CHAPTER 7

OPTIMAL SELECTION OF BASE CLASSIFIERS IN AN ENSEMBLE USING WHALE OPTIMIZATION ALGORITHM AND DIVERSITY MEASURES

This chapter discusses the proposed Optimized Super Learner Ensemble Model (OSLEM), where optimal selection of base models in SLEM is done by using Whale Optimization Algorithm (WOA) and diversity measure. The proposed OSLEM is modeled on Heart Datasets of both dimensions with features chosen by ModifiedBoostARoota (MBAR). The proposed OSLEM model is evaluated and its efficiency is compared with the other models.

7.1 INTRODUCTION

As the number of deaths caused by cardiovascular disease continues to rise at an alarming rate, it has emerged as one of the most pressing issues in modern medicine. Diseases can be detected earlier with the help of prediction, and deaths can be avoided with the help of treatment or lifestyle adjustments. Previously, the SLEM model was developed with a diverse combination of linear, probabilistic, bagging, boosting and stacking models, for heart disease prediction. Previous studies found that selecting the best base models from among many options will lead to optimal ensemble model performance.

Therefore, in this study, we employ the Whale Optimization Algorithm (WOA) to create an OSLEM that more accurately chooses the best ensemble classifiers. Disagreement pairwise diversity measure is a metric used to assess the dissimilarity or variation among the predictions made by individual classifiers within an ensemble. It quantifies the level of disagreement or diversity between pairs of classifiers. It provides insight into the differences in opinions or perspectives among the classifiers. So, on measuring the disagreement pairwise diversity, the subsets of classifiers with greater diversity measure are chosen as the base models for the SLEM classifier.

Several base classifiers are used initially depending on their performance accuracy gained through iterative stratified k-fold cross validation, as detailed in section 6.2. Whether or not these base classifiers are incorporated in the final ensemble is encoded by

the whale population in the WOA technique. Then, on applying WOA on the population of whales for different sets of iterations, a population with good fitness value is obtained. Then the diversity measure for each whale is computed and the whale with the maximum diversity is chosen as the best fitness whale. In the ensemble SLEM, the best fitness whaleis used to pick the base classifiers. The generated super learner transfers its knowledge from the training data to the test data. The performance of OSLEM is evaluated for the heart disease prediction.

7.2 OVERVIEW OF OPTIMIZING ENSEMBLE MODEL FOR CLASSIFICATION METHODS

Classifier ensembles have been increasingly popular in recent years due to the inefficiency of trying to identify a single classifier model that performs at a wide variety of tasks. In ensemble learning, weaker students' knowledge is combined with that of stronger ones to improve performance. Ensembles have been demonstrated to outperform single classifiers in a number of empirical experiments (Dietterich 2000; Duda et al., 2006). The primary difficulty with ensemble methods is that they tend to produce an excessively large number of classifiers in the final ensemble. It has been shown experimentally that the generalization performance of an ensemble may be maintained despite a reduction in the number of classifiers (Zhou et al., 2002). Ensemble selection is clearly one of the most critical processes in the development of ensemble classifiers.

The quickest and easiest method is to pick a single, top-performing classifier fromthe available training data and apply it to previously undiscovered patterns. The best outcomes are not necessarily achieved with the simplest methods (Roli & Giacinto, 2002).Researches have shifted their attention to the topic of ensemble pruning since a thorough search for the finest subset of an ensemble may become intractable for relatively modest ensemble sizes. The selection criterion is the most important component of selecting a classifier. It appears that combined performance is the most appropriate metric by which to judge the combiner, as compared to individual mean classifier performance. The exponential complexity of evaluating all conceivable combinations of classifiers is a clear disadvantage of using performance as a selection criterion. When selecting classifiers based on combiner performance, there is also a significant risk of overfitting.

The effective growth of classifier ensembles can also be aided by using a measure of variety as a selection criterion. Researchers dispute on whether or not diversity boosts the effectiveness of the combined classifier, as evidenced by many studies (Dietterich 2000; Li et al. 2012; Tsymbal et al. 2005; Oliveira et al. 2006). However, there is a positive correlation between ensemble performance and diversity between base classifiers, as shown by empirical evidence. For instance, Cavalcanti et al. (2016) presented a GA-based pruning method that took advantage of numerous types of pairwise diversity. They createdpotential ensembles using techniques from graph coloring theory. They showed that a combination of various diversity measures might be an efficient strategy for pruning an ensemble of classifiers and improving identification rates by comparing their results to those of five state-of-the-art strategies in ensemble pruning. When choosing a classifier triplet from a pool of five, Kim et al. (1997) used a similarity metric, although this did not ensure optimality.

