IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

IMPROVED LIGHTGBM MODEL PERFORMANCE ANALYSIS AND COMPARISON FOR CORONARY HEART DISEASE PREDICTION

P RAJYALAKSHMI[1], ALATHURU YAMINI[2], A DHANASEKHAR REDDY[3], G VISWANATH[4]
[1]Assistant Professor, Department of CSE, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: lakshmiprl83@gmail.com
[2]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: alathuryamini2000@gmail.com
[3]Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: dhanasekhar918@gmail.com
[4]Associate Professor, Department of CSE(AIML), Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: viswag111@gmail.com, ORCID: https://orcid.org/0009-0001-7822-4739

Abstract: Coronary heart disease (CHD) is a serious cardiovascular sickness with no therapy. Compelling patient treatment requires exact and early coronary course artery disease . Early distinguishing proof empowers early medicines and better persistent results. The "HY_OptGBM" model predicts CHD utilizing a superior LightGBM classifier. Gradient boosting framework LightGBM is proficient and exact in prescient demonstrating. Enhancements to the misfortune capability and hyperparameters streamline the LightGBM classifier. This streamlining approach further develops model preparation exactness and effectiveness. Model execution is surveyed utilizing Framingham Heart Institute coronary heart disease data. In light of this information, the model precisely predicts CHD, permitting early distinguishing proof and maybe lower treatment costs. Furthermore, presents a Voting Classifier (RF + AdaBoost) with close to 100% accuracy to analyze Coronary Heart Disease. This Random Forest-AdaBoost ensemble model recognizes CHD designs well. An easy to use Flask framework with SQLite combination works on user testing enlistment and signin to really take a look at convenience. The CHD discovery partners might utilize ML moves toward all the more effectively with this worked on interface.

Index terms *- Coronary heart disease, hyperparameter optimization, LightGBM, loss function, machine learning, OPTUNA.[48]*

1. INTRODUCTION

Atherosclerotic plaques in the coronary courses diminish blood stream to the heart muscle, causing CHD, a typical cardiovascular illness. Side effects incorporate chest uneasiness, angina, windedness, palpitations, and cardiovascular breakdown. CHD can cause a coronary episode, which can for all time harm the heart muscle and lower personal satisfaction. Perceiving and overseeing CHD by clinical mediation and way of life changes is pivotal [1].

Early CHD ID increments fix rates and brings down treatment costs. Because of advances in ML calculations and lower information capacity costs, a few ML calculations and data mining innovations have been generally applied in medication [2], [3], [4], [5], [6]. Sickness determination, valuable diagnostics, drug mining, and biomedicine require information mining

ISSN 2347–3657

Volume 12, Issue 3, 2024

advances. We can remove inert sickness data from huge measures of unstructured clinical information, build infection expectation models, and survey results utilizing information mining advancements.[50]

Medical care suppliers experience a few obstacles in giving superior grade, practical therapy. Medical clinics give quality medical services that expects doctors to have intensive information and make exact patient conclusions to forestall squandering cash. Data mining innovation is compelling and essential in clinical situations. The ideal hyperparameters [7], [8] for any arrangement strategy enormously influence execution. Picking the best hyperparameters further develops characterization calculation exactness. This study utilized OPTUNA [9] to upgrade LightGBM model hyperparameters. Consequently, our review chosen the best hyperparameters among the open ones. Random and network searches can upgrade hyperparameters. Another methodology is OPTUNA hyperparametric search. Since LightGBM execution relies upon how much hyperparameters, standard arbitrary and network search strategies don't gain from past enhancements, fooling around and being inefficient. OPTUNA gains from earlier enhancements and changes hyperparameters depending on the situation. OPTUNA was picked for hyperparameter enhancement in this work.

The misfortune capability influences model accuracy [10]. This work proposes a zeroed in misfortune capability in light of cross-entropy misfortune, including class weight α and test trouble weight changing variable γ. This study tended to lopsided positive and negative example extents. Likewise, the center misfortune capability can support model execution. This study utilized the engaged misfortune capability to further develop the LightGBM [11] model's default misfortune capability to anticipate CHD.

2. LITERATURE SURVEY

Overweight and corpulence are connected to standard and contemporary CVD risk factors, which increment the gamble of cardiovascular disease (CVD) and CHD [1]. Cardiovascular disease is likewise connected to weight. Focal weight and metabolic disorder are firmly connected to CVD, particularly CHD. There is solid epidemiologic proof connecting overweight and weight to CHD [2], [3], [4], [5], [6]. Posthumous and coronary corridor imaging examinations are less persuading. Ongoing examination suggest a stout conundrum in CHD mortality. Actual activity and cardiorespiratory wellness diminish stoutness' CVD gambles. There is little information on what deliberate weight reduction means for CVD results in overweight and hefty individuals.

Huge assets are being utilized to apply software engineering and measurements to clinical difficulties in Machine Learning (ML). Defenders of ML say it can deal with gigantic, convoluted, and different information found in medication and will advance biomedical exploration, altered therapy, and PC supported diagnostics [12,13]. ML thoughts are unfamiliar to numerous clinical professionals, and its examination potential is underutilized. In this exposition [2], we cover ML hypothesis, clinical ML calculations, their disadvantages, and the fate of ML in medication.

