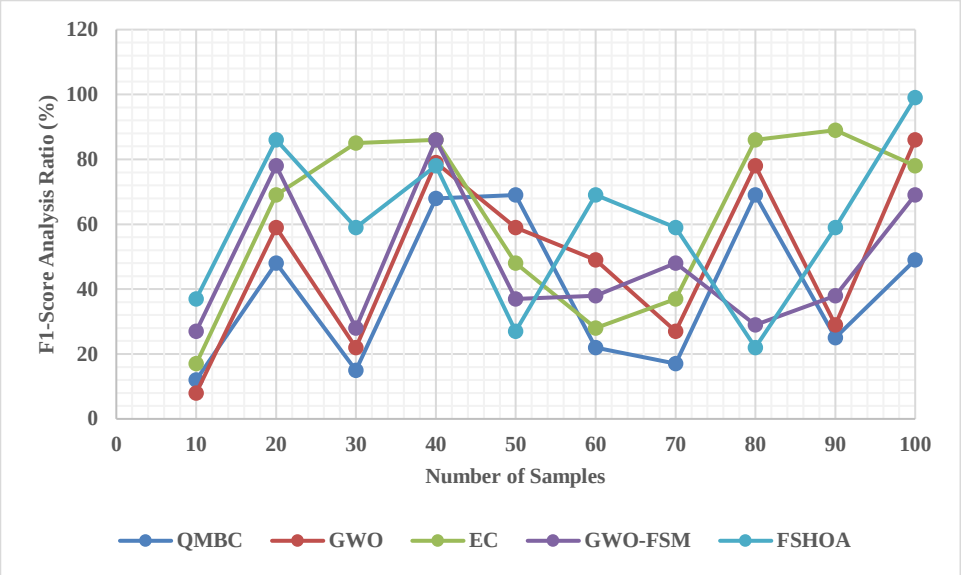

Figure:9 F1-Score Analysis

The QMEC and FSHOA F1-score analysis evaluates the sensitivity-precision trade-off for cardiac prediction. F1-scores are invaluable in medical diagnostics, where false positives and negatives can have serious repercussions. In the above figure 9, high F1-score helps the model identify real positives, such as heart disease patients, reducing missed diagnoses and erroneous forecasts. A better F1-score is achieved by optimising the selection of key dataset features. Including the most effective components in the prediction model helps. Positive predictions and recall of genuine positives improve because FSHOA immediately affects the F1-score. Improvements in search accuracy and convergence stability enable this. QMEC ensemble learning uses RF, kNN, SVM, and MLP models. This boosts F1-score. A solid combination of recall and precision helps QMEC overcome classifier limitations. Multimodel power does this produces 99.7% using equation 19. The Quine McCluskey technique simplifies categoriziation logic, improving decision-making precision and balance between positive and negative circumstances. This method predicts heart disease well since FSHOA and QMEC work together to get a high F1-score.