However, not all authors are convinced that diversity measurements help when creating ensemble classifiers. Kuncheva and Whitaker (2003) conducted numerous trials but were unable to establish an official relationship between diversity measures and ensemble accuracy improvement. In other words, while it is essential to develop varied classifiers, it is still difficult to evaluate this variety and effectively employ it when constructing stronger ensembles. Rogova (1994) discovered that the quality of the classifiers as a whole could not be inferred from their individual results. Ruta and Gabrys (2005) evaluate the value of diversity metrics in the context of combining classifiers. They found that using diversity metrics as selection criteria was less effective than using combiner errors directly.

Previous papers have reported a wealth of experimental evidence (Kuncheva, & Whitaker, 2003; Ruta and Gabrys 2005; Kuncheva, & Whitaker, 2001; Ruta & Gabrys 2005; Shipp & Kuncheva, 2002; Ko et al., 2007) that suggests measures of diversity have either nil effect on final performance or a very less effect. Diversity metrics prioritize diversity when making decisions. Using combiner performance as a criterion for selection is specific, significant, and permits continuous evaluations of different classifier subsets, independent of the number of classifiers and their own results. When it comes to the challenge of constructing an ensemble of classifiers, existing diversity measures are not

practical. For example, Tang et al. (2006) believed that choosing the right diversitymeasure can affect classification outcomes. Combining predictors and eliminating duplicates is a common topic of study in ML, with the goal of improving accuracy. To choose classifiers for an ensemble, Wu et al.'s (2001) GA based Selective ENsemble (GASEN) method uses evolved weights that could relate to the fitness of incorporating the classifiers in the ensemble. They conducted an experiment using neural networks as classifiers, GA, and several data sets.

There are a variety of dynamic selection models that greatly outperform, the best individual classifier (Giacinto & Roli, 2000; Pérez-Gállego et al. 2019; Cruz et al. 2017). Cluster and select-based algorithms have a challenging problem when it comes to the amount of clusters, which could affect the algorithms' efficiency. Clustering-based and dynamic selection approaches, while greatly reducing the complexity of the selection process, nonetheless increase the entire model's complexity and do not ensure the search is optimal, even on a local scale.

On addressing the above issues, the WOA is used for optimal selection of classifiers in ensemble classifiers. Ensemble pruning presents a ‘diversity based on accuracy’ pruning technique that takes into account both the accuracy of combined classifiers andthe pairwise diversity among these classifiers across datasets.

7.3 PROPOSED METHODOLOGY

The framework of OSLEM model is brief illustrated in below sections.

7.3.1 Constructing the base classifiers

The SVM, LR, GNB, RF, DT, KNN, MVE, XGB, and CatB ML base learners are employed in this framework; their brief descriptions may be found in section 6.2. These base learners are selected in this work based on the variety of learning models. Stratified ten-fold cross validation is performed three times, and the average accuracy of each classifier is determined. WOA's fitness function computation relies on the accuracy metric.

7.3.2 Utilizing WOA for selecting the subset of classifiers

The well-known classifier selection problems can be well-fit by the population-based evolutionary models. Using the Whale Optimization Algorithm (WOA) (Mirjalili S

& Lewis Andrew, 2016), we can eliminate unnecessary base classifiers and find the optimal ensemble. After doing extensive research on whale behavior, particularly humpback whales' sophisticated hunting techniques, the WOA population-based meta- heuristic algorithm was developed. Not only are whales the largest mammals on Earth, butthey are also extremely intelligent and have developed complex social behaviors, making them really remarkable creatures.

The WOA model incorporates the bubble-net feeding strategy, a unique form of humpback whale hunting. Figure 7.1 is a graphic representation of the bubble-net feeding hunting strategy. To improve the ensemble model, this approach is often applied to the task of subset selection within an ensemble model.

