Ensemble Learning Strategies for Enhancing Predictive Models in Cardiology

Group Name - Ensemble Medi-Tech

- AYANTANU LAHA
- RAJARSHI SAHA

Content:

GOAL :

Our aim is to develop a predictive model using machine learning techniques to accurately identify the presence or absence of heart disease based on various patient attributes and medical measurements.

By analyzing a comprehensive dataset containing some features, our goal is to train an **Ensemble model** capable of effectively predicting the likelihood of heart disease occurrence.

IMPORTANCE


The Economic Times | Panache

English Edition ▾ | Today's ePaper

Subscribe Sign In

Steal Deal on ETPrime

Home ETPrime Markets News Industry Rise Politics Wealth Mutual Funds Tech Careers Opinion NRI Panache Spotlight

Panache • ET Magazine

World Heart Day: 27% of deaths in India are caused by cardiovascular diseases

ET Online • Last Updated: Sep 29, 2023, 09:45:00 AM IST

FOLLOW US SHARE FONT SIZE SAVE PRINT COMMENT


bt Business Today

HOME MAGAZINE BT TV MARKET TODAY TECH TODAY MONEY TODAY MUTUAL FUNDS PMS TODAY INDUSTRY

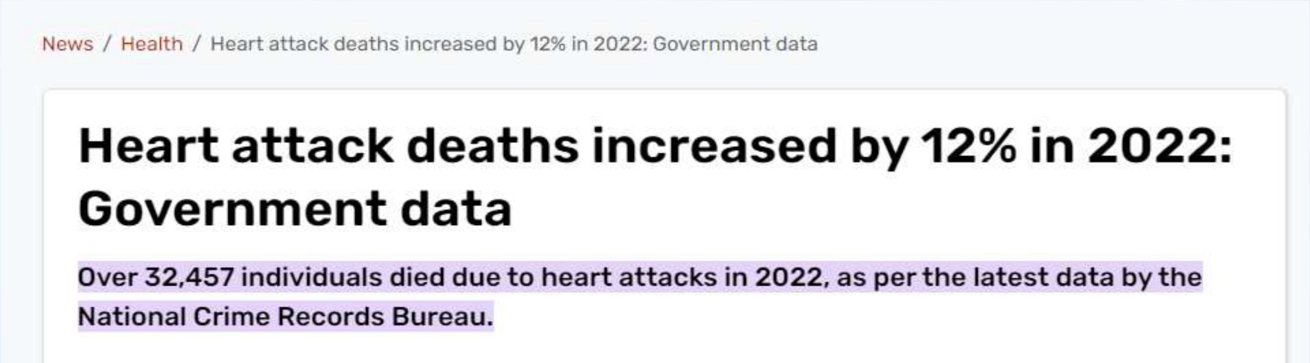
Budget 2024 Immersives UPSTART India BTShorts Weather BT Bazaar

News / LATEST / In Focus / 'Rising heart attacks among youth who had severe Covid,' warns health minister ...

Feedback

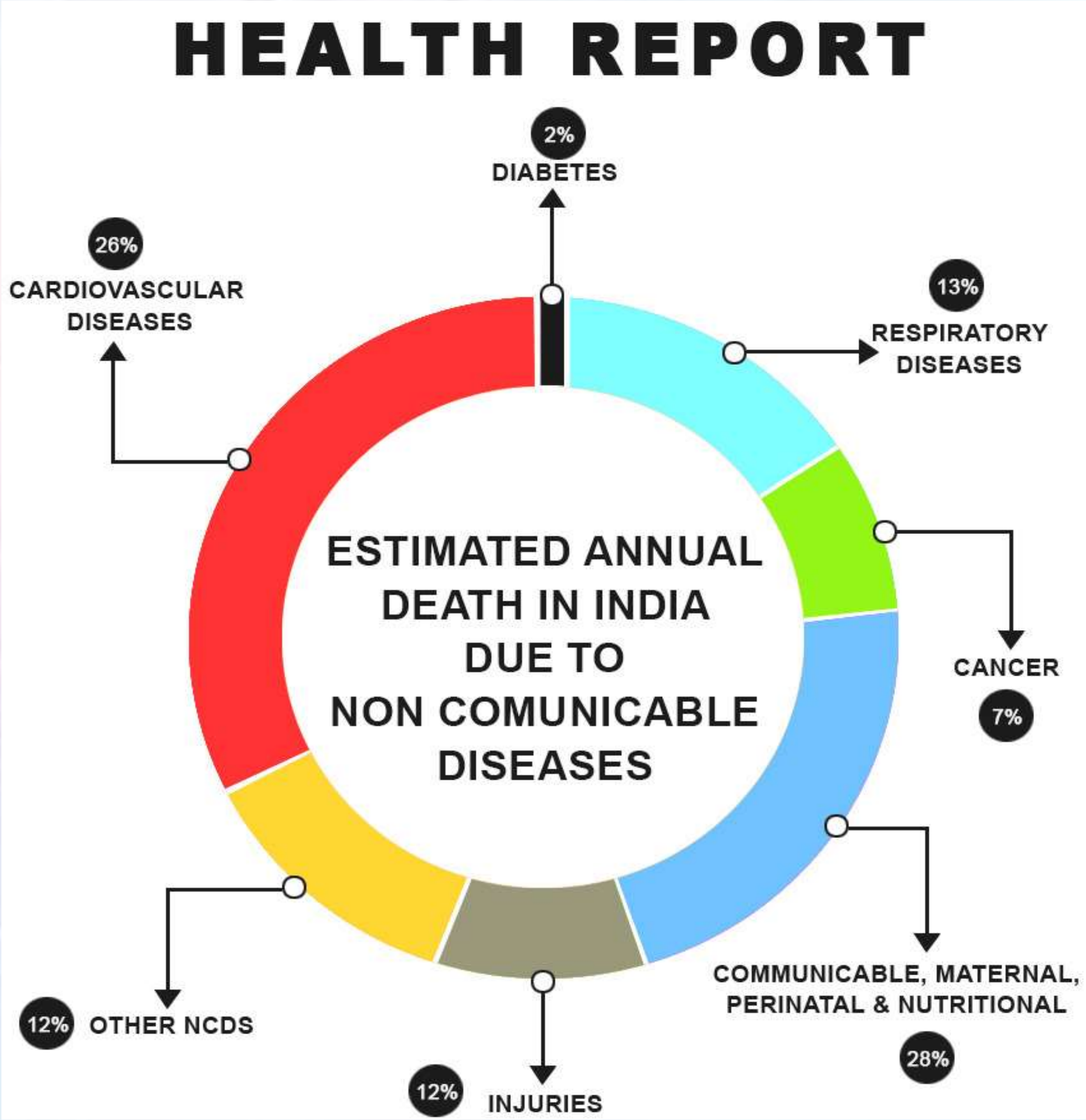
'Rising heart attacks among youth who had severe Covid,' warns health minister Mansukh Mandaviya

Mandaviya said at the sidelines of an event in Gujarat that a study by ICMR shows the rise of heart attacks among youth who have recovered from severe Covid-19 infection


News / Health / Heart attack deaths increased by 12% in 2022: Government data

Heart attack deaths increased by 12% in 2022: Government data

Over 32,457 individuals died due to heart attacks in 2022, as per the latest data by the National Crime Records Bureau.



METHODOLOGY:

- Raw dataset is used to maintain the authenticity, because pre-processing may tamper the raw dataset which may cause an impact in our study.
- Exploratory data analysis is done to understand the data’s characteristics.
- Ensemble learning technique is applied to develop predictive models.
- Finally the performance is compared with the previous works.

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
Alfredo Daza , Juana Bobadilla, Juan Carlos Herrera , Angelica Medina , Nemias Saboya , Karoline Zavaleta , Segundo Siguenas	Stacking ensemble based hyperparameters to diagnosing of heart disease: Future works [5]	**Stacking 1**: Level 0 (Decision Tree, Random Forest, Bayesian Networks, and Gradient Boosting) and Level 1 Logistic Regression **Stacking 2:** Level 0 (Decision Tree, Random Forest, Bayesian Networks, and Gradient Boosting) and Level 1 Random Forest; **Stacking 3:** Level 0 (Decision Tree, Random Forest, Bayesian Networks, and Gradient Boosting) and level 1 SVM; **Stacking 4:** Level 0 (Decision Tree, Random Forest, Bayesian Networks, and Gradient Boosting) and Level 1 KNN	Heart Failure Prediction Dataset (1190 instances) - Kaggle	Stacking 1 (Logistic Regression) combined with oversampling and AdaBoost-SVM with hyperparameter tuning exhibited superior performance. This combination achieved a high accuracy of **88.24%** and an impressive ROC Curve of 92.00%.	2024

Daza A., Bobadilla J. , Herrera J. C., Medina A., Saboya N., Zavaleta K., Siguenas S., Stacking ensemble based hyperparameters to diagnosing of heart disease: Future works, Results in Engineering, Volume 21,2024,101894, ISSN 2590-1230, https://doi.org/10.1016/j.rineng.2024.101894

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
Mohd Syafiq Asyraf Suhaimi , Nor Azuana Raml, Noryanti Muhammad	Heart disease prediction using ensemble of k-nearest neighbour, random forest and logistic regression method [4]	Introduced a novel ensemble approach, specifically employing stacking, to predict heart disease. **They utilized logistic regression as the meta-level classifier, alongside random forest and k-nearest neighbour methods as the base-level classifiers.**	Heart Disease Dataset (1025 instances)- Kaggle	The combined approach of Random Forest, logistic regression, and k-NN achieved a comparable accuracy of **82.42%** with Random Forest alone. Notably, k-NN exhibited the lowest accuracy among all methods.	2024

Suhaimi M. S. A, Ramli N. A., and Muhammad N. "Heart disease prediction using ensemble of k-nearest neighbour, random forest and logistic regression method." *AIP Conference Proceedings*. Vol. 2895. No. 1. AIP Publishing, 2024. https://doi.org/10.1063/5.0192203