Artificial intelligence is most regularly utilized in drug treatment to match patients to the best medication or mix of medicines, expect drug-target or medication

ISSN 2347–3657

Volume 12, Issue 3, 2024

drug collaborations, and improve treatment regimens. Some ongoing man-made intelligence approaches for pharmacological treatment and organization are surveyed [3]. Joining patient information like hereditary qualities or proteomics with pharmacological information like compound substance attributes to assess treatment adequacy is normal in understanding medication determination. Closeness measures are utilized to foresee drug communications by assuming that prescriptions with comparable designs or targets would act in basically the same manner or associate. Numerical models assess pharmacokinetic and pharmacodynamic information to streamline drug portion. The newly developed areas of strength for and for each occupation are examined and evaluated here [12].

The dataset and preparing methods significantly influence ML model execution. The right preparation calculation can change a model's story. A few calculations succeed in some datasets however battle in others. Execution can likewise be expanded by altering calculation hyperparameters that manage preparing. This work [7] utilizes Grey Wolf Optimization (GWO) and Genetic Algorithm (GA) metaheuristics to upgrade ML calculation hyperparameters. Additionally, 11 calculations like Averaged Perceptron, FastTree, FastForest, LGBM, and Limited Memory. Broyden Fletcher Goldfarb Shanno algorithm Maximum Entropy (LbfgsMxEnt), Linear Support Vector Machine (LinearSVM), and a Deep Neural Network (DNN) with four designs are utilized on 11 natural, biomedical, and nature datasets about sub-atomic cooperations, malignant growth, clinical determination, conduct forecasts, RGB pictures of human skin, and X-rays of Covid19 and cardiomegaly patients. We found that all preliminaries upgraded preparing stage execution. Additionally, GWO performs better with 2.6E-5 p-esteem. Most analyses in this study show that metaheuristic approaches beat Thorough Matrix Search and meet speedier. The proposed method takes a dataset and offers the best-investigated calculation with related contentions. Hence, datasets with questionable appropriation, ML calculations with complex way of behaving, and shoppers new to insightful measurements and information science techniques can use it.

ML might be the best apparatus for high-throughput sequencing genomic information investigation because of its expectation limit. The perplexing technique of tweaking hyperparameters significantly upsets ML's utilization in creature and plant rearing undertakings. To improve genomic expectation utilizing ML, we consolidated tree-structured Parzen estimator (TPE), an independent tuning hyperparameters approach. TPE advanced KRR and SVR hyperparameters in this work [8]. To evaluate TPE execution, we analyzed KRR-TPE and SVR-TPE prediction accuracy to genomic best linear unbiased prediction (GBLUP) and KRR-RS, KRR-Lattice, SVR-RS, and SVR-Matrix, which tuned KRR and SVR hyperparameters utilizing random search (RS) and grid search (Gird) in reproduction and genuine datasets [47]. KRR-TPE anticipated all populaces well and was generally advantageous. For Chinese Simmental meat cows and Loblolly pine populaces, KRR-TPE exhibited a 8.73% and 6.08% normal increment above GBLUP in prediction accuracy. Our work will support GP ML and breeding improvement.[52]

3. METHODOLOGY

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

i) Proposed Work:

The recommended framework streamlines a LightGBM model for coronary heart disease prediction, assesses its presentation, executes gathering draws near, permits client information, and adds an easy to understand frontend and confirmation. For exact coronary heart disease prediction, enhancement and ensemble strategies increment accuracy. LightGBM's boundaries and misfortune capabilities are upgraded for prediction accuracy. The framework's adaptability and pertinence make it helpful in a few medical care areas [11,26]. What's more, presents a Voting Classifier (RF + AdaBoost) with close to 100% precision to analyze Coronary Heart Disease. This Random Forest-AdaBoost ensemble model recognizes CHD designs well. An easy to use Flask framework with SQLite reconciliation improves on client testing enlistment and signin to really take a look at convenience. The CHD location partners might utilize ML moves toward all the more effectively with this worked on interface [2], [3], [4], [5], [6].

ii) System Architecture:

Less complex arrangements are ideal for ML models, particularly for gigantic preparation and datasets. All of the previously mentioned make OPTUNA an extraordinary hyperparametric improvement structure. The superior LightGBM model plan is displayed in Fig. 1. Fig. 1 shows every specialist playing out the objective capability during the hunt.

Fig 1 Proposed architecture

iii) Dataset collection:

Framingham Heart Disease data is stacked and inspected to decide its design, qualities, and content. The Framingham Heart Study (FHS) looks to uncover cardiovascular disease risk factors. Framingham, Mama enlisted 5,209 people matured 30-62 of every 1948. A 1971 Posterity Companion, 1994 Omni Partner, 2002 Third Era Companion, 2004 New Posterity Mate Accomplice, and 2003 Second Era Omni Accomplice started. Research on cardiovascular and cerebrovascular ailments rules the dataset. Natural examples, sub-atomic hereditary information, phenotypic information, tests, pictures, member vascular utilitarian information, physiological information, segment information, and ECG information are included.A Boston College Public Heart, Lung, and Blood Establishment coordinated effort.