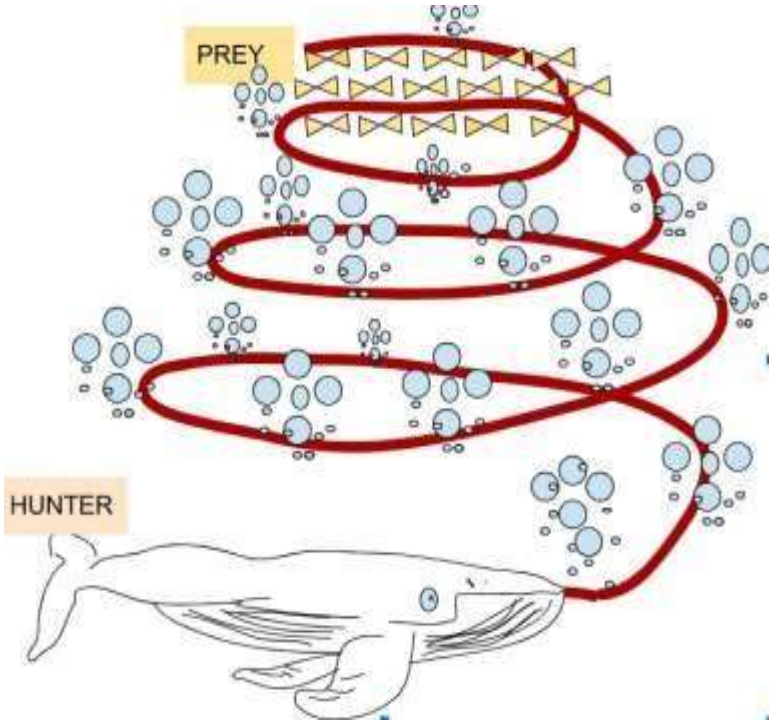

Figure 7.1 Visualization of bubble-net feeding hunting method

Below is an illustration of the mathematical model of feeding behavior in a spiral bubble net.

(i) Encircling prey

With pinpoint accuracy, the humpback whale can home in on prey and encircle it. Because the precise coordinates of the optimal plan in the search space are unknown in advance, the WOA algorithm proceeds from the premise that the best candidate solution is relatively close to the target prey or the global optimum. Once the top search agent has been

identified, the remaining search agents will naturally gravitate toward it by updating their positions. The term "exploitation process" describes this action. The subsequent equations model this behavior.

$$\vec{D} = \left|\vec{C} \bullet \vec{X}^*(t) - \vec{X}(t)\right| \quad (7.1)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \bullet \vec{D} \quad (7.2)$$

In equations (7.1) and (7.2), $\vec{X}$ is the position vector, $\vec{X}^*$ denotes the best solution obtained so far, which will be updated in each iteration if there is a better solution, t represents the current iteration, '.' denotes the absolute value operation $|.|$, and $\bullet$ is an element-by-element multiplication. Here, $\vec{A}$ and $\vec{C}$ are two parameters, and they are calculated as shown in (7.3) and (7.4),

$$\vec{A} = 2\,\vec{a} \bullet \vec{r} - \vec{a} \quad (7.3)$$

$$\vec{C} = 2 \bullet \vec{r} \quad (7.4)$$

The symbol $\vec{r}$ is a random number in the interval $[0,1]$. The decreasing enclosing behavior is achieved by $\vec{a}$ dropping linearly from 2 to 0 during the course of exploration and exploitation repetitions.

(ii) Bubble-net attacking method

When using a bubble net for hunting, the attack is the primary component. Step two, exploitation, is represented in WOA parlance by a decreasing encircling mechanism and a spiral updating position. The first behavior is modeled with a vector $\vec{a}$ whose value decreases as the iteration number increases, and the second behavior is modeled by introducing a specific equation to model the whales' spiraling motion, as shown in Eq. (7.6). In reality, when hunting, both of these actions happen at the same time as the whales spiral upwards towards the surface, decreasing their body size. This behavior can be quantified as follows:

$$\vec{D}' = \left|\vec{X}^*(t) - \vec{X}(t)\right| \quad (7.5)$$

$$\overrightarrow{X}(t+1) = \overrightarrow{D'} \bullet e^{bl} \bullet cos\,(2\pi\, l) + \overrightarrow{X^*}\,(t) \tag{7.6}$$

$\overrightarrow{D'}$ in Eq. (7.5) is the distance between the best solution so far and the i[th] whale, where 'l' takes random values between -1 and 1 and the profile of the logarithmic spiral is defined by the constant 'b'. When whales hunt, they simultaneously employ a spiraling course and a shrinking encirclement to secure a kill. Each mechanism is deployed with an equal chance to simulate this behavior.

$$\overrightarrow{X}(t+1) = \begin{cases} \overrightarrow{X}^*(t) - \overrightarrow{A} \bullet \overrightarrow{D} & if\ p < 0.5 \\ \overrightarrow{D'} \bullet e^{bl} \bullet cos\,(2\pi\, l) + \overrightarrow{X^*}\,(t) & if\ p \geq 0.5 \end{cases} \tag{7.7}$$

where p is a random number in [0,1].