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
G. Manikandan, B. Pragadeesh, V. Manojkumar, A.L. Karthikeyan, R. Manikandan, Amir H. Gandomi	Classification models combined with Boruta feature selection for heart disease prediction [9]	They compared the performance of **logistic regression**, **decision tree**, and **support vector machine (SVM)** methods with and without **Boruta feature selection** (automatically determines any thresholds and returns features that are most meaningful in the dataset (Algorithm).	Cleveland Clinic Heart Disease Dataset (303 instances) - Kaggle	Logistic Regression - 88.52%.	2024

G. Manikandan, B. Pragadeesh, V. Manojkumar, A.L. Karthikeyan, R. Manikandan, Amir H. Gandomi,Classification models combined with Boruta feature selection for heart disease prediction,Informatics in Medicine Unlocked,Volume 44,2024,101442,ISSN 2352-9148,
https://doi.org/10.1016/j.imu.2023.101442

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
Ahmed Al Ahdal, Moaiad Ahmad Khder,Manik Rakhra,Rahul R. Rajendran,Binit Patel,Rituraj Jain, Farrukh Arslan, Balaji Ramkumar Rajagopal	Monitoring Cardiovascular Problems in Heart Patients Using Machine Learning [1]	Random forests and extreme gradient boost	UCI Heart Disease Data (920 instances) – Kaggle	Random forest - 96.72%,	2023

A. A. Ahdal, M. A. Khder,M. Rakhra,R. R. Rajendran,B. Patel,R. Jain, F. Arslan, B. R. Rajagopal , "Monitoring Cardiovascular Problems in Heart Patients Using Machine Learning", 2040-2295, Journal of Healthcare Engineering, , Hindawi, 10.1155/2023/9738123, https://doi.org/10.1155/2023/9738123

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
Subhash Mondal, Yachang Omo, Ranjan Maity, Soumadip Ghosh, Amitava Nag	An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches [10]	Five machine learning classifiers - **Logistic Regression, Support Vector Machine, Random Forest, Decision Tree**, and **Extreme Gradient Boost** - were used to create the initial prediction model, and their performance was **optimized using hyperparameter tuning techniques** like **RandomizedSearchCV and GridSearchCV**.	Dataset 1 : Cleveland Cardiovascular Disease Dataset (1190 instances) - IEEE Data Dataset 2: Cardiovascular_Disease_Dataset (1000 instances) - Mendeley Data	Dataset 1 : 96.00 % Dataset 2 : 96.88 %	2024

S. Mondal, R. Maity, Y. Omo, S. Ghosh and A. Nag, "An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches," in *IEEE Access*, vol. 12, pp. 7255-7270, 2024, doi: 10.1109/ACCESS.2024.3350996.

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
B.Rupa Devi, U.Sivaji, Thammisetty Swetha, Dr.J.Avanija, Dr.A.Suresh, Dr. K. Reddy Madhavi	Advanced Cardiovascular Disease Prediction: A Comparative Analysis of Ensemble Stacking and Deep Neural Networks. [7]	A neural network model is defined using TensorFlow/Keras, and two instances ('network1' and 'network2') are created. Both networks are trained on the data for a specified number of epochs. Predictions are obtained from both networks on the validation set and stacked horizontally to create features for the meta-model. A meta-model using the XGBoost classifier is defined and trained on the stacked predictions from the neural networks. Predictions on the test set are made using both 'network1' and 'network2,' stacked horizontally, and the meta-model (XGBoost) is used to make final ensemble predictions on the test set.	They gathered a dataset with several characteristics some crucial factors of heart disease.	The proposed Model achieved accuracy of **93%**	2023

Devi, B. R. ., Sivaji, U. ., Swetha, T. ., Avanija, J. ., Suresh, A. ., & Madhavi, K. R. . (2023). "Advanced Cardiovascular Disease Prediction: A Comparative Analysis of Ensemble Stacking and Deep Neural Networks.", International Journal of Intelligent Systems and Applications in Engineering, 12(6s), 46–55. https://ijisae.org/index.php/IJISAE/article/view/3937

2. Literature Survey

Author	Paper	Method	Dataset	Result	Year
Abdullah Alqahtani,Shtwai Alsubai,Mohemmed Sha,Lucia Vilcekova, Talha Javed	Cardiovascular Disease Detection using Ensemble Learning. [3]	Six classification method algorithms is used to predict cardiovascular disease. They proposed an ensemble learning method for detecting CV disease including several classifiers in voting-based decision making. The Random Forest **features extraction approach** is applied to identify the most informative feature. Besides machine learning algorithms they used a neural network known as KDNN, a robust medical disease detection approach.	Cardiovascular Disease dataset (70,000 instances) - Kaggle	ML ensemble classifier achieved an accuracy of **88.70 %** The stacked classifier achieved an accuracy of **86.49%.** Among all the classifier algorithms RF classifier obtained the best accuracy of **88.65%.**	2022

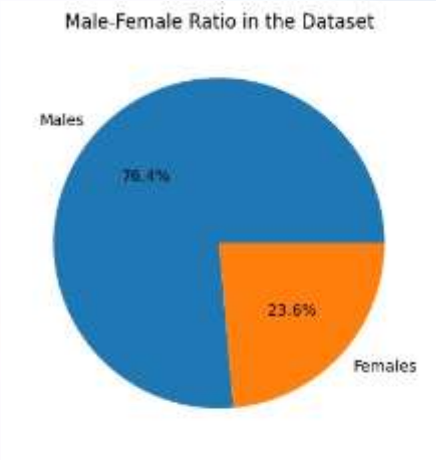
A. Alqahtani,S. Alsubai,M. Sha,L. Vilcekova, T. Javed , "Cardiovascular Disease Detection using Ensemble Learning. ", 1687-5265, Computational Intelligence and Neuroscience, Hindawi, 10.1155/2022/5267498, https://doi.org/10.1155/2022/5267498

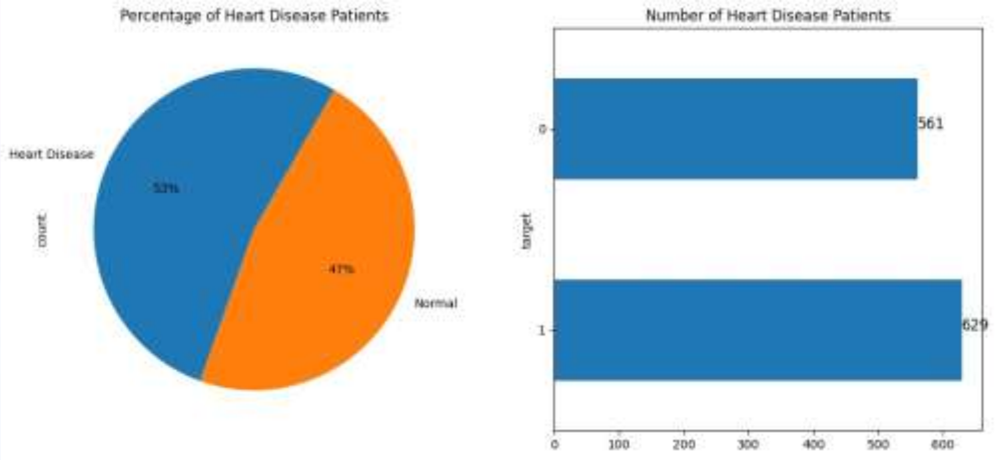
3. Data Acquisition

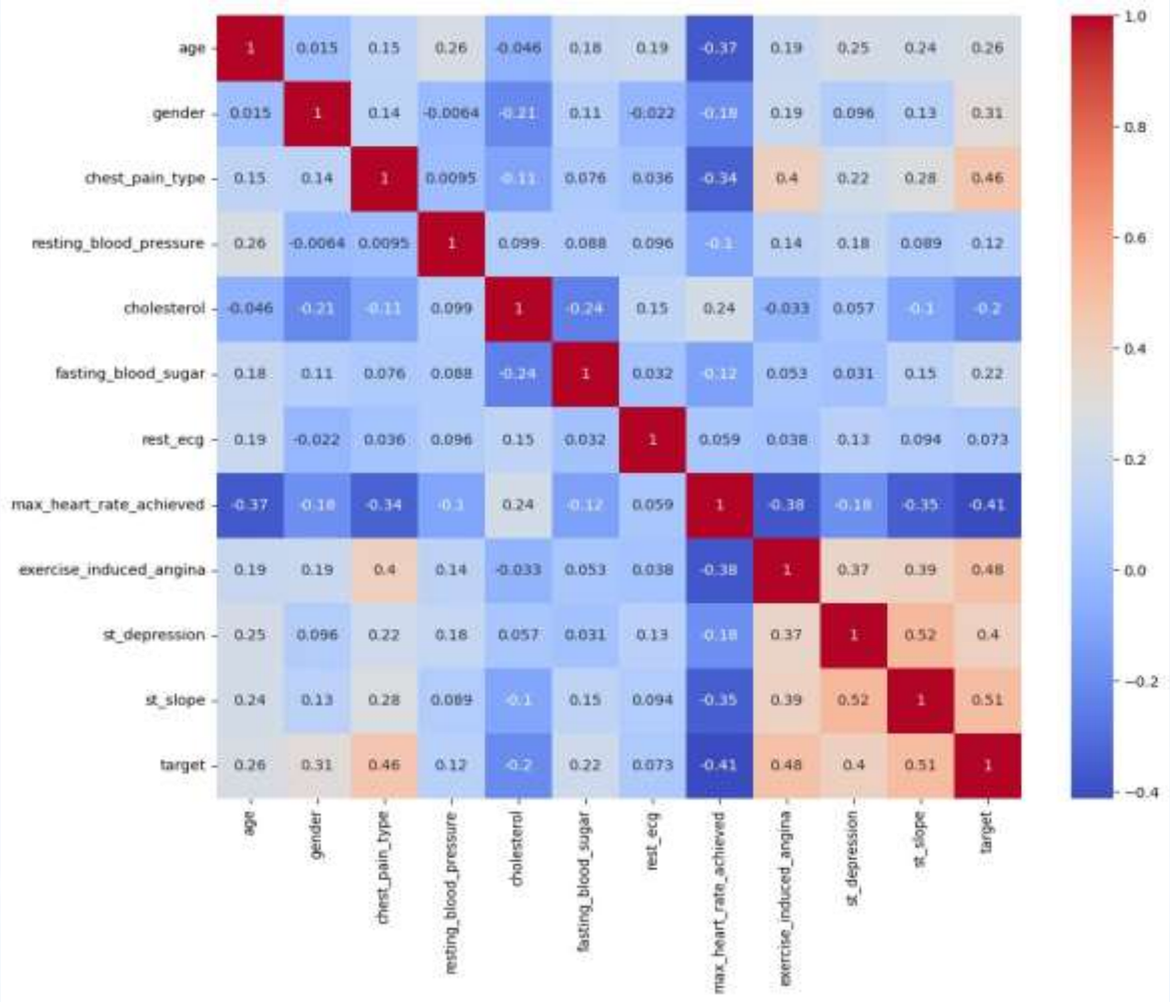
Dataset 1	Dataset 2	Dataset 3	Dataset 4
• 1190 observations • 12 attributes (11 predictor variables and 1 predicted variable) • Dataset 1	• 1025 observations • 14 attributes (13 predictor variables and 1 predicted variable) • Dataset 2	• 1000 observations • 14 attributes (13 predictor variables and 1 predicted variable) • Dataset 3	• 70000 observations • 12 attributes (11 predictor variables and 1 predicted variable) • Dataset 4