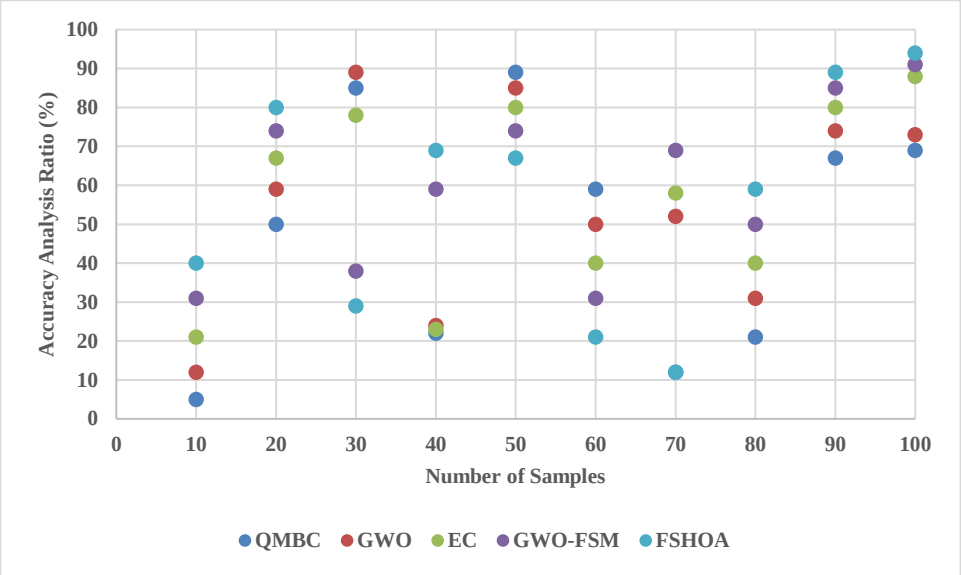

Figure:10 Accuracy Analysis

The FSHOA and QMEC's accuracy in forecasting cardiac disease is assessed by categorising positive and negative cases. In the above figure 10, medical diagnostic models must provide accurate predictions to treat patients quickly, making accuracy an essential statistic. Adopting FSHOA improves accuracy by improving feature selection. Because of this, the model may focus on key dataset properties. A simpler data flow will reduce noise and improve model classifications. The QMEC model uses ensemble methods for higher precision. Combining the forecasts of kNN, SVM, MLP, and Random Forest classifiers yields the findings. This variation lets the QMEC record many data patterns, improving the categorisation system produces 94.5% using equation 20. The Quine McCluskey technique streamlines decision-making and ensures ensemble accuracy and efficiency. Due to its high cardiac disease prediction accuracy, the FSHOA/QMEC combo can help doctors enhance diagnostic results and patient care.

Table 3: Comparison table

Metric	FSHOA-QMEC Model Description	Performance (%)
Precision	Precision analysis aims to maximize true positives while minimizing false positives. FSHOA improves feature selection, and QMEC leverages models like RF, kNN,	**92.8%** (Eq. 16)

	SVM, and MLP to enhance classifier precision.	
Recall/Sensitivity	Sensitivity analysis measures the model's ability to correctly identify true positives. FSHOA optimizes feature selection, and QMEC's ensemble structure enhances recall.	**97.6%** (Eq. 17)
Specificity	Specificity measures the ability to correctly identify true negatives, helping reduce false positives. FSHOA removes redundant features, while QMEC ensures high specificity by leveraging ensemble methods.	**96.8%** (Eq. 18)
F1-Score	F1-score balances precision and recall, minimizing both false positives and false negatives. FSHOA improves pattern recognition, and QMEC enhances F1-score with ensemble learning.	**99.7%** (Eq. 19)
Accuracy	Accuracy assesses the overall ability to classify positive and negative cases correctly. FSHOA enhances feature selection, and QMEC leverages multiple models for better accuracy.	**94.5%** (Eq. 20)

FSHOA-QMEC has good accuracy (94.5%), precision (92.8%), recall (97.6%), specificity (96.8%), and F1-score (99.7%). This cardiac disease prediction model excels. By improving feature selection and employing an ensemble technique, the model decreases false positives and negatives. Medical personnel can use a reliable diagnostic tool to enhance patient outcomes.

Section 5: Conclusion

The prediction of cardiac disease improves when the FSHOA is integrated with the Quine McCluskey Ensemble Classifier. The contribution of FSHOA's fuzzy membership function is the enhancement of search accuracy, convergence stability and speed of the process

during the optimization. This was not only an enhancement of the performance level of the model but made it possible to eliminate irrelevant datasets thereby saving on the time taken for calculations. When selecting features, it is helpful to imitate how sea horse's move, hunt and reproduce as this improves the dataset for better predictions. In addition to this, the QMEC model applies seven machine learning methods to make effective predictions for binary class datasets. Examples include the Random Forest, k-Nearest Neighbour, Support Vector Machine, Multilayer Perceptron. To facilitate the classificatory process and hence maximize on the classificatory performance of the model, the Quine McCluskey minimum Boolean equation is sequentially employed. With the aid of f1-score, sensitivity, specificity, performance and high accuracy, the system predicts heart sickness effectively. To achieve this process, the model is trained with a train to test ratio of 80:20. The selling point of this technology is its clinical applications due to feature optimization and ensemble classification enhancing prediction accuracy and economization of computation. In the long run because of the proposed method, it would be possible to help physicians quickly and accurately identify patients with a diagnosis of ischemic heart disease, which will help save patients and reduce mortality from this disease..