The exploration phase is a new mechanism introduced by WOA. Its primary purpose is to look for alternatives to the greatest option currently available. As a result, WOA can search in a worldwide context. It's mathematically identical to equation (7.7),but instead of selecting the optimal agent to direct the search, it picks an agent at random. To determine whether the position should be updated through exploration (the hunt for food) or exploitation (a tightening of the encircling mechanism), the random variable |A| is used.

$$\overrightarrow{D} = \left| \overrightarrow{C} \bullet \overrightarrow{X_{rand}} - \overrightarrow{X} \right| \tag{7.8}$$

$$\overrightarrow{X}\,(\mathrm{t}+1) = \overrightarrow{X_{rand}} - \overrightarrow{A} \bullet \overrightarrow{D} \tag{7.9}$$

Where $\overrightarrow{X_{rand}}$ is a location vector selected from the current set at random.

Algorithm for whale optimization algorithm (WOA)

Get the population of whales started $X_i\,(i = 1, 2, 3 \ldots \mathrm{n})$

Each whale's fitness levels must be calculated

X^*= the best whale

while $(t < maximum\ number\ of\ iterations)$

for each whale

Update l , p , A , a and C

If 1 (p is $less$ $than$ 0.5)

if 2 ($|A|$ is $less$ $than$ 1)

Apply the equation (7.1) to update the location of the current of the whale

else if 2 ($|A|$ is $greater$ $than$ or $equal$ to 1)

Pick a whale at random (X_{rand})

Apply the Equation (7.9) to update the location of the current whale.

end if 2

else if 1 (p is $greater$ $than$ or $equal$ to 0.5)

Apply the equation (7.7) to update the location of the current whale.

end if1

end for

Verify the authenticity of each whale by checking for and fixing any instances of duplicates

Use Eq. (7.10) to determine the fitness of each whale.

Update the value of X^* if there is a better answer

$t = t + 1$

end while

return value of X^*

In this work, in WOA, the size of each whales are similar to the number of base classifiers. Classifier incidences make up each whale, with zeros representing the absence of that classifier from the combination and ones representing its presence. The following Eq. (7.10), is used for calculating the fitness function of WOA for selecting the subset of classifier

$$F(i) = \max_{\vartheta} \left(accuracy\ (X) - \gamma * \frac{selected\ base\ classifier\ \ (X)}{Total\ number\ of\ base\ classifers\ (Y)} \right) \quad (7.10)$$

$F(i)$ = fitness function; P is the whale population, $accuracy\ (X)$ depicts classification accuracy of selected base classifier. The sum of the base classifiers is denoted by Y. γ is a parameter with weighted value. There are two steps to the Eq. (7.10). Both the classification accuracy and the fraction of base classifiers used in the final decision are taken into account during the first step. Variable γ has a value between zero and one. If they put γ close to 1, it shows that model correctness is more essential than the number of basic classifiers used. If two whales have identical accuracy values, the one with fewer base classifiers is chosen. When the average fitness of an infinite number of populations does not vary, we have the convergence condition.

7.3.3 Diversity measures

Pair-wise and non-pair-wise diversity measurements are distinguished in ML analysis.

- The purpose of pair-wise metrics is to evaluate the effectiveness of paired classifiers. Each pair of classifiers contributes to the ultimate value of the measure of pair-wise diversity, which is supplied as the average of their contributions. Notable pair-wise diversity metrics include the Q statistic (Kuncheva & Whitaker, 2003), measure of disagreement (Skalak, 1996), and measure of double-fault (Giacinto & Roli, 2001).
- Larger sets of classifiers can be evaluated with non-pair-wise metrics like the entropy measure (Cunningham & Carney, 2000), making them helpful for evaluating ensembles.

The number of contradictory assessments as a percentage of total observations is used as the disagreement pair-wise measure in this investigation.