4. Data Distribution

4.1 Dataset 1 :



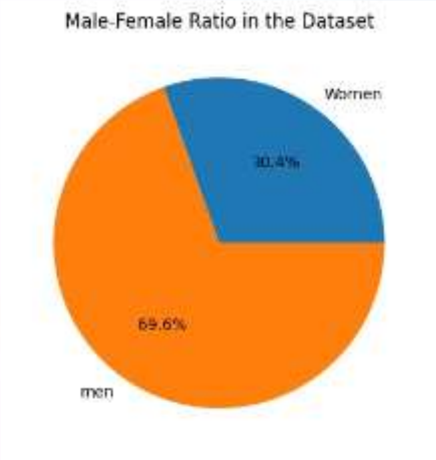


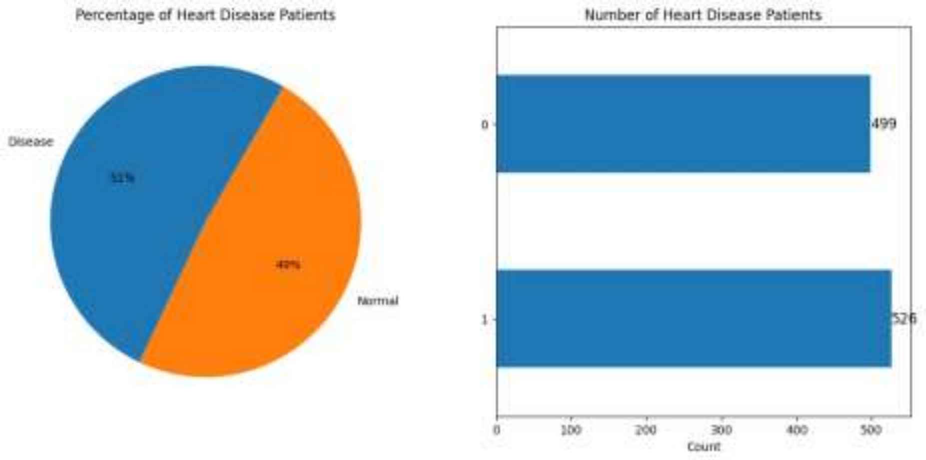


Confusion Matrix

4. Data Distribution

4.2 Dataset 2 :

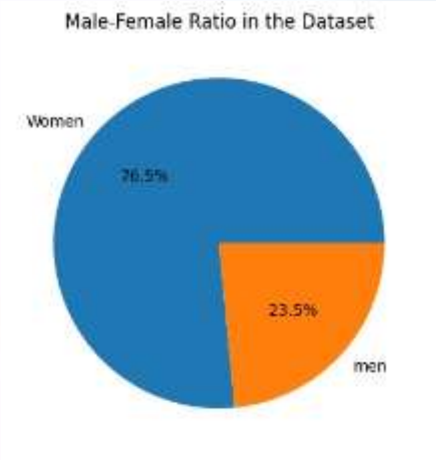


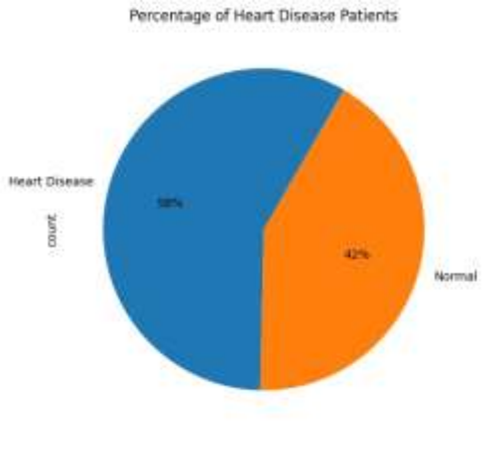


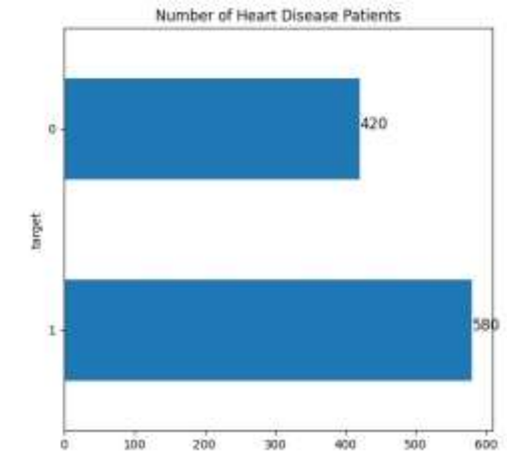
Confusion Matrix

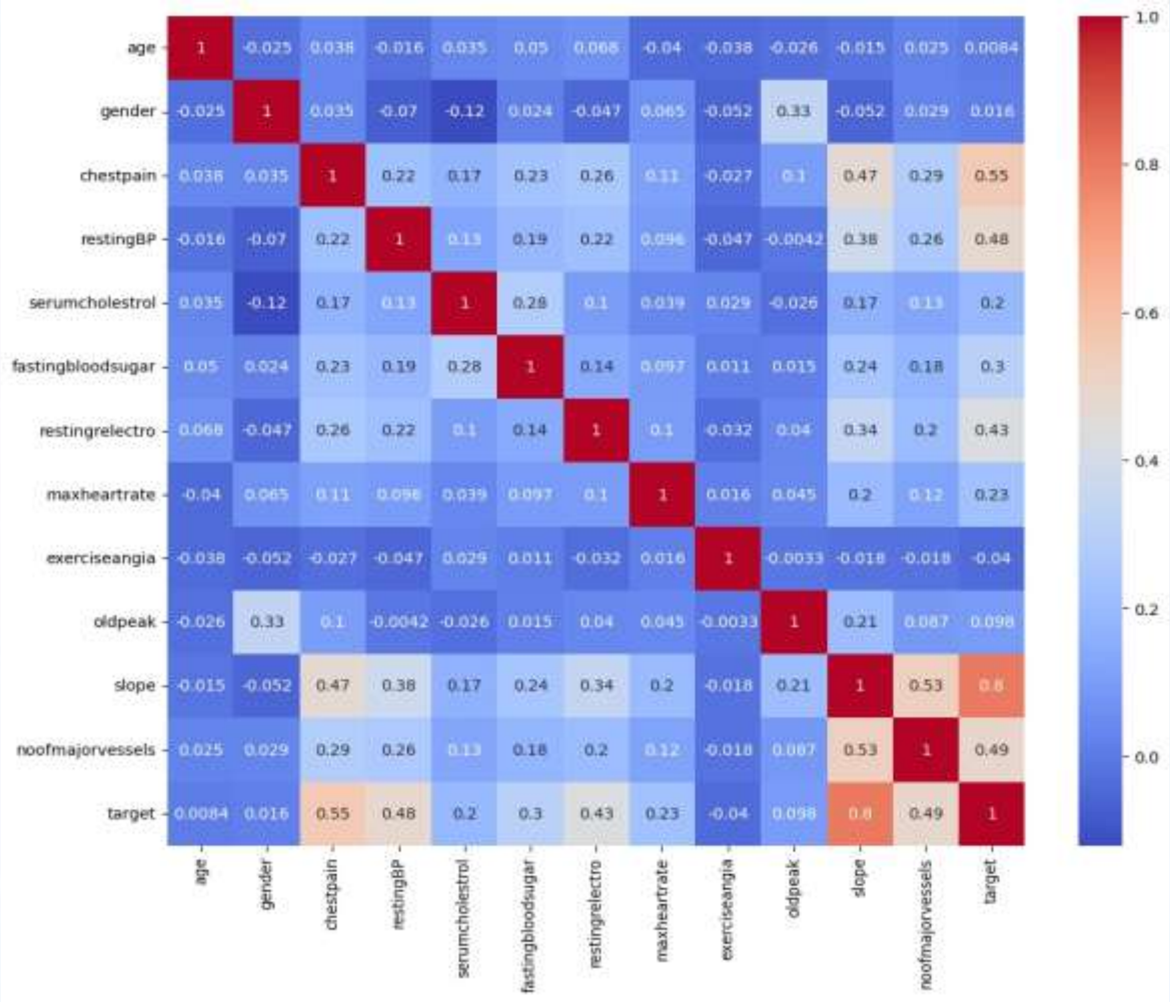
4. Data Distribution

4.3 Dataset 3 :





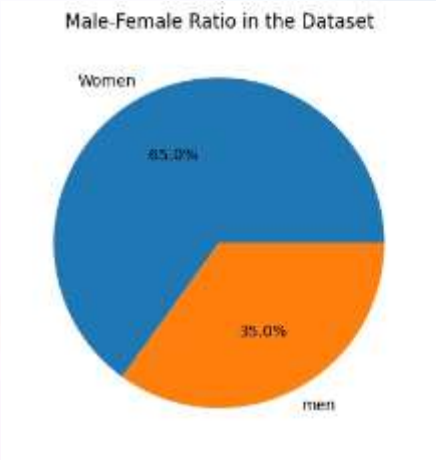


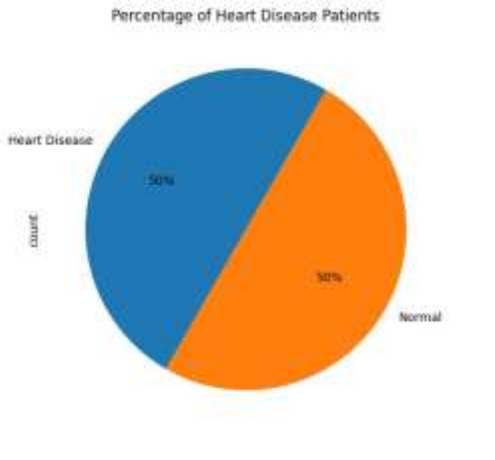


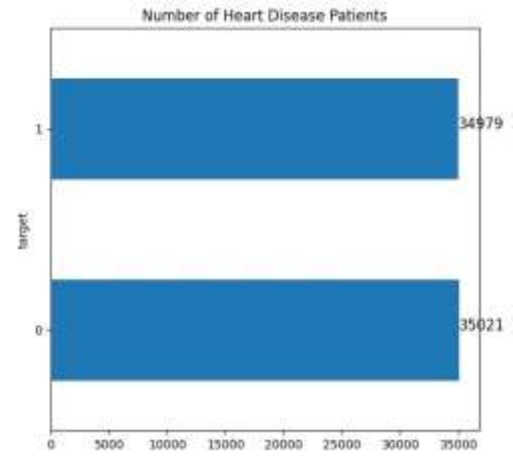
Confusion Matrix

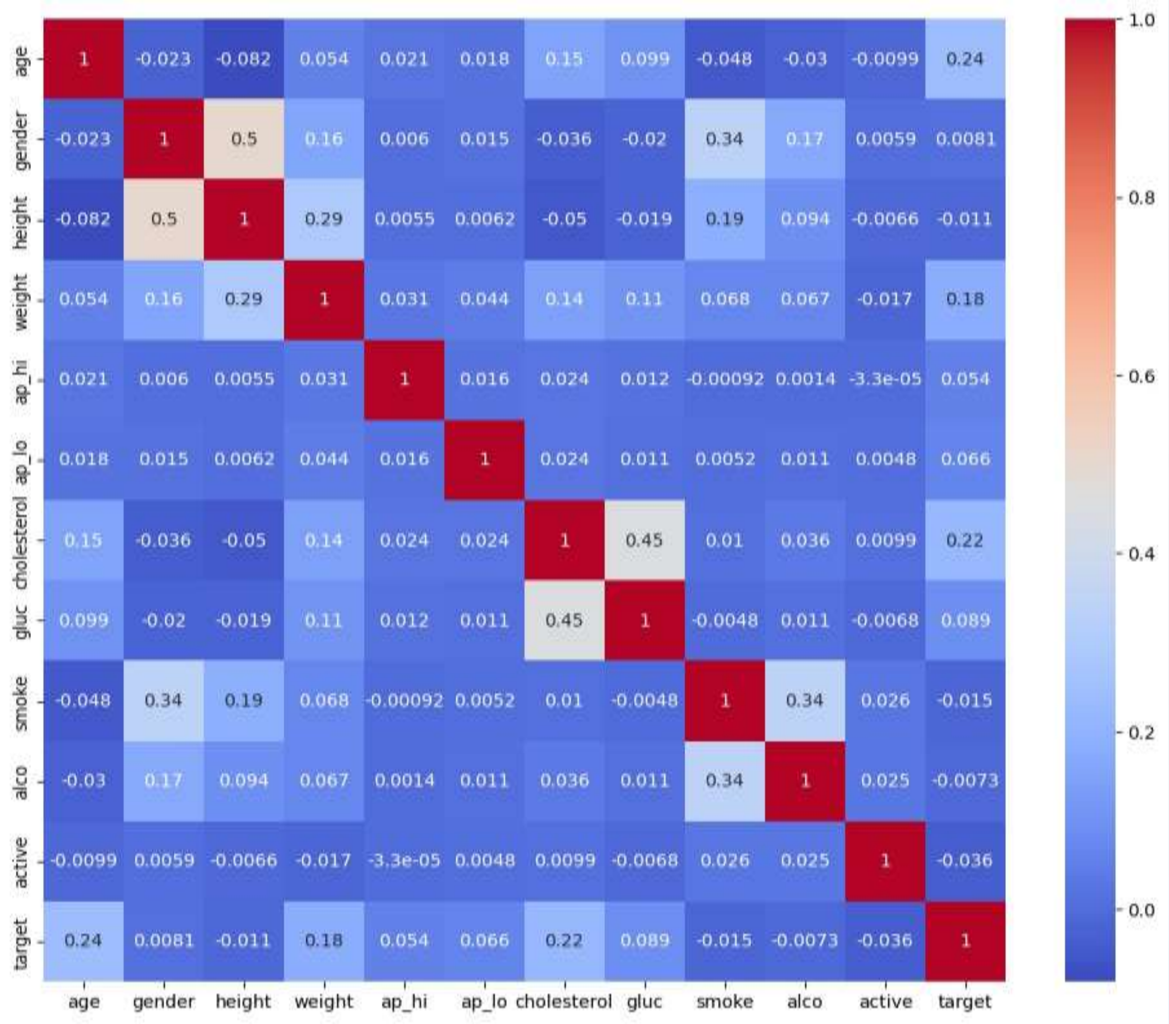
4. Data Distribution

4.4 Dataset 4 :









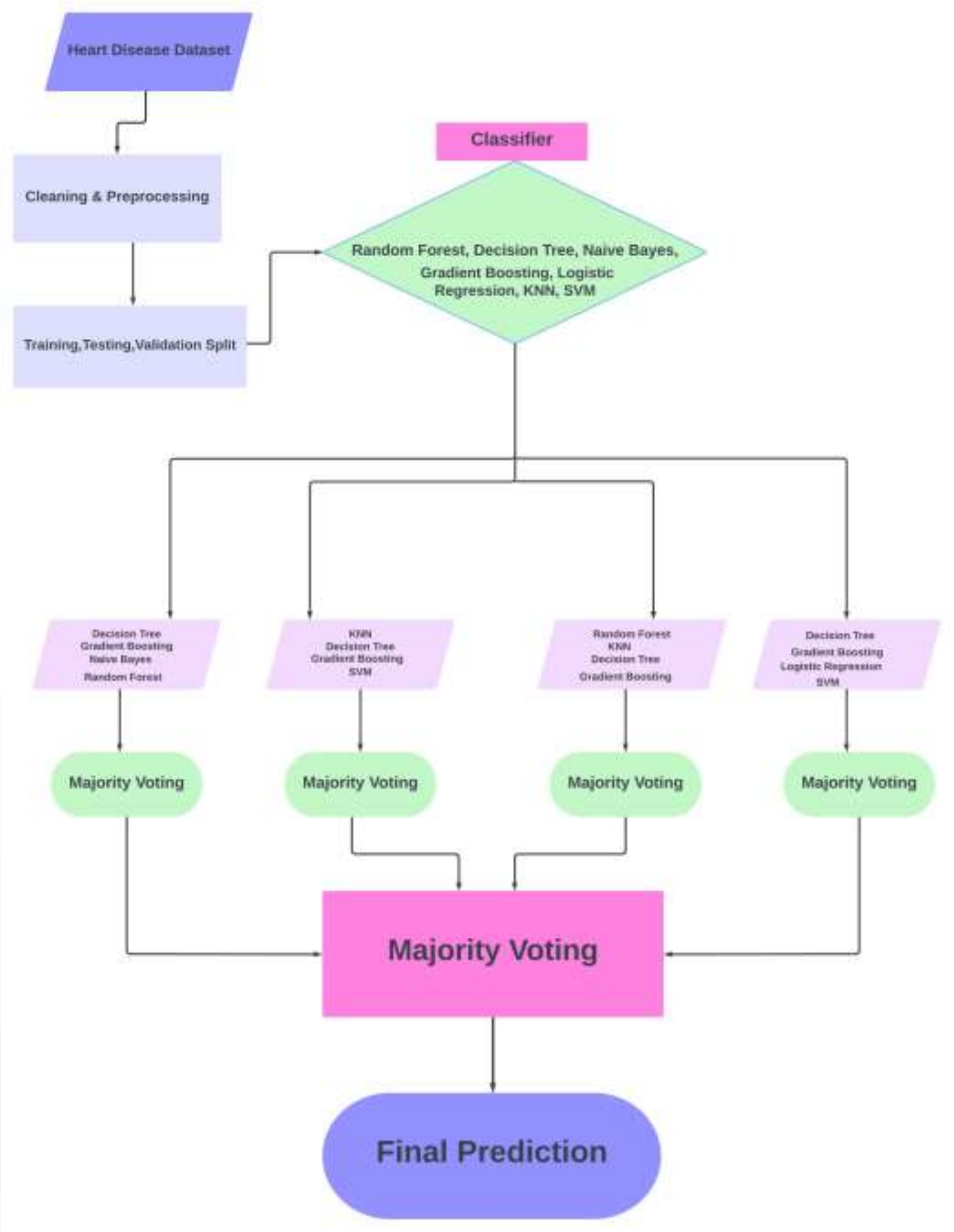
Confusion Matrix

5. Model Selection & Ensemble Modelling

5.1 Data Split :

- Training set: 64% of the original dataset
- Validation set: 16% of the original dataset
- Testing set: 20% of the original dataset

5.2 Flow Chart:



5.3 Stacking Classifier

Install User Guide API Examples Community More

Go

Prev Up Next

scikit-learn 1.4.2
Other versions

Please **cite us** if you use the software.

sklearn.ensemble.StackingClassifier
`StackingClassifier`
Examples using
`sklearn.ensemble.StackingClassi`

sklearn.ensemble.StackingClassifier

class `sklearn.ensemble.`**`StackingClassifier`**(*estimators*, *final_estimator=None*, ***, *cv=None*, *stack_method='auto'*, *n_jobs=None*, *passthrough=False*, *verbose=0*) [source]

Stack of estimators with a final classifier.

