*Grenze International Journal of Engineering and Technology, January Issue*

# Multi Agent based Prediction of Heart Disease: Efficient Calculation of Execution Time and Accuracy

Chandanita Thakur[1], Shibakali Gupta[2] and Somsubhra Gupta[3]
[1,3]Department of Computer Science and Engineering, Swami Vivekananda University, Barrackpore, West Bengal, India,
Email: thakur.chandanita@gmail.com
[2]Department of Computer Science and Engineering, UIT, Burdwan University, West Bengal, India

***Abstract*—The present work has used Twenty-Six different Machine Learning Algorithms for prediction of heart disease by using Hybrid dataset (for heart disease). Accuracy, Balance Accuracy, ROC-AUC, F1-Score and execution time (time complexity) is calculated and compared for all Twenty-Six Algorithms. Multi Agent System is developed and Multiple Agents are trained with the best fit Algorithm. It is tested that the trained Agents can predict the heart disease with almost the same accuracy. Five heart disease datasets Cleveland, Hungarian, Long beach, Stalog (heart) and Switzerland are combined to achieve better accuracy for the prediction of heart disease. Feature selection is done by ANOVA f-test at the time of preparation of data. Here the dataset is split into two segments. One is training the model and another one is testing.70-80 percent of data has been kept for training and 20-30 percent for testing. The present work has used 67% data for training and the rest 33% of data for testing. Twenty Six Machine Learning based Algorithm is implemented by Jupyter notebook. Accuracy, Balance Accuracy, ROC-AUC, F1-Score and Execution time is calculated and compared for all Twenty-Six Algorithms. RandomForestClassifier() is proved as the best accurate algorithm (93.8%) for the present dataset. The best performing algorithm is implemented using Multi Agent System and achieved accuracy of approximately 96.8%. The work compares the accuracy of twenty-six different ML algorithms applied to the heart disease dataset. The best performing algorithm, achieving 93.8% accuracy, is integrated into the Multi Agent System and achieved accuracy of approximately 96.8% in predicting heart disease. Heart disease prediction in combination with Multi Agent System and Twenty-six different Machine Learning Algorithms is an innovation in the field of Machine Learning. The work has used a large dataset combining five standard sources and implemented using Jupyter Notebook.**

***Index Terms*—Multi Agent System (MAS), Machine Learning Algorithms, Accuracy, Heart Disease, Artificial Neural Network.**

## I. INTRODUCTION

Heart disease remains a leading cause of mortality worldwide, underscoring the urgent need for effective prediction and diagnosis methods. In recent years, machine learning (ML) has emerged as a powerful tool for predicting heart disease, with numerous studies demonstrating its potential to enhance diagnostic accuracy. Various machine learning algorithms, including Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), have been extensively employed to predict heart disease [1][2].

*Grenze ID: 01.GIJET.11.1.419_1*

However, these studies typically rely on a limited number of algorithms and standard datasets, which may not fully exploit the potential of machine learning for this critical application.

Despite significant advancements, there remains a gap in the literature regarding the integration of intelligent agents and multi-agent systems (MAS)[3][4] for heart disease prediction. Intelligent agents, capable of autonomous decision-making and learning, can enhance the prediction process by implementing and optimizing multiple machine learning algorithms. The use of a multi-agent system can further improve accuracy by enabling collaboration and coordination among different agents, each trained on various algorithms and datasets.

To address these gaps, the present research introduces an innovative approach by merging five different datasets to enhance the predictive accuracy of heart disease. This approach aims to leverage the diverse characteristics of each dataset, thus providing a more comprehensive and robust prediction model. Additionally, the research incorporates intelligent agents to execute and optimize twenty-six different machine learning algorithms, identifying the algorithm with the highest accuracy. The introduction of intelligent agents and the creation of a multi-agent system represent a significant advancement in the field, enabling a more sophisticated and dynamic prediction model.

The core contribution of this research is the development of a multi-agent system trained with the best-performing algorithms on a hybrid dataset. Each agent within the system is trained and tested to ensure high accuracy and reliability in predicting heart disease. By employing intelligent agents and a multi-agent system, the research achieves an unprecedented prediction accuracy of 96.8%, demonstrating the potential of this approach to significantly improve heart disease diagnostics. This research merges five different datasets to improve prediction accuracy and introduces Intelligent Agents to create a MAS. These agents, sharing the same memory, enable parallel computation and collectively determine the best model accuracy and Calculate the Execution time i.e. time complexity. This paper provides a comprehensive implementation of machine learning algorithms using intelligent agents and a multi-agent system for heart disease prediction, utilizing Python and Jupyter Notebook. The results indicate that this novel approach not only enhances predictive accuracy but also offers a scalable and efficient framework for integrating multiple datasets and algorithms in the prediction process.