7.3.4 Overall framework for the proposed OSLEM model

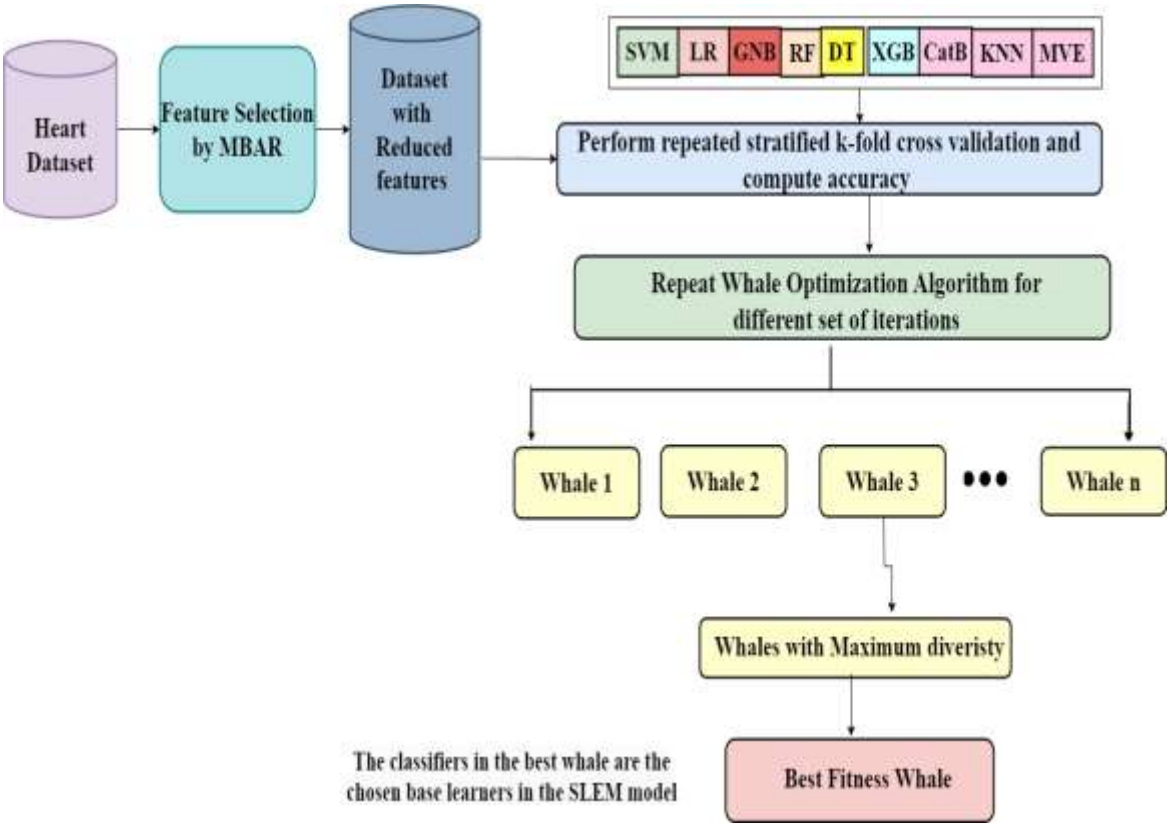

Figure 7.2 Overall framework for the proposed OSLEM model

The OSLEM model's recommended technique is shown in Figure 7.2. Initially, base classifiers SVM, LR, GNB, RF, KNN, DT, MVE, XGB and CatB are chosen based on the diversity of the learning models. The absence or presence of these base classifiers is coded as whales in WOA algorithm. The WOA outputs a whale with good fitness value. This is repeated for different sets of iterations. Then, the diversity measure value is calculated for each whale, and the most diverse whale is chosen. Super Learner Ensemble Model (SLEM) (shown in Chapter 6) uses the selected whale (base classifiers) as its foundation. The meta-learner, LR, learns to better integrate predictions from various base models, and the ensemble classifier, SLEM, uses this information as input. Finally, the proposedOSLEM model's performance is evaluated on test data. The proposed OSLEM model is devised to improve the ensemble model SLEM's performance (previous work) by selecting the optimal combination of base classifiers using WOA.

7.4 RESULTS AND DISCUSSION

Using Ubuntu OS and Python in Jupyter Notebook, the experiment was done on a system with an i5 processor and 4 GB of RAM. For this research work, the experiments were performed on five different low-dimensional heart datasets namely, Cleveland HD Dataset, Statlog HD Dataset, South African (SA) HD dataset, Cardio Vascular Disease Dataset, Cardiac Biomarkers Dataset and two different high-dimensional heart datasets namely, Z-Alizadeh Sani dataset and Arrhythmia dataset. Section 3.4 includes descriptions of all datasets. In addition, section 3.5 details the assessment measures employed for scoring the efficiency of the suggested algorithms.