Stacked generalization consists in stacking the output of individual estimator and use a classifier to compute the final prediction. Stacking allows to use the strength of each individual estimator by using their output as input of a final estimator.

Note that `estimators_` are fitted on the full `X` while `final_estimator_` is trained using cross-validated predictions of the base estimators using `cross_val_predict`.

Read more in the User Guide.

New in version 0.22.

Parameters:

estimators : *list of (str, estimator)*
Base estimators which will be stacked together. Each element of the list is defined as a tuple of string (i.e. name) and an estimator instance. An estimator can be set to 'drop' using `set_params`.

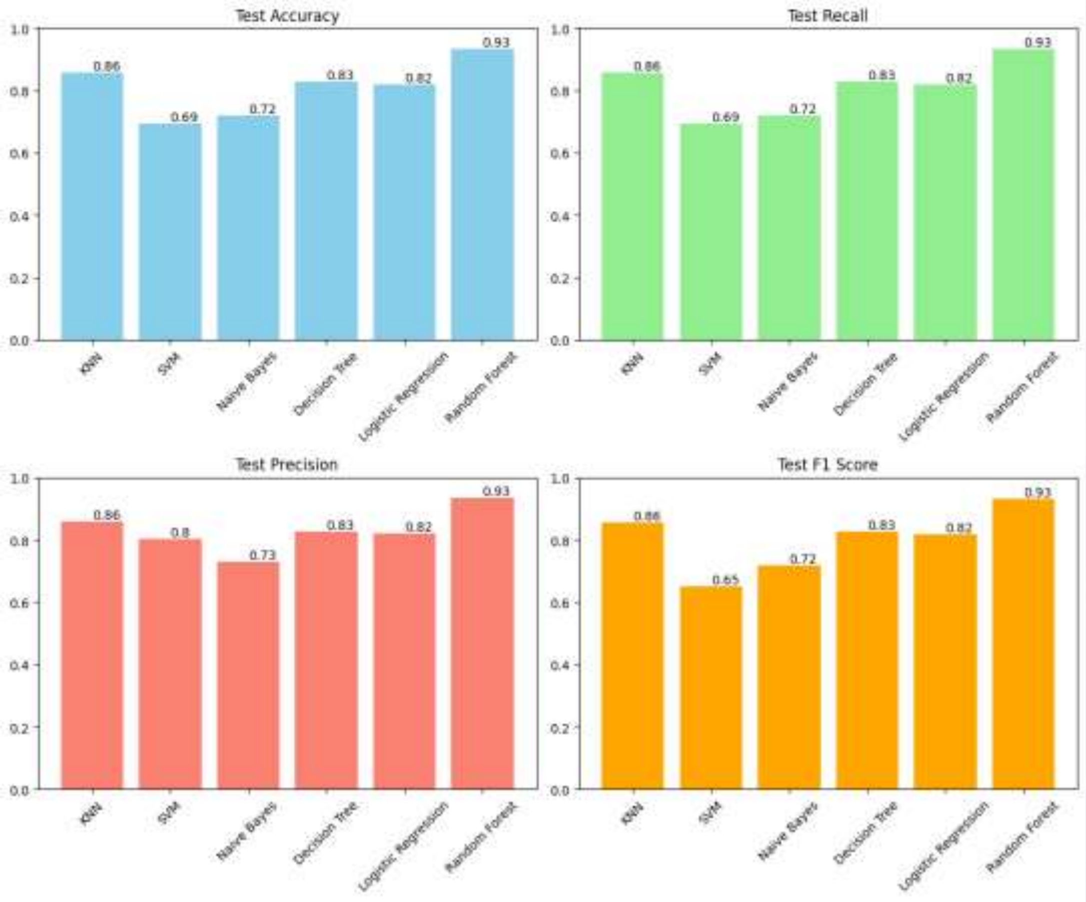
The type of estimator is generally expected to be a classifier. However, one can pass a regressor for some use case (e.g. ordinal regression).

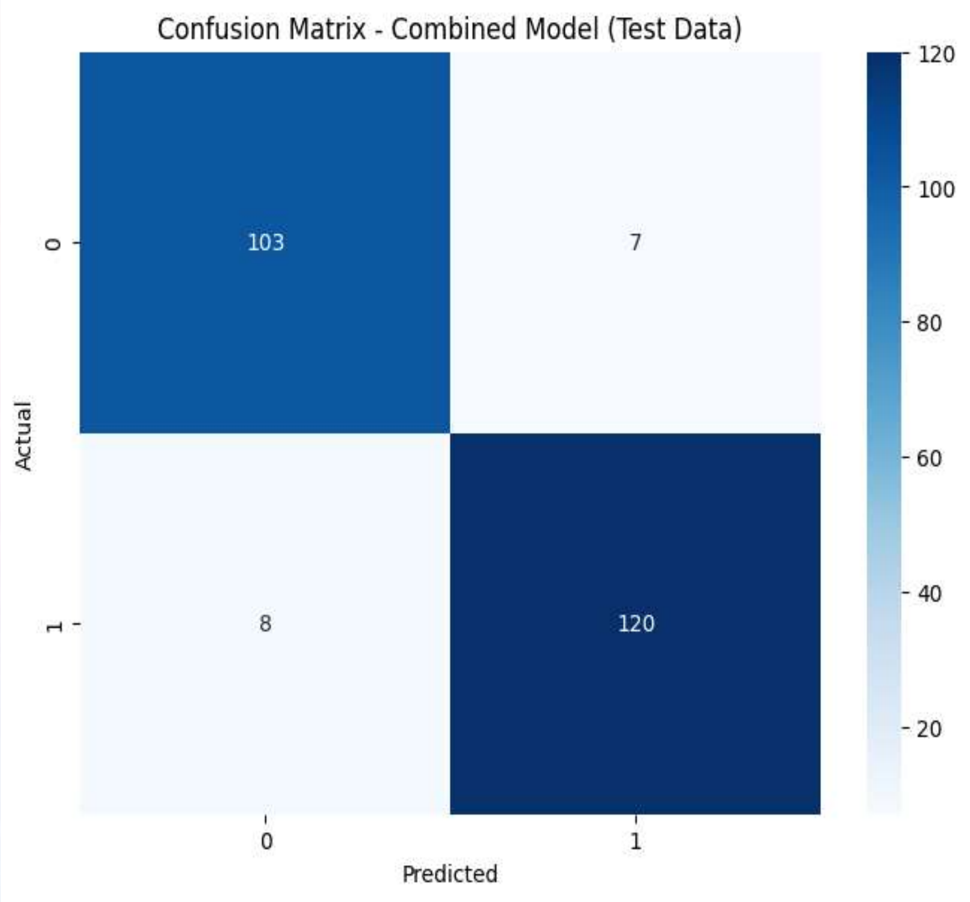
final_estimator : *estimator, default=None*
A classifier which will be used to combine the base estimators. The default classifier is a `LogisticRegression`.

```
rf=RandomForestClassifier(criterion='entropy', max_features='sqrt')
dt= DecisionTreeClassifier (criterion='entropy', max_depth=5, min_samples_leaf=5,
splitter='random')
gb=GradientBoostingClassifier(max_depth=4, min_samples_leaf=20)
gnb-GaussianNB(var_smoothing=0.004328761281083057)
svm=SVC(C=100, gamma 'auto', probability=True)
knn=KNeighbors Classifier (metric='manhattan', n_neighbors=9, weights='distance')
lr=LogisticRegression (C=0.615848211066026, max_iter=500, penalty='11', solver='liblinear')
```