	Sex	Age	Education	CurrentSmoker	CigsPerDay	BPMeds	PrevalentStroke	PrevalentHyp
0	1	39	1	0	0.0	0.0	0	0
1	0	46	0	0	0.0	0.0	0	0
2	1	48	0	1	20.0	0.0	0	0
3	0	61	1	1	30.0	0.0	0	1
4	0	46	1	1	23.0	0.0	0	0

Fig 2 Framingham Heart Disease Data

iv) Data Processing:

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

Data processing transforms crude information into business-valuable data. Information researchers assemble, coordinate, clean, check, examine, and orchestrate information into charts or papers. Information can be handled physically, precisely, or electronically. Data ought to be more important and decision-production simpler. Organizations might improve tasks and pursue basic decisions quicker. PC programming advancement and other computerized data processing innovations add to this. Enormous information can be transformed into important experiences for quality administration and independent direction.[54]

v) Feature selection:

Feature selection chooses the steadiest, non-repetitive, and pertinent elements for model turn of events. As data sets extend in amount and assortment, purposefully bringing down their size is significant. The fundamental reason for feature selection is to increment prescient model execution and limit processing cost.

One of the vital pieces of feature engineering is picking the main attributes for machine learning algorithms. To diminish input factors, feature selection methodologies take out copy or superfluous elements and limit the assortment to those generally critical to the ML model. Rather than permitting the ML model pick the main qualities, feature selection ahead of time enjoys a few benefits.

vi) Algorithms:

AdaBoost is a strategy for ensemble discovering that forms strong classifiers by joining frail students, for the most part decision trees. By improving the exhibition of frail students, (for example, decision trees) in the ensemble, AdaBoost can expand the exactness of coronary heart disease prediction [25].

Fig 3 Adaboost

Decision Tree is a construction that looks like a flowchart, with each leaf hub addressing an outcome, the branch addressing a decision rule, and within hub addressing a feature. To work on the expectation of coronary heart disease, Decision Trees were utilized as base students in troupe procedures, for example, AdaBoost and Bagging [22].

Fig 4 Decision tree

To increment model accuracy, **bagging (otherwise called bootstrap aggregating)** involves building

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

various models utilizing different training dataset subsets and averaging the predictions. With regards to coronary heart disease prediction, bagging was utilized to create a ensemble of models, further developing prediction accuracy [26].

Fig 5 Bagging

Gradient Boosting limits a misfortune capability by more than once coordinating the predictions of feeble models to areas of strength for make models. An ensemble of models was created utilizing gradient boosting, which iteratively expanded the accuracy of coronary heart disease prediction [25].

Fig 6 Gradient boosting

XGBoost (Extreme Gradient Boosting) is a versatile and successful gradient boosting arrangement. To further develop coronary heart disease prediction accuracy, XGBoost was utilized as a boosting algorithm [25].

Fig 7 XGBoost

CatBoost is a proficient gradient boosting library for downright features. Classification information is taken care of consequently without pre-handling like one-hot encoding. CatBoost took care of absolute qualities in the dataset, working on demonstrating and further developing predictions [24].

Fig 8 Catboost

LightGBM is a gradient boosting, and Focal Loss is a changed loss function that objectives hard-to-order

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

tests to lighten class unevenness. LightGBM with Focal Loss zeroed in on predicaments to upgrade coronary heart disease prediction, particularly in lopsided information.

```
from scipy.misc import derivative

def focal_loss(ytrue, ypred, gamma=2.0):
    p = 1 / (1 + np.exp(-ypred))
    loss = -(1 - ytrue) * p**gamma * np.log(1 - p) - ytrue * (1 - p)**gamma * np
    return loss


def focal_loss_metric(ytrue, ypred):
    return 'focal_loss_metric', np.mean(focal_loss(ytrue, ypred)), False


def focal_loss_objective(ytrue, ypred):
    func = lambda z: focal_loss(ytrue, z)
    grad = derivative(func, ypred, n=1, dx=1e-6)
    hess = derivative(func, ypred, n=2, dx=1e-6)
    return grad, hess
```

Fig 9 Light GBM

This implies that LightGBM's ordinary misfortune capabilities ought to be utilized instead of the Focal Loss function. To think about and survey the impact of Focal Loss on the expectation execution for coronary heart disease, LightGBM without Focal Loss was used as the benchmark.

```
import lightgbm as lgb
clf = lgb.LGBMClassifier(boosting_type='gbdt',verbosity=-1,metric='auc',
clf.fit(X_train, y_train, verbose=0)

y_pred = clf.predict(X_test)

confusion = confusion_matrix(y_pred, y_test)
TP = confusion[1,1] # true positive
TN = confusion[0,0] # true negatives
FP = confusion[0,1] # false positives
FN = confusion[1,0] # false negatives

lgb_acc = accuracy_score(y_pred, y_test)
lgb_prec = precision_score(y_pred, y_test)
lgb_rec = recall_score(y_pred, y_test)
lgb_f1 = f1_score(y_pred, y_test)
lgb_auprc = average_precision_score(y_pred, y_test)
lgb_auroc = roc_auc_score(y_test, clf.predict_proba(X_test)[:, 1])
lgb_mcc = matthews_corrcoef(y_pred, y_test)

lgb_sens = TP / (TP + FN)
lgb_spec = TN / (TN + FP)

storeResults('LightGBM w/o Focal Loss',lgb_acc,lgb_prec,lgb_rec,lgb_f1,
```

Fig 10 LightGBM without Focal Loss

A Voting Classifier is an ensemble method that predicts the class name by greater part vote from many models. This review utilized a Voting Classifier with Random Forest (RF) and AdaBoost models to increment coronary heart disease prediction accuracy.