7.4.1 Analysis of OSLEM Model on Low-Dimensional Heart Datasets

The Low-dimensional Heart Datasets are preprocessed and balanced using SMOTE. Then the significant features contributing to the heart disease are selected using MBAR. Then OSLEM is modeled on these selected features of each dataset and the performance is evaluated.

Table 7.1 demonstrates the accuracy attained by individual classifiers on experimenting 3 rounds of stratified cross-validation with 5 folds on the selected datasets. From Table 7.1, it is evident that the tree-based learners show significantly higher performance compared to other base learners. Also, SVM and KNN show lesser performance compared to their counterparts. These models would perform better if experimented on hyperparameter tuning. On the real-world cardiac biomarkers dataset, as only the highly significant features for the prediction of heart attack are chosen by MBAR, the proposed model and tree-based models have shown remarkable performance.

The proposed OSLEM model provides more efficient classification results than other individual classifiers considered. On optimizing the choice of base models in SLEM using WOA, the resultant OSLEM model's performance has improved considerably.

Table 7.1 Comparison of performance accuracy of the classifiers on low-dimensional datasets with Features Chosen by MBAR

Classifiers	Accuracy on datasets (with Feature selection by MBAR)				
	Cardiac Biomarkers Dataset	Cleveland HD Dataset	SA HD Dataset	Statlog HD Dataset	Cardio-vascular Dataset
CatBoost (CB)	100.0	85.5	81.8	83.3	99.0
Gaussian Naïve Bayes (GNB)	87.50	86.8	70.2	77.8	94.5
XGBoost (XGB)	100.0	84.2	76.9	84.4	97.2
Random Forest (RF)	100.0	85.5	76.8	83.3	98.6
Decision Tree (DT)	95.83	80.3	76.0	83.3	95.1
Majority Voting (MV)	93.75	88.2	71.9	83.3	98.3
Support Vector Machine (SVM)	75.0	85.5	72.7	77.8	79.7
K Nearest Neighbors (KNN)	81.25	82.9	72.7	83.3	83.1
Logistic Regression (LR)	93.75	86.8	69.4	77.8	96.9
Super learner Ensemble Model(SLEM)	97.9	85.5	81.0	83.3	98.6
SLEM with CB & DT	100	93.4	81.8	83.3	99.0
Proposed Optimized Superlearner Ensemble Model (OSLEM)	**100**	**93.4**	**89.2**	**92.6**	**99.0**

Table 7.2 displays the model evaluation metrics of the OSLEM on the low-dimensional heart datasets. Results show that the OSLEM has high precision and has steadily achieved better metric values than other classifiers across all the heart datasets.

Table 7.2 Performance Metrics of OSLEM on Low Dimensional Heart Datasets

Dataset	Classifier	Performance Metrics in %			
		Precision	Recall	f1-score	Accuracy
Cardiac Biomarkers	MBAR+ OSLEM	100	100	100	**100**
Cleveland Heart	MBAR+ OSLEM	95.80	88.50	92.00	**93.40**
SA Heart	MBAR+ OSLEM	92.00	74.10	82.10	**89.30**
Statlog Heart	MBAR+ OSLEM	88.90	88.90	88.90	**92.60**
Cardio vascular disease	MBAR+ OSLEM	98.50	99.20	98.80	**99. 0**

7.4.2 Analysis of OSLEM Model on High-Dimensional Heart Datasets

SMOTE is used to first fill in missing values and then to balance two high- dimensional datasets on cardiac disease: two datasets, Arrhythmia and Z-Alizadeh Sani (described in Section 3.4), were used. Then relevant features are selected using MBAR. The proposed OSLEM model when applied on the datasets exhibits high performance compared to other individual classifiers. The results of three iterations of stratified 5-fold cross validation on the given datasets are displayed in Table 7.3 for each classifier. The proposed OSLEM model is compared to the accuracy measurements of other models by experimenting on high-dimensional heart datasets. Astabulated in Table 7.3, OSLEM model demonstrates superior performance. The prediction time of the proposed model was 0.015 seconds.