6. Experimental Results

6.1 Dataset 1:



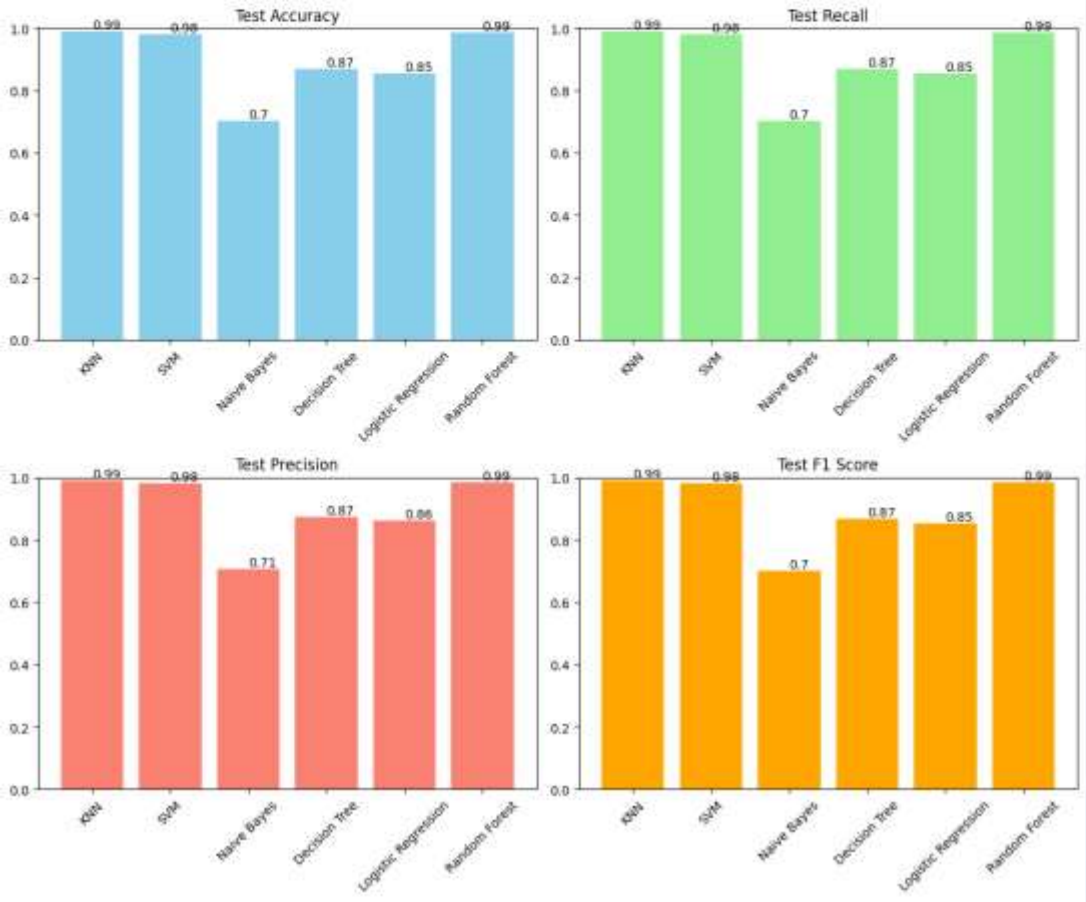


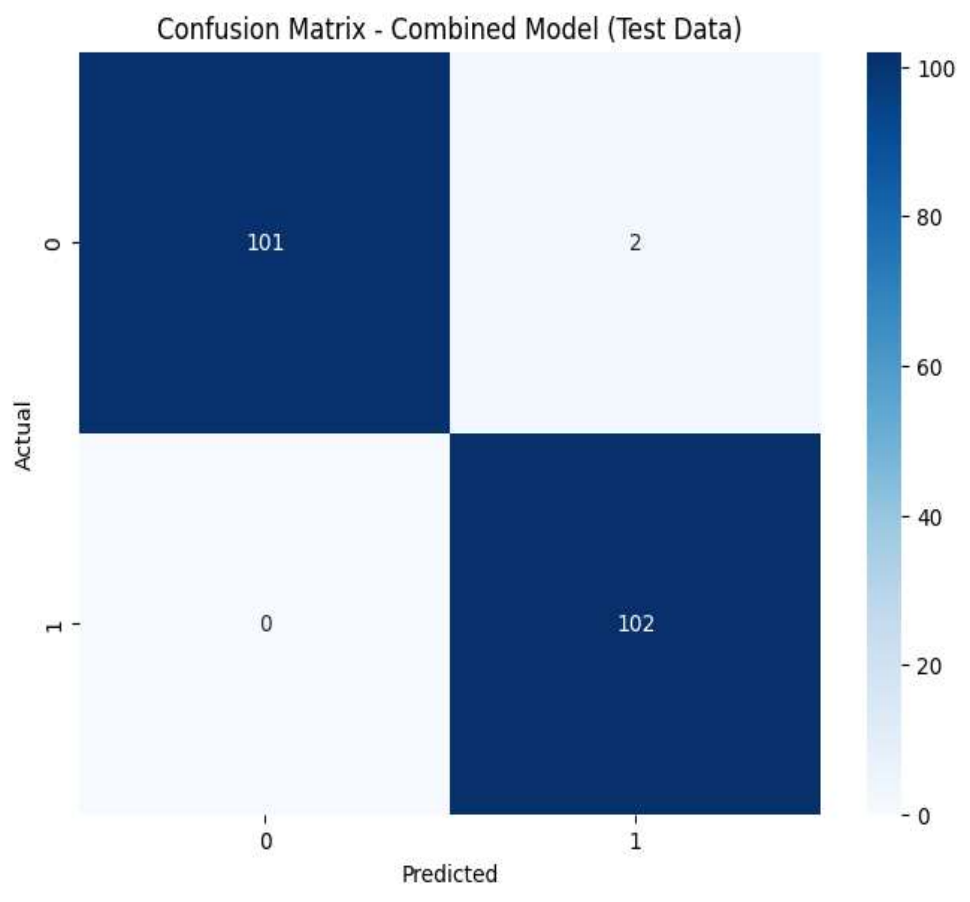
Dataset 1 (1190)

Model Type	Classification Algorithm	Accuracy		Precision		Recall		F1 Score	
		Validation	Test	Validation	Test	Validation	Test	Validation	Test
Individual Model	K-Nearest Neighbors	80.63%	85.71%	80.81%	85.75%	80.63%	85.71%	80.54%	85.68%
	Support Vector Machine	68.06%	69.33%	80.16%	80.47%	68.06%	69.33%	63.74%	65.11%
	Naïve Bayes Classifier	73.40%	71.85%	73.62%	72.91%	73.29%	71.85%	73.29%	71.82%
	Decision Tree Classifier	81.15%	82.77%	81.17%	82.76%	81.15%	82.77%	81.12%	82.75%
	Logistic Regression Classifier	81.16%	81.93%	81.30%	81.94%	81.15%	81.93%	81.08%	81.94%
	Random Forest Classifier	89%	93.28%	89.02%	93.41%	89%	93.28%	88.99%	93.29%
Stack Classifier	Stacking 1	89.53%	93.70%	89.66%	93.70%	89.53%	93.70%	89.50%	93.70%
	Stacking 2	88.48%	91.60%	88.48%	91.73%	88.48%	91.60%	88.48%	91.61%
	Stacking 3	89.53%	91.18%	89.54%	91.28%	89.53%	91.18%	89.52%	91.19%
	Stacking 4	90.05%	92.02%	90.08%	92.02%	90.05%	92.02%	90.04%	92.02%
Majority Voting	Combined Stacking Model	90.58%	93.68%	90.63%	93.70%	90.57%	93.70%	90.56%	93.70%