Fig 11 Voting classifier

4. EXPERIMENTAL RESULTS

Precision: Precision estimates the level of positive cases or tests precisely sorted. Precision is determined utilizing the recipe:

$$\text{Precision} = \text{True positives}/ (\text{True positives} + \text{False positives}) = TP/(TP + FP)$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

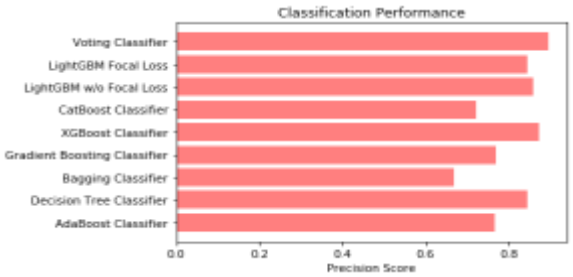

Fig 6 Precision comparison graph

Recall: Machine learning recall assesses a model's ability to perceive all significant examples of a class. It shows a model's culmination in catching occasions

ISSN 2347–3657

Volume 12, Issue 3, 2024

of a class by contrasting accurately anticipated positive perceptions with complete positives.

$$Recall = \frac{TP}{TP + FN}$$

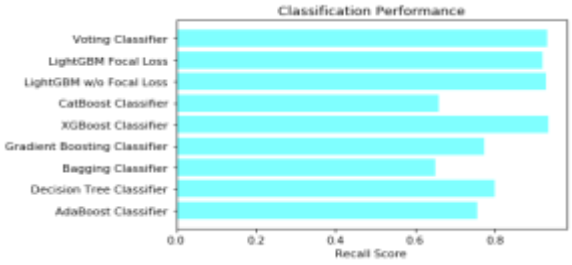

Fig 7 Recall comparison graph

Accuracy: The level of accurate expectations spread the word about in a grouping position is as accuracy, and it demonstrates how accurate a model's forecasts are by and large.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

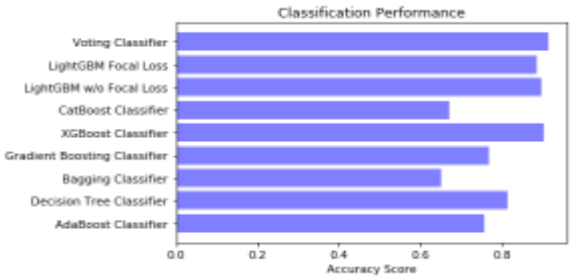

Fig 8 Accuracy graph

F1 Score: The F1 Score is fitting for unequal datasets on the grounds that it gives a reasonable metric that considers both bogus up-sides and misleading negatives. It is determined as the consonant mean of accuracy and recall.[56]s

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

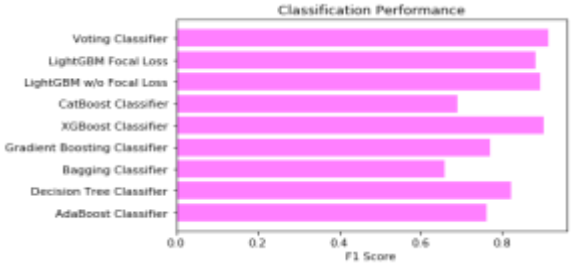

Fig 9 F1Score

ML Model	Accuracy	Precision	Recall	F1 Score
AdaBoost Classifier	0.758	0.767	0.757	0.762
Decision Tree Classifier	0.815	0.846	0.799	0.822
Bagging Classifier	0.651	0.668	0.651	0.659
Gradient Boosting Classifier	0.769	0.769	0.772	0.771
XGBoost Classifier	0.904	0.873	0.933	0.902
CatBoost Classifier	0.671	0.721	0.660	0.689
LightGBM w/o Focal Loss	0.896	0.860	0.929	0.893
LightGBM Focal Loss	0.885	0.845	0.921	0.881
Extension Voting Classifier	0.913	0.894	0.931	0.912