Table 7.3 Comparison of the Performance of OSLEM with Other Classifiers on High-Dimensional Datasets

Classifiers	Performance Accuracy (in %) of the classifiers on	
	Arrhythmia Heart dataset	**Z-Alizadeh Sani heart dataset**
Logistic Regression	74.5	94.2
Gaussian Naïve Bayes	75.5	87.4
Decision Tree (DT)	71.4	90.8
Support Vector Machine	73.5	74.0
Random Forest	89.8	92.0
XGBoost	86.7	93.1
K Nearest Neighbors	68.4	61.0
CatBoost (CB)	89.8	96.1
Majority Voting Ensemble	82.7	94.3
Super Learner EnsembleModel (SLEM)	88.8	93.1
SLEM with CB & DT	90.0	96.1
Optimized Super Learner Ensemble Model (OSLEM)	**90.0**	**97.7**

Table 7.4 displays the model evaluation metrics of the OSLEM on the two high dimensional Heart Datasets. The table exhibits high recall performance by the proposed OSLEM model.

Table 7.4 Performance Metrics of OSLEM on High Dimensional Heart Datasets

Dataset	Classifier	Precision	Recall	f1-score	Accuracy
Arrhythmia heart dataset	MBAR+ OSLEM	88.00%	91.70%	89.80%	**90.00%**
Z-Alizadeh Sani heart dataset	MBAR+ OSLEM	97.87%	97.87%	97.87%	**97.70%**

7.4.3 Visualization of the performances of the classifiers on both low and high dimensional heart datasets

Accuracy values achieved by the proposed OSLEM and other classifiers, on low dimensional heart disorders datasets with specified features using MBAR, are graphically depicted in Figure 7.3. In case of Cardiac Biomarkers Dataset, the OSLEM model attains high accuracy of 100%. Similarly, for Cleveland HD Dataset, SA HD dataset, Statlog HD Dataset and Cardio Vascular disease dataset, the OSLEM model achieves 93.4%, 89.2%, 92.6% and 99.0% prediction accuracy. The proposed OSLEM outperforms other models.

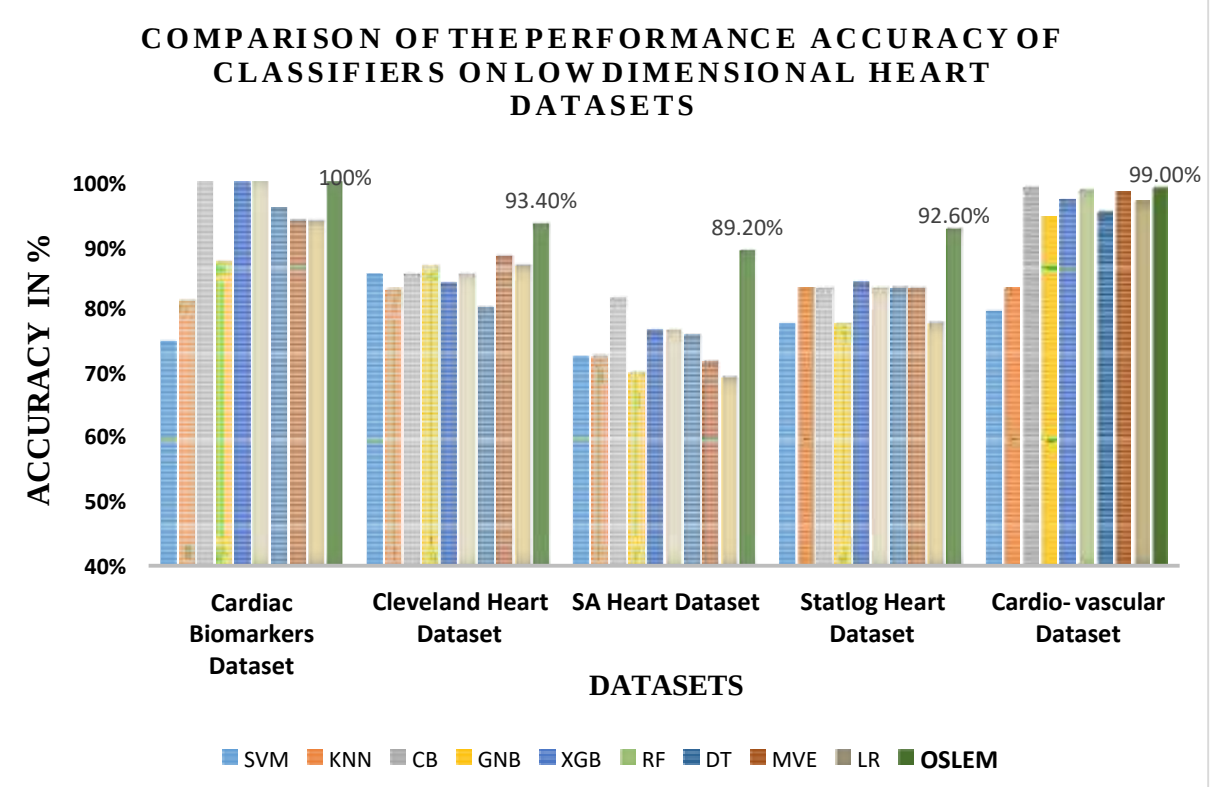

Figure 7.3 Performance Accuracy of the classifiers on Low Dimensional Heart Datasets

.