6. Experimental Results

6.2 Dataset 2:

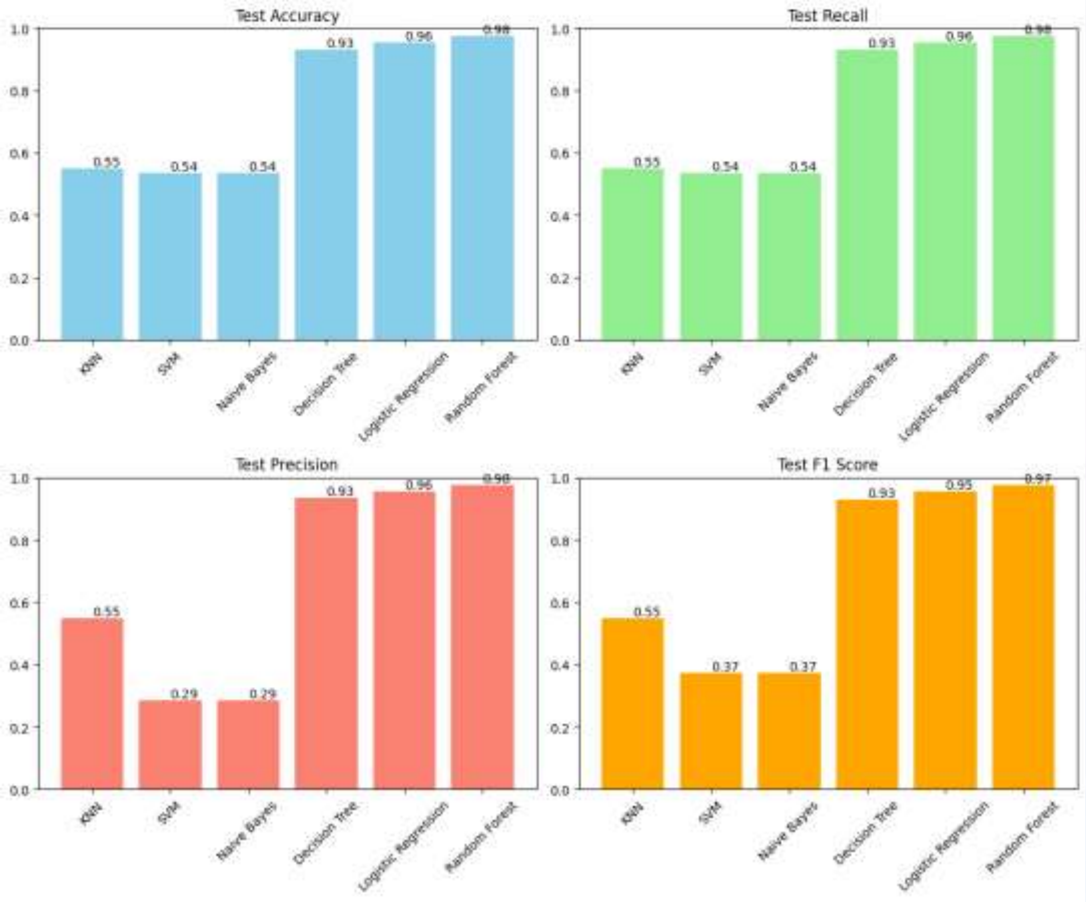


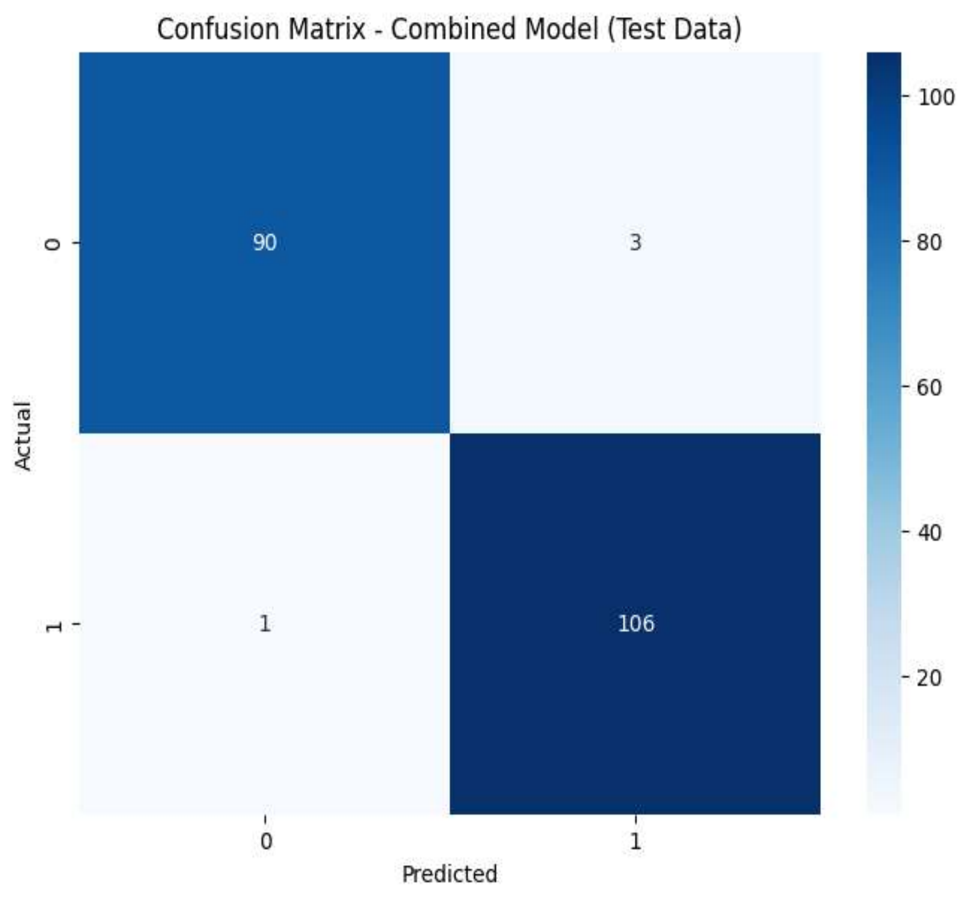


		Dataset 2 (1025)							
Model Type	Classification Algorithm	Accuracy		Precision		Recall		F1 Score	
		Validation	Test	Validation	Test	Validation	Test	Validation	Test
Individual Model	K-Nearest Neighbors	96.95%	99.02%	97.12%	99.04%	96.95%	99.02%	96.95%	99.02%
	Support Vector Machine	96.34%	98.04%	96.59%	98.12%	96.34%	98.05%	96.34%	98.05%
	Naïve Bayes Classifier	70.73%	70.24%	70.75%	70.59%	70.73%	70.24%	70.73%	70.13%
	Decision Tree Classifier	88.41%	86.83%	88.42%	87.26%	88.42%	86.83%	88.41%	86.79%
	Logistic Regression Classifier	81.09%	85.37%	81.22%	86.26%	81.09%	85.36%	81.08%	85.28%
	Random Forest Classifier	97%	98.54%	96.96%	98.54%	97%	98.54%	96.95%	98.54%
Stack Classifier	Stacking 1	98.78%	99.02%	98.81%	99.04%	98.78%	99.02%	98.78%	99.02%
	Stacking 2	98.78%	99.02%	98.81%	99.04%	98.78%	99.02%	98.78%	99.02%
	Stacking 3	96.95%	98.54%	96.96%	98.54%	96.95%	98.54%	96.95%	98.54%
	Stacking 4	98.78%	99.02%	98.81%	99.04%	98.78%	99.02%	98.78%	99.02%
Majority Voting	Combined Stacking Model	98.78%	99.02%	98.80%	99.04%	98.78%	99.02%	98.78%	99.02%

6. Experimental Results

6.3 Dataset 3:



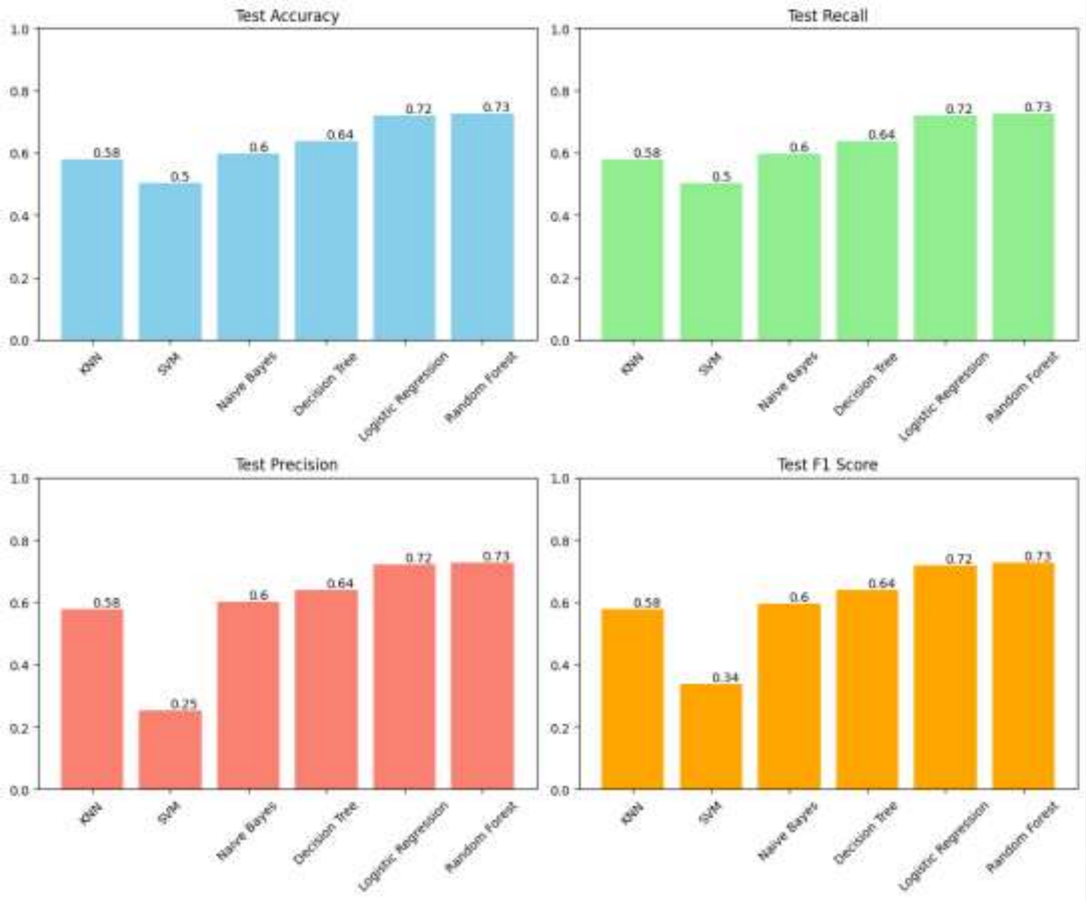


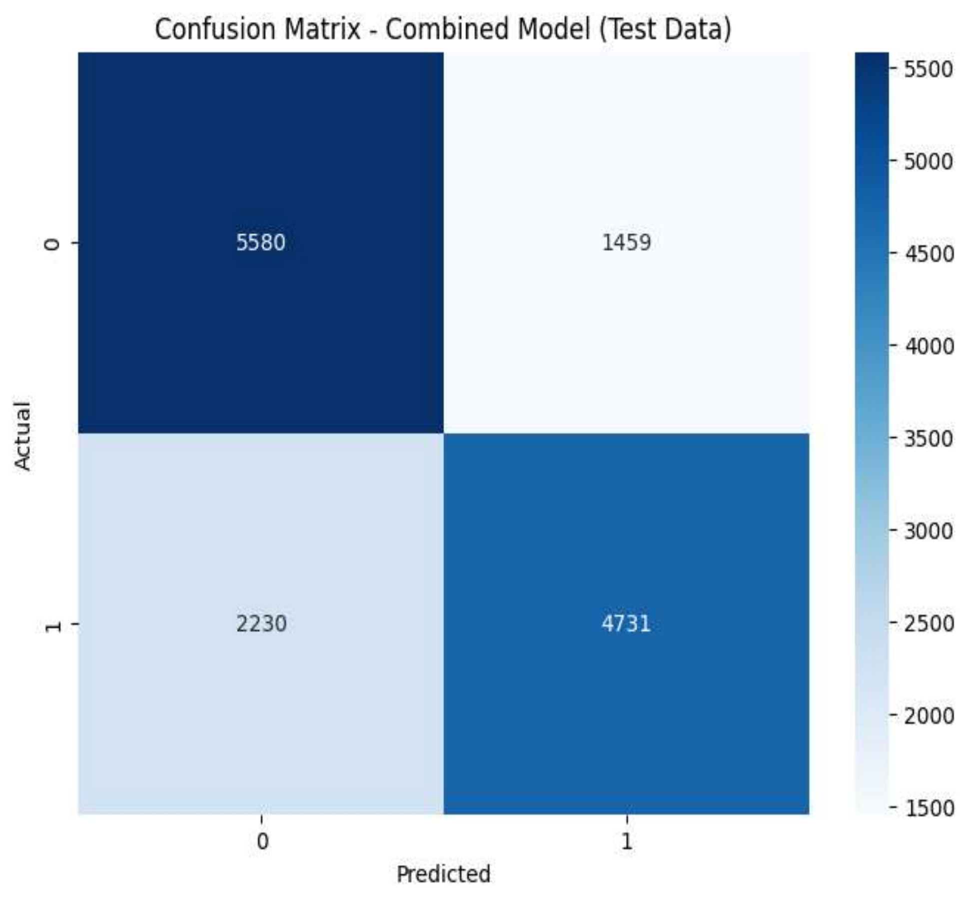
Dataset 3 (1000)

Model Type	Classification Algorithm	Accuracy		Precision		Recall		F1 Score	
		Validation	Test	Validation	Test	Validation	Test	Validation	Test
Individual Model	K-Nearest Neighbors	49.37%	55.00%	47.44%	54.88%	49.38%	55.00%	48.19%	54.91%
	Support Vector Machine	61.87%	53.50%	38.28%	28.62%	61.87%	53.50%	47.30%	37.29%
	Naïve Bayes Classifier	61.87%	53.50%	38.28%	28.62%	61.87%	53.50%	47.30%	37.29%
	Decision Tree Classifier	93.75%	93.00%	93.94%	93.47%	93.75%	93.00%	93.75%	93.00%
	Logistic Regression Classifier	95.62%	95.50%	95.65%	95.50%	95.62%	95.50%	95.63%	95.49%
	Random Forest Classifier	97%	97.50%	96.87%	97.53%	97%	97.50%	96.87%	95.49%
Stack Classifier	Stacking 1	96.25%	98.00%	96.25%	98.02%	96.25%	98.00%	96.25%	98.00%
	Stacking 2	98.12%	97.00%	98.18%	97.01%	98.12%	97.00%	98.12%	97.00%
	Stacking 3	98.12%	97.00%	98.18%	97.07%	98.12%	97.00%	98.12%	96.99%
	Stacking 4	97.50%	98.50%	97.52%	98.50%	97.50%	98.50%	97.49%	98.50%
Majority Voting	Combined Stacking Model	97.50%	98.00%	97.52%	98.01%	97.50%	98.00%	97.49%	97.99%