Fig 10 Performance Evaluation

Fig 11 Home page

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

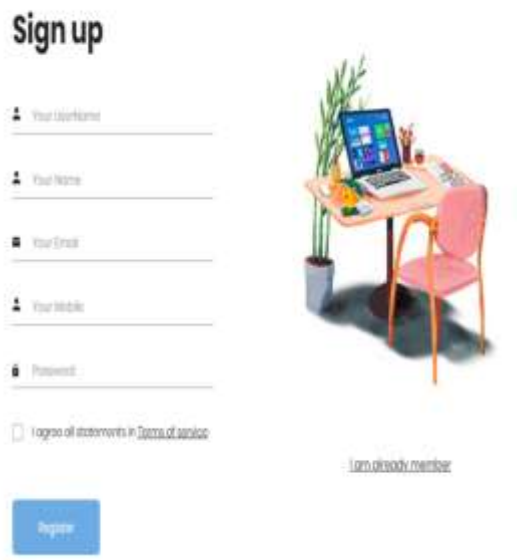

Fig 12 Signin page

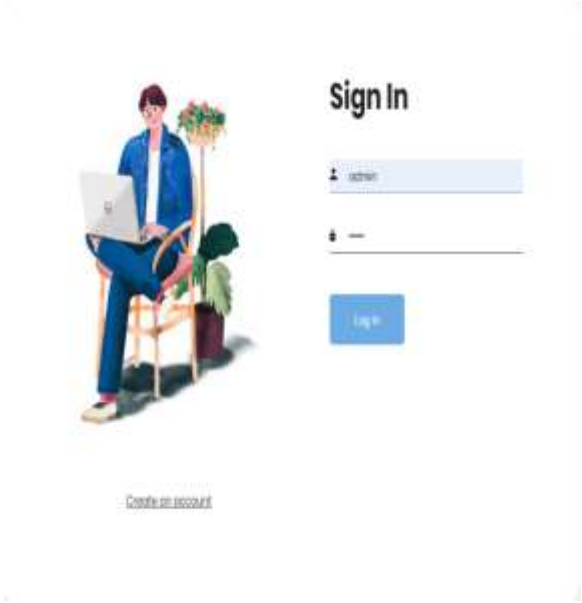

Fig 13 Login page

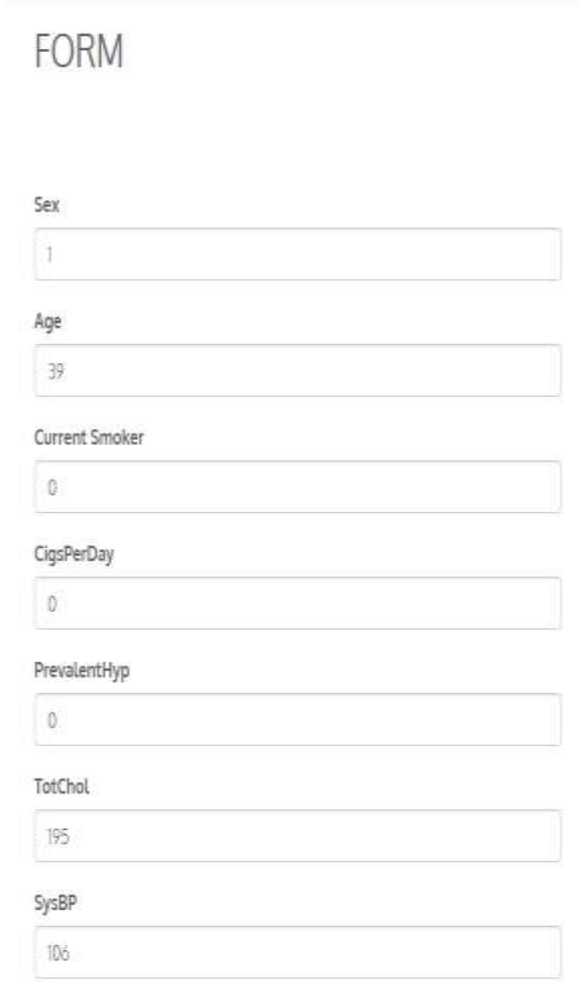

Fig 14 User input

Fig 15 Predict result for given input

5. CONCLUSION

Utilizing an upgraded LightGBM classifier and a reconsidered loss function, the HY_OptGBM expectation model precisely predicts CHD. Precision, recall, F1-score, and accuracy are utilized to assess the

ISSN 2347–3657

Volume 12, Issue 3, 2024

model's expectation abilities. Analyzers utilize strong classifiers and misfortune capabilities to further develop the HY_OptGBM model. These progressions increment the model's CHD detection and prediction accuracy [2], [3], [4], [5], [6]. An ensemble method joins expectations from various models to further develop system accuracy and flexibility. High level ensemble draws near, such the Voting Classifier, accomplish close to 100% accuracy, showing that changed models support prescient execution. A simple to-utilize Flask communicate with secure validation further develops system testing. This connection point works on information section for framework execution assessment, ensuring convenience and security.

6. FUTURE SCOPE

To further develop the HY_OptGBM model's coronary heart disease prediction, future exploration can add qualities or information sources. Clinical information might be incorporated for a total comprehension. Further review ought to test the model's generalizability and strength on greater and more differed datasets. This will uncover how actually the model adjusts to information conveyances. Contrasting the HY_OptGBM model with other modern ML techniques [12, 13] for CHD expectation can assist with deciding its viability and prevalence. You might utilize the proposed method to conjecture coronary illness as well as other cardiovascular diseases or circumstances. This expansion can change cardiology by giving an adaptable determining device.

REFERENCES

[1] N. Katta, T. Loethen, C. J. Lavie, and M. A. Alpert, ‘‘Obesity and coronary heart disease: Epidemiology, pathology, and coronary artery imaging,’’ Current Problems Cardiol., vol. 46, no. 3, Mar. 2021, Art. no. 100655, doi: 10.1016/j.cpcardiol.2020.100655.

[2] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, M. J. Lee, and H. Asadi, ‘‘EDoctor: Machine learning and the future of medicine,’’ J. Internal Med., vol. 284, no. 6, pp. 603–619, Sep. 2018, doi: 10.1111/joim.12822.

[3] E. L. Romm and I. F. Tsigelny, ‘‘Artificial intelligence in drug treatment,’’ Annu. Rev. Pharmacol. Toxicol., vol. 60, no. 1, pp. 353–369, Jan. 2020, doi: 10.1146/annurev-pharmtox-010919-023746.

[4] L. Lo Vercio, K. Amador, J. J. Bannister, S. Crites, A. Gutierrez, M. E. MacDonald, J. Moore, P. Mouches, D. Rajashekar, S. Schimert, N. Subbanna, A. Tuladhar, N. Wang, M. Wilms, A. Winder, and N. D. Forkert, ‘‘Supervised machine learning tools: A tutorial for clinicians,’’ J. Neural Eng., vol. 17, no. 6, Dec. 2020, Art. no. 062001, doi: 10.1088/1741-2552/abbff2.

[5] S. Rauschert, K. Raubenheimer, P. E. Melton, and R. C. Huang, ‘‘Machine learning and clinical epigenetics: A review of challenges for diagnosis and classification,’’ Clin. Epigenetics, vol. 12, no. 1, p. 51, Apr. 2020, doi: 10.1186/s13148-020-00842-4.

[6] Y. Arfat, G. Mittone, R. Esposito, B. Cantalupo, G. M. De Ferrari, and M. Aldinucci, ‘‘Machine learning for cardiology,’’ Minerva Cardiol. Angiol., vol. 70, no. 1, pp. 75–91, Mar. 2022, doi: 10.23736/s2724-5683.21.05709-4.

[7] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, and N. Aydin, ‘‘Tuning hyperparameters of machine

learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases,'' Comput. Biol. Chem., vol. 97, Apr. 2022, Art. no. 107619, doi: 10.1016/j.compbiolchem.2021.107619.

[8] M. Liang, B. An, K. Li, L. Du, T. Deng, S. Cao, Y. Du, L. Xu, X. Gao, L. Zhang, J. Li, and H. Gao, ''Improving genomic prediction with machine learning incorporating TPE for hyperparameters optimization,'' Biology, vol. 11, no. 11, p. 1647, Nov. 2022, doi: 10.3390/ biology11111647.

[9] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, ''OPTUNA: A nextgeneration hyperparameter optimization framework,'' in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Anchorage, AK, USA, 2019, pp. 2623–2631.

[10] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, ''Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation,'' Computerized Med. Imag. Graph., vol. 95, Jan. 2022, Art. no. 102026, doi: 10.1016/j.compmedimag.2021.102026.

[11] G. Ke et al., ''LightGBM: A highly efficient gradient boosting decision tree,'' in Proc. 31st Int. Conf. Neural Inf. Process. Syst., Long Beach, CA, USA, 2017, pp. 3149–3157.

[12] O. Goldman, O. Raphaeli, E. Goldman, and M. Leshno, ''Improvement in the prediction of coronary heart disease risk by using artificial neural networks,'' Qual. Manage. Health Care, vol. 30, no. 4, pp. 244–250, Jul. 2021, doi: 10.1097/qmh.0000000000000309.

[13] Z. Du, Y. Yang, J. Zheng, Q. Li, D. Lin, Y. Li, J. Fan, W. Cheng, X.-H. Chen, and Y. Cai, ''Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: Model development and performance evaluation,'' JMIR Med. Informat., vol. 8, no. 7, Jul. 2020, Art. no. e17257, doi: 10.2196/17257.

[14] J. K. Kim and S. Kang, ''Neural network-based coronary heart disease risk prediction using feature correlation analysis,'' J. Healthcare Eng., vol. 2017, Sep. 2017, Art. no. 2780501, doi: 10.1155/2017/2780501.

[15] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, ''Artificial intelligence in precision cardiovascular medicine,'' J. Amer. College Cardiol., vol. 69, no. 21, pp. 2657–2664, 2017, doi: 10.1016/j.jacc.2017.03.571.

[16] A. Akella and S. Akella, ''Machine learning algorithms for predicting coronary artery disease: Efforts toward an open source solution,'' Future Sci. OA, vol. 7, no. 6, Jul. 2021, Art. no. FSO698, doi: 10.2144/fsoa-2020- 0206.

[17] L. J. Muhammad, I. Al-Shourbaji, A. A. Haruna, I. A. Mohammed, A. Ahmad, and M. B. Jibrin, ''Machine learning predictive models for coronary artery disease,'' Social Netw. Comput. Sci., vol. 2, no. 5, p. 350, Sep. 2021, doi: 10.1007/s42979-021-00731-4.

[18] C. A. U. Hassan, J. Iqbal, R. Irfan, S. Hussain, A. D. Algarni, S. S. H. Bukhari, N. Alturki, and S. S. Ullah, ''Effectively predicting the presence of coronary heart disease using machine learning

ISSN 2347–3657

Volume 12, Issue 3, 2024

classifiers,'' Sensors, vol. 22, no. 19, p. 7227, Sep. 2022, doi: 10.3390/s22197227.

[19] Captainozlem. Framingham_CHD_Preprocessed_Data. Version 1. Accessed: May 5, 2020. [Online]. Available: https://www.kaggle. com/-datasets/captainozlem/framingham-chd-preprocesseddata/download?datasetVersionNumber=1

[20] V. Voillet, P. Besse, L. Liaubet, M. San Cristobal, and I. González, ''Handling missing rows in multi-omics data integration: Multiple imputation in multiple factor analysis framework,'' BMC Bioinf., vol. 17, no. 1, p. 402, Oct. 2016, doi: 10.1186/s12859-016-1273-5.

[21] G. Douzas and F. Bacao, ''Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE,'' Inf. Sci., vol. 501, pp. 118–135, Oct. 2019, doi: 10.1016/j.ins.2019.06.007.

[22] D. Che, Q. Liu, K. Rasheed, and X. Tao, ''Decision tree and ensemble learning algorithms with their applications in bioinformatics,'' in Software Tools and Algorithms for Biological Systems (Advances in Experimental Medicine and Biology), H. Arabnia and Q. N. Tran, Eds. New York, NY, USA: Springer, 2011, pp. 191–199.

[23] L. Yang, H. Wu, X. Jin, P. Zheng, S. Hu, X. Xu, W. Yu, and J. Yan, ''Study of cardiovascular disease prediction model based on random forest in eastern China,'' Sci. Rep., vol. 10, no. 1, p. 5245, Mar. 2020, doi: 10.1038/s41598-020-62133-5.

[24] J. T. Hancock and T. M. Khoshgoftaar, ''CatBoost for big data: An interdisciplinary review,'' J. Big Data, vol. 7, no. 1, p. 94, Nov. 2020, doi: 10.1186/s40537-020-00369-8.

[25] W. Wenbo, S. Yang, and C. Guici, ''Blood glucose concentration prediction based on VMD-KELM-adaboost,'' Med. Biol. Eng. Comput., vol. 59, nos. 11–12, pp. 2219–2235, Sep. 2021, doi: 10.1007/s11517-021-02430- x.

[26] X. Mi, F. Zou, and R. Zhu, ''Bagging and deep learning in optimal individualized treatment rules,'' Biometrics, vol. 75, no. 2, pp. 674–684, Mar. 2019, doi: 10.1111/biom.12990.

[27] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, ''Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM),'' Diagnostics, vol. 11, no. 9, p. 1714, Sep. 2021, doi: 10.3390/diagnostics11091714.

[28] J. Feng, B. Ni, D. Xu, and S. Yan, ''Histogram contextualization,'' IEEE Trans. Image Process., vol. 21, no. 2, pp. 778–788, Feb. 2012, doi: 10.1109/TIP.2011.2163521.

[29] P. Łabędź, K. Skabek, P. Ozimek, and M. Nytko, ''Histogram adjustment of images for improving photogrammetric reconstruction,'' Sensors, vol. 21, no. 14, p. 4654, Jul. 2021, doi: 10.3390/s21144654.

[30] L. Lin, J. Zhang, N. Zhang, J. Shi, and C. Chen, ''Optimized LightGBM power fingerprint identification based on entropy features,'' Entropy, vol. 24, no. 11, p. 1558, Oct. 2022, doi: 10.3390/e24111558.

[31] O. Krivorotko, M. Sosnovskaia, I. Vashchenko, C. Kerr, and D. Lesnic, ‘‘Agent-based modeling of COVID-19 outbreaks for New York state and U.K.: Parameter identification algorithm,’’ Infectious Disease Model., vol. 7, no. 1, pp. 30–44, Mar. 2022, doi: 10.1016/j.idm.2021.11.004.

[32] A. Namoun, B. R. Hussein, A. Tufail, A. Alrehaili, T. A. Syed, and O. BenRhouma, ‘‘An ensemble learning based classification approach for the prediction of household solid waste generation,’’ Sensors, vol. 22, no. 9, p. 3506, May 2022, doi: 10.3390/s22093506.

[33] M. M. Arifin, M. A. Based, K. M. Mumenin, A. Imran, M. A. Azim, Z. Alom, and M. A. Awal, ‘‘OLGBM: Optuna optimized light gradient boosting machine for intrusion detection,’’ in Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (IC4ME2), Rajshahi, Bangladesh, Dec. 2021, pp. 1–4.

[34] P. Srinivas and R. Katarya, ‘‘HyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost,’’ Biomed. Signal Process. Control, vol. 73, Mar. 2022, Art. no. 103456, doi: 10.1016/j.bspc.2021.103456.

[35] D. Jensen and J. Neville, ‘‘Correlation and sampling in relational data mining,’’ in Proc. 33rd Symp. Interface Comput. Sci. Statist., 2001, pp. 1–14.

[36] S. Yan, J. M. Peck, M. Ilgu, M. Nilsen-Hamilton, and M. H. Lamm, ‘‘Sampling performance of multiple independent molecular dynamics simulations of an RNA aptamer,’’ ACS Omega, vol. 5, no. 32, pp. 20187–20201, Aug. 2020, doi: 10.1021/acsomega.0c01867.

[37] M. Komorowski, D. C. Marshall, J. D. Salciccioli, and Y. Crutain, ‘‘Exploratory data analysis,’’ in Secondary Analysis of Electronic Health Records. Cham: Springer, 2016, pp. 185–203.

[38] T. R. Vetter, ‘‘Descriptive statistics: Reporting the answers to the 5 basic questions of who, what, why, when, where, and a sixth, so what?’’ Anesthesia Analgesia, vol. 125, no. 5, pp. 1797–1802, Nov. 2017, doi: 10.1213/ane.0000000000002471

[39] B. Wang, J. J. Klemeš, P. S. Varbanov, and M. Zeng, ‘‘An extended grid diagram for heat exchanger network retrofit considering heat exchanger types,’’ Energies, vol. 13, no. 10, p. 2656, May 2020, doi: 10.3390/en13102656.

[40] M. W. Browne, ‘‘Cross-validation methods,’’ J. Math. Psychol., vol. 44, no. 1, pp. 108–132, 2000, doi: 10.1006/jmps.1999.1279.

[41] S. Parvandeh, H.-W. Yeh, M. P. Paulus, and B. A. McKinney, ‘‘Consensus features nested cross-validation,’’ Bioinformatics, vol. 36, no. 10, pp. 3093–3098, May 2020, doi: 10.1093/bioinformatics/btaa046.

[42] S. Kucheryavskiy, S. Zhilin, O. Rodionova, and A. Pomerantsev, ‘‘Procrustes cross-validation—A bridge between cross-validation and independent validation sets,’’ Anal. Chem., vol. 92, no. 17, pp. 11842–11850, Aug. 2020, doi: 10.1021/acs.analchem.0c02175.

[43] J.-J. Beunza, E. Puertas, E. García-Ovejero, G. Villalba, E. Condes, G. Koleva, C. Hurtado, and M. F. Landecho, ‘‘Comparison of machine learning algorithms for clinical event prediction (risk of

coronary heart disease),'' J. Biomed. Informat., vol. 97, Sep. 2019, Art. no. 103257, doi: 10.1016/j.jbi.2019.103257.

[44] M. V. Dogan, I. M. Grumbach, J. J. Michaelson, and R. A. Philibert, ''Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham heart study,'' PLoS ONE, vol. 13, no. 1, Jan. 2018, Art. no. e0190549, doi: 10.1371/journal.pone.0190549.

[45] M. V. Dogan, S. Knight, T. K. Dogan, K. U. Knowlton, and R. Philibert, ''External validation of integrated genetic-epigenetic biomarkers for predicting incident coronary heart disease,'' Epigenomics, vol. 13, no. 14, pp. 1095–1112, Jul. 2021, doi: 10.2217/epi-2021-0123.

[46] S. Simon, D. Mandair, A. Albakri, A. Fohner, N. Simon, L. Lange, M. Biggs, K. Mukamal, B. Psaty, and M. Rosenberg, ''The impact of time horizon on classification accuracy: Application of machine learning to prediction of incident coronary heart disease,'' JMIR Cardio, vol. 6, no. 2, Nov. 2022, Art. no. e38040, doi: 10.2196/38040.

[47] S. Prabu, B. Thiyaneswaran, M. Sujatha, C. Nalini, and S. Rajkumar, ''Grid search for predicting coronary heart disease by tuning hyper-parameters,'' Comput. Syst. Sci. Eng., vol. 43, no. 2, pp. 737–749, 2022.

[48] G.Viswanath, "Hybrid encryption framework for securing big data storage in multi-cloud environment", Evolutionary intelligence, vol.14, 2021, pp.691-698.

[49] Viswanath Gudditi, "Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage", Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol.12, 2021, pp.545-552.

[50] Viswanath Gudditi, "A Smart Recommendation System for Medicine using Intelligent NLP Techniques", 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022, pp.1081-1084.

[51] G.Viswanath, "Enhancing power unbiased cooperative media access control protocol in manets", International Journal of Engineering Inventions, 2014, vol.4, pp.8-12.

[52] Viswanath G, "A Hybrid Particle Swarm Optimization and C4.5 for Network Intrusion Detection and Prevention System", 2024, International Journal of Computing, DOI: https://doi.org/10.47839/ijc.23.1.3442, vol.23, 2024, pp.109-115.

[53] G.Viswanath, "A Real Time online Food Ording application based DJANGO Restfull Framework", Juni Khyat, vol.13, 2023, pp.154-162.

[54] Gudditi Viswanath, "Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS", 2014, International Journal of Engineering Inventions, vol.4, pp.08-12.

[55] G.Viswanath," A Real-Time Video Based Vehicle Classification, Detection And Counting System", 2023, Industrial Engineering Journal, vol.52, pp.474-480.

[56] G.Viswanath, "A Real- Time Case Scenario Based On Url Phishing Detection Through Login Urls

ISSN 2347–3657

Volume 12, Issue 3, 2024

", 2023, Material Science Technology, vol.22, pp.103-108.

[57] Manmohan Singh,Susheel Kumar Tiwari, G. Swapna, Kirti Verma, Vikas Prasad, Vinod Patidar, Dharmendra Sharma and Hemant Mewada, "A Drug-Target Interaction Prediction Based on Supervised Probabilistic Classification" published in Journal of Computer Science, Available at: https://pdfs.semanticscholar.org/69ac/f07f2e756b79181e4f1e75f9e0f275a56b8e.pdf