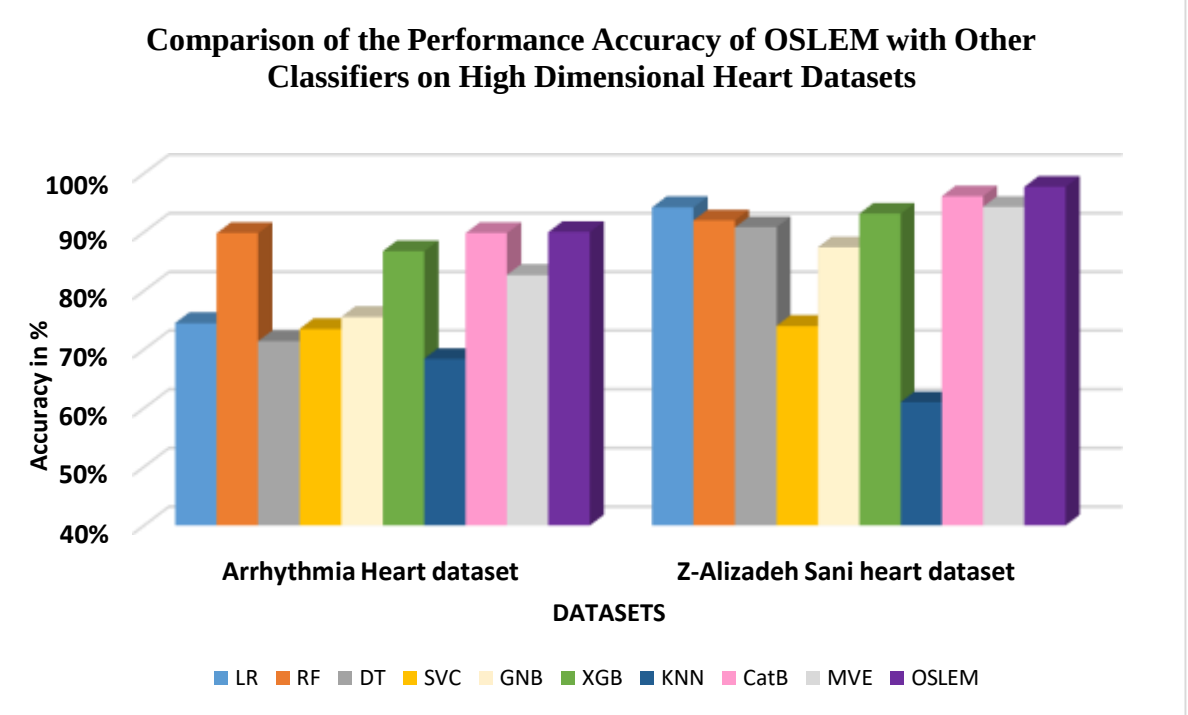

Figure 7.4 Performance Accuracy of the classifiers on High dimensional Heart Datasets

Figure 7.4 demonstrates that the proposed OSLEM model outperforms the other classifiers when it comes to classification accuracy. For the high-dimensional datasets like Alizadeh Sani dataset and Arrhythmia dataset, the OSLEM achieves 97.7% and 90% prediction accuracy.

Overall, the model evaluation metric values show that the proposed OSLEMmodel performs very well on the Heart Datasets.

7.5 CHAPTER SUMMARY

In this section, the OSLEM model is proposed for accurately predicting cardiovascular problems. To determine which SLEM (prior work) base classifiers are the most effective, we use the Whale Optimization Algorithm (WOA) in conjunction with the disagreement pairwise diversity metric. A meta-learner, LR, learns to better integrate predictions from these underlying models, and this is what the ensemble classifier, SLEM, uses as input. Across the Heart Datasets, the proposed OSLEM model outperforms the classifiers considered.