6. Experimental Results

6.4 Dataset 4:





Dataset 4 (70000)

Classification Algorithm	Accuracy		Precision		Recall		F1 Score	
	Validation	Test	Validation	Test	Validation	Test	Validation	Test
K-Nearest Neighbors	57.96%	57.91%	58.03%	57.91%	57.96%	57.91%	57.94%	57.89%
Support Vector Machine	49.59%	50.26%	75.00%	25.28%	49.58%	50.26%	32.88%	33.63%
Naïve Bayes Classifier	60.19%	59.80%	60.29%	59.98%	60.19%	59.80%	60.03%	59.65%
Decision Tree Classifier	63.84%	63.90%	63.85%	63.94%	63.84%	63.90%	63.82%	63.89%
Logistic Regression Classifier	72.13%	71.91%	73.33%	72.01%	72.13%	71.91%	72.09%	71.87%
Random Forest Classifier	72%	72.68%	72.37%	72.23%	72%	71.27%	72.25%	72.65%
Stacking 1	73.62%	73.66%	73.97%	73.89%	73.62%	73.66%	73.54%	73.59%
Stacking 2	73.40%	73.68%	73.85%	73.98%	73.40%	73.68%	73.30%	73.58%
Stacking 3	73.13%	73.29%	73.84%	73.77%	73.13%	73.29%	72.96%	73.13%
Stacking 4	73.63%	73.69%	74.19%	74.06%	73.63%	73.69%	73.51%	73.58%
Combined Stacking Model	73.64%	73.66%	74.09%	73.95%	73.64%	73.65%	73.54%	73.56%

7. Comparison with different Literature Survey

DATASET 1

Paper		Algorithm		Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Doppala et al. [8]		DT		82.56	83	79	81
		RF		90.75	88	93	90
		NB		84.24	82	85	84
		LR		84.03	81	87	84
		SVM		81.52	82	84	83
		Ensemble Model		93.39	99	88	90
Daza et al. [5]		RF		88.5	88.95	88.8	88.5
		Ensemble Model		91.5	91.5	91.6	91.49
Ozcan et al. [12]		Cart Algorithm		87.25	88.24	84.51	-
Tiwari et al [13]		RF		90.21	87.31	-	91.05
		KNN		80.85	78.67	-	82.62
		SVM		82.55	80.14	-	84.16
		Ensemble Model		92.34	92	-	92.74
OUR METHOD		KNN		85.71	85.75	85.71	85.68
		SVM		69.33	80.47	69.33	65.11
		NB		71.85	72.91	71.85	71.82
		DT		82.77	82.76	82.77	82.75
		LR		81.93	81.94	81.93	81.94
		RF		93.28	93.41	93.28	93.29
		Proposed Method		93.68	93.7	93.7	93.7

7. Comparison with different Literature Survey

DATASET 2							
Paper		**Algorithm**		**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Asyraf et al. [4]		LR		81.32	-	91.49	83.5
		KNN		74.73	-	89.36	78.5
		RF		82.42	-	85.11	83.3
		Ensemble Model		82.42	-	91.49	84.3
Nawaz et al. [11]		SVM		75	-	75	-
		KNN		66.73	-	68	-
		NB		83	-	78	-
		RF		92.48	-	91.17	-
		Gradient Decent Optimization		98.54	-	99.43	-
OUR METHOD		KNN		99.02	99.04	99.02	99.02
		SVM		98.04	98.12	98.05	98.05
		NB		70.24	70.59	70.24	70.13
		DT		86.83	87.26	86.83	86.79
		LR		85.37	86.26	85.36	85.28
		RF		98.54	98.54	98.54	98.54
		Proposed Method		99.02	99.04	99.02	99.02

7. Comparison with different Literature Survey

DATASET 3							
Paper		Algorithm		Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Doppala et al. [8]		DT		95	96	95	95
		RF		95.12	97	94	96
		NB		94.25	94	95	94
		LR		95.25	97	97	97
		SVM		93.15	93	95	93
		Ensemble Model		96.75	98	96	97
Mondal et al. [10]		RF		96	96	96	96
		DT		93	92	92	92
		LR		87	87	86	86
		SVM		78	77	78	77
		Ensemble Model		96.88	96	97	97
OUR METHOD		KNN		55	54.88	55	54.91
		SVM		53.5	28.62	53.5	37.29
		NB		53.5	28.62	53.5	37.29
		DT		93	93.47	93	93
		LR		95.5	95.5	95.5	95.49
		RF		97.5	97.53	97.5	95.49
		Proposed Method		98	98.01	98	97.99

7. Comparison with different Literature Survey

DATASET 4

Paper	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Delima et al [6]	LR	72.35	-	-	-
	DT	61.72	-	-	-
	RF	68.94	-	-	-
	SVM	72.16	-	-	-
	KNN	68.34	-	-	-
	Genetic Algorithm ANN	73.43	-	-	-
Alfaidi et al. [2]	LR	85.54	90.01	84.34	87.09
	RF	86.03	92.82	82.19	87.19
	DT	85.93	93.57	81.24	86.97
	NB	83.38	93.65	76.44	84.18
	KNN	84.56	88.97	83.66	86.24
	SVM	86.63	96.06	80.16	87.39
	MLP	87.23	95.23	82.01	88.13
Alqahtani et al. [3]	RF	88.65	90.03	88.03	88.02
	KNN	86.45	87.53	86.21	86.25
	DT	86.35	86.23	86.22	86.22
	Stacking	86.49	87.32	86.02	86.01
	ML Ensemble(DNN+KDNN+Stack)	88.7	88.02	88.02	88.01
OUR METHOD	KNN	57.91	57.91	57.91	57.89
	SVM	50.26	25.28	50.26	33.63
	NB	59.8	59.98	59.8	59.65
	DT	63.9	63.94	63.9	63.89
	LR	71.91	72.01	71.91	71.87
	RF	72.68	72.23	71.27	72.65
	Proposed Method	73.66	73.95	73.65	73.56

8. Summary

- Four different datasets are taken and six different classification algorithms are applied on each dataset.
- Significant improvements are shown using our approach on three of the datasets.
- For the dataset 4, we have not achieved satisfactory accuracy.
- Future work will focus on improving the model for this dataset 4 as well as other three datasets.

9. Reference

1. Ahmed Al Ahdal, Manik Rakhra, Rahul R. Rajendran, Farrukh Arslan, Moaiad Ahmad Khder,Binit Patel, Balaji Ramkumar Rajagopal, and Rituraj Jain. [retracted] monitoring cardiovascular problems in heart patients using machine learning. Journal of Healthcare Engineering,2023:9738123, Feb 2023. Paper link
2. Aseel Alfaidi, Reem Aljuhani, Bushra Alshehri, Hajer Alwadei, and Sahar Sabbeh. Machine learning: Assisted cardiovascular diseases diagnosis. International Journal of Advanced Computer Science and Applications, 13, 01 2022. Paper link
3. Abdullah Alqahtani, Shtwai Alsubai, Mohemmed Sha, Lucia Vilcekova, and Talha Javed. Cardiovascular disease detection using ensemble learning. Computational Intelligence and Neuroscience, 2022:5267498, Aug 2022 Paper link
4. Syafiq Asyraf, Nor Azuana Ramli, and Noryanti Muhammad. Heart disease prediction using ensemble of k-nearest neighbour, random forest and logistic regression method. page 040009, 03 2024. Paper link
5. Alfredo Daza, Juana Bobadilla, Juan Carlos Herrera, Angelica Medina, Nemias Saboya, Karoline Zavaleta, and Segundo Siguenas. Stacking ensemble based hyperparameters to diagnosing of heart disease: Future works. Results in Engineering, 21:101894, 2024. Paper link

9. Reference

6. Allemar Jhone Delima and Jan Carlo Arroyo. An optimized neural network using genetic algorithm for cardiovascular disease prediction. Journal of Advances in Information Technology, 13:95–99, 01 2022. Paper link
7. B. Rupa Devi, U. Sivaji, Thammisetty Swetha, J. Avanija, A. Suresh, and K. Reddy Madhavi. Advanced cardiovascular disease prediction: A comparative analysis of ensemble stacking and deep neural networks. International Journal of Intelligent Systems and Applications in Engineering, 12(6s):46–55, Nov. 2023. Paper link
8. Bhanu Prakash Doppala, Debnath Bhattacharyya, Midhunchakkaravarthy Janarthanan, and Namkyun Baik. A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques. Journal of Healthcare Engineering, 2022:2585235, Mar 2022. Paper link
9. G. Manikandan, B. Pragadeesh, V. Manojkumar, A.L. Karthikeyan, R. MANIKANDAN, and Amir Gandomi. Classification models combined with boruta feature selection for heart disease prediction. Informatics in Medicine Unlocked, 44, 01 2024. Paper link

9. Reference

10. Subhash Mondal, Ranjan Maity, Yachang Omo, Soumadip Ghosh, and Amitava Nag. An efficient computational risk prediction model of heart diseases based on dual-stage stacked machine learning approaches. IEEE Access, 12:72557270, 2024. Paper link

11. Muhammad Saqib Nawaz, Bilal Shoaib, and Muhammad Adeel Ashraf. Intelligent cardiovascular disease prediction empowered with gradient descent optimization. Heliyon, 7(5):e06948, 2021. Paper link

12. Mert Ozcan and Serhat Peker. A classification and regression tree algorithm for heart disease modeling and prediction. Healthcare Analytics, 3:100130, 2023. Paper link

13. Achyut Tiwari, Aryan Chugh, and Aman Sharma. Ensemble framework for cardiovascular disease prediction. Computers in Biology and Medicine, 146:105624, 2022. Paper link

10. Acknowledgment

We are deeply grateful to our esteemed supervisor, Dr. Soumitra Samanta, for his constructive academic advice, constant encouragement, and valuable suggestions at crucial junctures. His guidance and support have been instrumental in the successful completion of this project. We truly benefited from his excellent supervision.

We also extend our sincere thanks to our respected Head of the Department, Swami Punyeshwaranandaji maharaj, for providing us with the necessary facilities and support.

Additionally, we would like to thank our PhD scholar, Rajdeep Mandal and other seniors in our department for giving us this wonderful opportunity to work on this project.

Thank You