## II. Literature Survey /State of Art Review on Heart Disease Prediction Algorithms (2021-2023)

TABLE I

| Ref No | Disease | Algorithm Used | Accuracy% | Year | Dataset |
|---|---|---|---|---|---|
| [1] | Cardiac | KNN, DT, NB, LR, RF | LR -96.6 | 2023 | Cleveland Dataset |
| [2] | Cardio vascular | RF,DT,MP,XGB | MLP-87.28 | 2023 | 70000 instances of Kaggle |
| [3] | Heart | LR, SVM,DT | LR -96.6 | 2023 | UCI |
| [4] | Heart | DT,GB,LR,SVM,RF | 86.7 | 2023 | Cleveland Dataset |
| [5] | Heart | RF,KNN,LR,NB,GB,AdaB | 93.44 | 2023 | Cleveland Dataset |
| [6] | Heart | LR,NN,SVM,DT,RF | RF-82.18 | 2023 | Cleveland dataset from UCI ML |
| [7] | Heart | LR,NN,SVM,DT,RF | RF-90 | 2023 | Data have been collected from hospitals, diagnostic centres and clinic centres in Bangladesh |
| [8] | Heart | DT,NB,LR,SVM,RF | 86.7 | 2023 | Realtime |
| [9] | Heart Attack | XGB,LR | XGB-94,LR-92 | 2022 | UCL |
| [10] | Heart attack | NB,DT ,WARM | DT-99.5 | 2022 | Raw Clinical |
| [11] | Heart | MLP,SVM,RF,NB | 91.67 | 2022 | NA |
| [12] | Cardiovascular | SVM | NA | 2022 | NA |
| [13] | Heart failure | LR,KNN,MNB,ET,RF,DT,GBC,DT LGBM, HGBC and XGB | RF-87.03 | 2022 | NA |
| [14] | Heart | LR,RF | RF-83 | 2022 | NA |
| [15] | Heart | LR,SVM,KNN | KNN-83<br>SVM-92.1<br>DT-89.6<br>LR-92 | 2022 | NA |
| [16] | Heart | LR, SVM, and KNN, | 93.44 | 2022 | NA |
| [17] | Cardiovascular | NB,KNN,LR | 87.5 | 2021 | UCI |
| [18] | Heart | KNN,RF | KNN-86.885<br>RF-81.967 | 2021 | NA |
| [19] | Heart (CVD) | SGD(SHDML frame work) | NA | 2021 | NA |

## III. METHODOLOGY

### *A. Problem statement*

The present work has used Twenty Six different Machine Learning Algorithms for prediction of Heart disease by using mixed dataset (for heart disease). Accuracy, Balance Accuracy, ROC-AUC,F1-Score is calculated and compared for all Twenty Six Algorithms. Multi Agent System is developed and Multiple Agents are trained with the best fit Algorithm. It is tested that the trained Agents can predict the heart disease with almost the same accuracy.

### *B. Description of Methodology*

*Data Collection*

In order to perform a heart disease prediction, the proposed method has adapted the High Dimensional dataset. At first, the heart disease data is created from five different datasets and used for analysis. This Heart Disease dataset has been presented by integrating various datasets already available individually but not combined before. Hence, the integration makes it the largest heart disease dataset accessible for clinical research purposes. The five datasets employed for its curation are provided below

- Cleveland – 303
- Hungarian – 324
- Long beach VA – 280
- Stalog (heart) – 290
- Switzerland– 143

The Heart Disease dataset (mixture of Hungarian, Cleveland, Stalog, long beach VA, Switzerland datasets) comprised a total of 1280 records of patients from Hungary, UK, US, and Switzerland. It encompasses 11 features. These datasets are collected together to make clinical diagnosis accurately.

*Data Preparation*

Data preparation is a very important consideration in Machine Learning. To check for Null values df.isnull().sum() function is used. It returns the number of Null values in the dataset. The present hybrid dataset is free from Null values. This is all about cleaning data and removing outliers from data. The present data is free from Null values. To make the dataset Scalable, StandardScaler () function is used. StandardScaler () function converts values from higher order to lower order. Conversion of values from higher to lower order reduces the possibility of error. It is required to select only those features that are non-correlated among themselves and highly dependent on target variables. Main intention of feature selection method is to remove non-informative or redundant predictors from the model and to retain most useful features in order to predict the target variables. Feature selection is done by ANOVA f-test.

*Model Selection*

A machine learning model determines the output generated after running a machine learning algorithm on the collected data. Choosing a relevant model is important for the task in hand. The present work creates on two different types of models. In first category, every model is separately trained with Twenty-seven different algorithms by using Agent and all algorithms are separately passed in agent model. In second category, Multi Agent model is developed and implemented the best fit algorithm among Twenty Six algorithms.

*Model Training*

One of the important step for machine learning is training. In training, prepared data needs to be passed to a machine learning model to find patterns and make predictions. It results in the model learning from the given data to accomplish the task set. Over the time, with required training, the model becomes better at predicting.
Here the dataset is split into two segments. One is training the model and another one is testing.70-80 percent of data can be kept for training and 20-30 percent for testing. The present work has used 67% data for training and the rest 33% of data for testing.

*Model Evaluation*

This phase checks the performance of the model which is done by testing the performance of the model on previously unseen data. The hybrid data was already splitted and the Model was tested with testing data (33%). Twenty Six different Machine Learning algorithms got tested and predicted by using Agent and predicts Accuracy, Balance Accuracy, ROC-AUC, F1-Score and Execution time for heart disease detection. Those scores of all the algorithm is given below.

```
In [174]: import pandas as pd
db={"Model":"DummyClassifier"  ,"Accuracy":accuracy,"Balance Accuracy":bs,"ROC AUC":rc,"F1-Score":score_dummy,'exec_time':[time_
#result_df=result_df.append(db,ignore_index=True)
result_df=pd.concat([result_df,pd.DataFrame(db)],ignore_index=True)
result_df
```

Out[174]:

| | Model | Accuracy | Balance Accuracy | ROC AUC | F1-Score | exec_time |
|---|---|---|---|---|---|---|
| 0 | Logistic Regr. | 0.853428 | 0.853571 | 0.853571 | 0.853428 | 0.015957 |
| 1 | SVC | 0.924350 | 0.922705 | 0.922705 | 0.924350 | 0.030921 |
| 2 | RF | 0.933806 | 0.932747 | 0.932747 | 0.933806 | 0.254318 |
| 3 | DT | 0.860520 | 0.861912 | 0.861912 | 0.860520 | 0.009971 |
| 4 | GNB | 0.855792 | 0.856225 | 0.856225 | 0.855792 | 0.002992 |
| 5 | knn | 0.867612 | 0.867017 | 0.867017 | 0.867612 | 0.010005 |
| 6 | sgdc | 0.770686 | 0.772694 | 0.772694 | 0.770686 | 0.012001 |
| 7 | perceptron | 0.735225 | 0.731941 | 0.731941 | 0.735225 | 0.001997 |
| 8 | lsvc | 0.853428 | 0.853571 | 0.853571 | 0.853428 | 0.040858 |
| 9 | Nusvc | 0.862884 | 0.861330 | 0.861330 | 0.862884 | 0.035905 |
| 10 | QuadraticDiscriminantAnalysis | 0.843972 | 0.844290 | 0.844290 | 0.843972 | 0.023934 |
| 11 | NearestCentroid | 0.853428 | 0.853952 | 0.853952 | 0.850000 | 0.001994 |
| 12 | CalibratedClassifierCV | 0.853428 | 0.853762 | 0.853762 | 0.853428 | 0.026929 |
| 13 | BernoulliNB | 0.848700 | 0.848645 | 0.848645 | 0.848700 | 0.004986 |
| 14 | PassiveAggressiveClassifier | 0.782506 | 0.782725 | 0.782725 | 0.782506 | 0.002993 |
| 15 | RidgeClassifierCV | 0.851064 | 0.851299 | 0.851299 | 0.851064 | 0.591657 |
| 16 | RidgeClassifier | 0.851064 | 0.851299 | 0.851299 | 0.851064 | 0.075798 |
| 17 | LinearDiscriminantAnalysis | 0.851064 | 0.851299 | 0.851299 | 0.851064 | 0.005987 |
| 18 | AdaBoostClassifier | 0.851064 | 0.851108 | 0.851108 | 0.851064 | 0.233417 |
| 19 | BaggingClassifier | 0.900709 | 0.899597 | 0.899597 | 0.900709 | 0.328122 |
| 20 | ExtraTreesClassifier | 0.936170 | 0.935401 | 0.935401 | 0.936170 | 0.201459 |
| 21 | LGBMClassifier | 0.921986 | 0.922335 | 0.922335 | 0.921986 | 0.423145 |
| 22 | XGBClassifier | 0.931442 | 0.931236 | 0.931236 | 0.931442 | 0.251867 |
| 23 | LabelSpreading | 0.912530 | 0.912483 | 0.912483 | 0.912530 | 0.042886 |
| 24 | LabelPropagation | 0.912530 | 0.912483 | 0.912483 | 0.912530 | 0.028922 |
| 25 | DummyClassifier | 0.520095 | 0.500000 | 0.500000 | 0.520095 | 0.000999 |

Fig 1. Accuracy, Balance Accuracy, ROC-AUC, F1-Score and Execution time is calculated for all the algorithms used to predict Heart disease

Graphical Representation of Accuracy of different Algorithms are given below

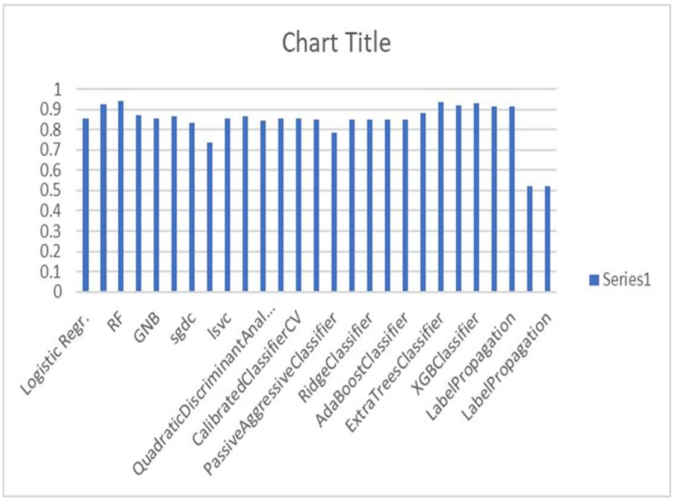


Fig 2. Graphical Representation of Accuracy of different Algorithms are given below

*Parameter Tuning*

Once the model is created and evaluated, this phase is an attempt to check whether the accuracy can be improved in any way. This is done by tuning the parameters present in the model. The present work has developed a Multi Agent System where agents have trained the Hybrid data and executed RandomForestClassifier() Algorithm ( the best algorithm among twenty six algorithms tested before) with almost same accuracy.

*Making Predictions*

Multiple Agents have been created and trained with RandomForestClassifier() as it is proved as the best accurate algorithm for the present dataset. Newly trained Multi Agent System is able to detect the possibility of heart disease with almost the same accuracy.

## IV. Results and Discussion

The research work merges five different datasets to create a comprehensive dataset, which is then used to train multiple intelligent agents, each implementing a different machine learning algorithm. The agents work collaboratively to evaluate the performance of each algorithm and identify the most accurate one for the given dataset. The present work implemented twenty-six ML algorithms with the assistance of Intelligent Agents, comparing their efficiencies to develop a MAS. Three Consecutive prediction details with improvement of accuracy are given below.

```
class Agent1:
    def __init__(self):
        self.model = RandomForestClassifier(n_estimators=20,criterion='gini')

    def train(self, X, y):
        self.model.fit(X, y)
    def predict(self, X):
        return self.model.predict(X)
# Define Agent 2
class Agent2:
    def __init__(self):
        self.model = RandomForestClassifier()
    def train(self, X, y):
        self.model.fit(X, y)
    def predict(self, X):
        return self.model.predict(X)
# Train and predict using agents
agent1 = Agent1()
agent2 = Agent2()
# Train Agent 1
agent1.train(X_train, y_train)
# Get predictions from Agent 1
predictions_agent1 = agent1.predict(X_test)
# Train Agent 2
agent2.train(X_train, y_train)
# Get predictions from Agent 2
predictions_agent2 = agent2.predict(X_test)
print("Agent1 accuracy=",accuracy_score(y_test, predictions_agent1))
print("Agent2 accuracy=",accuracy_score(y_test, predictions_agent2))
```

```
Agent1 accuracy= 0.96875
Agent2 accuracy= 0.96484375
```

## V. Conclusion

Social contribution of this paper is early detection of Heart disease using the Multi Agent System. Multi Agent System is applied with Machine Learning Algorithm for analysis of Heart Disease. This up to the minute work manifests future Scientists and Researchers in two ways. The first one is use of Hybrid dataset i.e. blending of five different standard dataset to improve accuracy of prediction and another one is implementation of Multi Agent System in Machine Learning Algorithm using Jupyter Notebook.RandomForestClassifier() has been proved as the best for diagnosis of heart disease with 93.8% accuracy where as by using Multiagent model, RandomForestClassifier() only has produced accuracy of 96.8 %. Hybrid dataset Implementation and incorporation of Multi Agent System using Machine Learning Algorithm to achieve highest level of accuracy is the novelty of the present research work which can explore a new era in the Field of Artificial Intelligence .This research work can be extended by implementing Multi Agent System by Artificial neural Network to achieve more accuracy for Heart Disease Prediction.

## References


[1] Islam Daoud Suliman,D. Vasumathi,"Prediction of Heart Disease Using Machine Learning Algorithms",IEEE,23 November,2023,10.1109/ICCCNT56998.2023.10308096,https://ieeexplore.ieee.org/xpl/conhome/10306338/proceeding, doi.org/10.1109/ICCCNT56998.2023.10308096.

[2] Chintan Bhatt,ParthPatel,Tarang Ghetia "Effective Heart Disease Prediction Using Machine Learning Techniques" https://doi.org/10.3390/a16020088,https://www.mdpi.com/1999-4893/16/2/88.

[3] Niloy Biswas,Md Mamun Ali,Md Abdur Rahaman,Minhajul Islam,Md. Rajib Mia,"Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques",BioMed Research ,Volume 2023, ArticleID-6864343 ,Int.https://doi.org/10.1155/2023/6864343.

[4] Dhanunjaya Rao,B. Amani Rama Padma Priya,B. Venkata Sai Sowmyal,T. Chandrasekhar,"Machine Learning Algorithms For Prediction Of Heart Disease",International Research Journal of Modernization in Engineering Technology and Science,e-ISSN: 2582-5208,DOI : https://www.doi.org/10.56726/IRJMETS35616.

[5] Nadikatla Chandrasekhar et al.,"Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization",2023, https://doi.org/10.3390/pr11041210.

[6] Huating Sun, Jianan Pan,“Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators”,Journal of Data Analysis and Information Processing, 2023,ISSN Online: 2327-7203,doi.org/10.4236/jdaip.2023.111001.
[7] Md. Imam Hossain,Mehadi Hassan Maruf,Md Asikur Rahman Khan,“Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison”,Springer Link,12 June 2023,https://doi.org/10.1007/s42044-023-00148-7
[8] G. Harinadha Babu,Gunda Jayasree,Chattu Ashika,Vajja Ahalya,Katta Asha Niroopa.,“Heart Disease Prediction System Using Random Forest Technique”https://doi.org/10.22214/ijraset.2023.48764.
[9] Neha Nandal,Rohit Tanwar,“Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis”,This article is included in the Computational Modelling and Numerical Aspects in Engineering collection,https://doi.org/f1000research.com/articles/11-1126.
[10] Janaraniani N,Lipika Goel,Rohit Tanwar,“Heart Attack Prediction using Machine Learning”,IEEE,21-23 September 2022,https://doi.org/10.1109/ICIRCA54612.2022.9985736.
[11] Chaimaa Boukhatem,Heba Yahia youssef,Ali Bou Nassif, “Heart Disease Prediction Using Machine Learning”,IEEE, 18 March ,2022,https://doi.org/10.1109/ASET53988.2022.9734880.
[12] P.Chinnasamy,,“Machine learning based cardiovascular disease prediction”,IEEE,21-23 September 2022,https://doi.org/10.1109/ICIRCA54612.2022.9985736.
[13] Sumaira Ahmed,Salahuddin Shaikh,Farwa Ikram,“Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models”,Volume 2022 | Article ID 3730303 https://doi.org/10.1155/2022/3730303.
[14] Victor Chang,Vallabhanent Rupa Bhavani,Ariel Qianwen Xu,,“An artificial intelligence model for heart disease detection using machine learning algorithms”,Science Direct,Volume,2,Nov 2022,https://doi.org/ 10.1016/j.health.2022.100016.
[15] Mihir Patel,Rahul Patange,Chaitanya Patil,“Predicting Heart Disease Using Machine Learning Algorithms”,IRJET,Volume: 09 Issue: 04 Apr 2022,e-ISSN: 2395-0056,p-ISSN: 2395-0072.
[16] Jian Yang,Jinhan Guan.,“A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm”https://doi.org/10.3390/info13100475.
[17] Harshit Jindal,Sarthak Agaewal,Rishabh Khera,Rachna Jain,Preeti Nagrath,“Heart disease prediction using machine learning algorithms” IOP Conf. Series: Materials Science and Engineering,https://doi.10.1088/1757-899X/1022/1/012072.
[18] Apurv Garg.Bhartendu Sharma,Rijwan Khan,“Heart disease prediction using machine learning techniques”,IOP Conf. Series: Materials Science and Engineering,doi.org/10.1088/1757-899X/1022/1/012046.
[19] Elwahsh H,“A new smart healthcare framework for real-time heart disease detection based on deep and machine learning”,doi.org/10.1088/1757-899X/1022/1/012046.