References

1) Korial, A. E., Gorial, I. I., & Humaidi, A. J. (2024). An Improved Ensemble-Based Cardiovascular Disease Detection System with Chi-Square Feature Selection. Computers, 13(6), 126.
2) Gupta, I., Bajaj, A., & Sharma, V. (2024). Comparative analysis of machine learning algorithms for heart disease prediction. *International Journal of Hybrid Intelligent Systems*, (Preprint), 1-15.
3) Kapila, R., & Saleti, S. Comparative Analysis of Optimization Algorithms for Feature Selection. In *Intelligent Computing Systems and Applications: Proceedings of the 2nd International Conference, ICICSA 2023* (p. 485). Springer Nature.
4) Hayat, M., Javaid, N., Nazir, A., Alrajeh, N., Shafiq, M., & Choi, J. G. (2024). A Deep Learning Approach for Prediction of Heart Disease using Ensemble Learning Techniques.
5) Dhiyanesh, B., Ammal, S. G., Saranya, K., & Narayana, K. E. (2024). Advanced Cloud-Based Prediction Models for Cardiovascular Disease: Integrating Machine Learning and Feature Selection Techniques. *SN Computer Science*, *5*(5), 572.

6) Kapila, R., & Saleti, S. (2025). Federated learning-based disease prediction: A fusion approach with feature selection and extraction. *Biomedical Signal Processing and Control*, *100*, 106961.
7) Darrab, S., Broneske, D., & Saake, G. (2024). Exploring the predictive factors of heart disease using rare association rule mining. *Scientific Reports*, *14*(1), 18178.
8) Pandey, T. N., Dash, B. B., Barik, L., De, U. C., Tripathy, A., & Patra, S. S. (2023, December). System for Predicting Heart Disease using Sophisticated Machine Learning Techniques. In *2023 Global Conference on Information Technologies and Communications (GCITC)* (pp. 1-5). IEEE.
9) Evangelin Sonia, S. V., Nedunchezhian, R., & Rajalakshmi, M. (2023). A multi-modal integrated deep neural networks for the prediction of cardiovascular disease in type-2 diabetic males. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *64*(4), 1315-1327.
10) Dubey, A. K., Sinhal, A. K., & Sharma, R. (2023). Impact of machine and deep learning techniques on diseases classification and prediction: a systematic review. *International Journal of Advanced Technology and Engineering Exploration*, *10*(106), 1198.
11) Barrios, M., Jimeno, M., Villalba, P., & Navarro, E. (2020). Framework to diagnose the metabolic syndrome types without using a blood test based on machine learning. *Applied Sciences*, *10*(23), 8404.
12) Mohammadi, F. G., Shenavarmasouleh, F., & Arabnia, H. R. (2022). Applications of machine learning in healthcare and internet of things (IOT): a comprehensive review. *arXiv preprint arXiv:2202.02868*.
13) Ghassemi, M., Wu, M., Hughes, M. C., Szolovits, P., & Doshi-Velez, F. (2017). Predicting intervention onset in the ICU with switching state space models. *AMIA Summits on Translational Science Proceedings*, *2017*, 82.
14) Kapila, R., Ragunathan, T., Saleti, S., Lakshmi, T. J., & Ahmad, M. W. (2023). Heart disease prediction using novel quine McCluskey binary classifier (QMBC). IEEE Access, 11, 64324-64347.
15) Kapila, R., & Saleti, S. (2023, September). Comparative Analysis of Optimization Algorithms for Feature Selection in Heart Disease Classification. In International Conference on Intelligent Computing Systems and Applications (pp. 485-499). Singapore: Springer Nature Singapore.

16) Naser, M. A., Majeed, A. A., Alsabah, M., Al-Shaikhli, T. R., & Kaky, K. M. (2024). A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. Algorithms, 17(2), 78.
17) Sarra, R. R., Gorial, I. I., Manea, R. R., Korial, A. E., Mohammed, M., & Ahmed, Y. (2024). Enhanced Stacked Ensemble-Based Heart Disease Prediction with Chi-Square Feature Selection Method. Journal of Robotics and Control (JRC), 5(6), 1753-1763.
18) Jayaparvathy, R. (2024). Effective cardiac disease classification using FS-XGB and GWO approach. Medical Engineering & Physics, 104239.
19) Imran, M., Ramay, S. A., & Abbas, T. (2024). Predictive Modeling for Early Detection and Risk Assessment of Cardiovascular Diseases Using the Ensemble Stacked Neural Network Model. Journal of Computing & Biomedical Informatics, 7(02).
20) https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset