Grenze International Journal of Engineering and Technology, June Issue

Enhancing Healthcare with Early Heart Disease Prediction

Roshni Bhave[1], Shilpa Dhopte[2], Gayatri Raut[3], Shweta P Sondawale[4] and Nikita Katariya[5]
[1-2]Yeshwantrao Chavan College of Engineering, Nagpur, India
Email: roshnibhave12@gmail.com, shilpa.dhopte@gmail.com
[3]R.H.Sapat College of Engineering, Management studies and Research, Nashik, India
Email: gayatri.wahane@gmail.com
[4]Sinhgad Academy of Engineering Pune
Email: shweta.sondawale123@gmail.com
[5]Shri Ramdeobaba College of Engineering and Management, Department of Computer Science & Engineering, Nagpur,India
Email: katariyanikita08@gmail.com

***Abstract*— heart disease, a widespread ailment that affects communities all over the world, is one of the most important global health concerns. The goal of this essay is to construct a machine learning model that can detect cardiac disease in its early stages because early detection is crucial for saving lives. The importance of this project rests in its potential to help the general population by enabling early intervention before heart issues worsen, potentially lowering the need for pricy medical procedures and treatments in hospitals and clinics. For a very long time, cardiac disease prediction and detection have been extremely difficult for healthcare professionals. These problems frequently result from lifestyle choices like bad eating habits, binge drinking, smoking, high blood pressure, raised cholesterol, diabetes, obesity, stress, and a sedentary lifestyle. As a result, creating a reliable prediction model has enormous potential for solving this important healthcare problem. Different supervised machine learning techniques have been used to predict cardiac disease, with an excellent accuracy rate of 94%.the performance of the proposed technique is evaluated on a real dataset known as mortality-rate-heart-patient-hospital.**

***Index Terms*— Heart disease, Random Forest, Artificial Neural Network, Support Vector**

I. INTRODUCTION

The second most important organ in the human body, after the brain, is the heart. Blood circulation throughout the body is its main purpose. Different disorders that can impair the heart's ability to efficiently pump blood are collectively referred to as heart disease. The two most common cardiac conditions are heart failure (HF) and coronary artery disease (CAD), despite the fact that they can take many different forms. The main cause of coronary heart disease (CAD) is the blockage or narrowing of the coronary arteries, which provide blood to the heart muscle [1]. The blood supply to the heart is crucially maintained by these arteries. Numerous variables that can affect the risk of heart failure are divided into two categories by medical experts: those that can be changed and those that cannot. Family history, gender, and age are uncontrollable risk factors, whereas diabetes, high cholesterol, hypertension, smoking, obesity, and inactivity are modifiable risk factors [3]. Early diagnosis or prognosis is essential given the severity of cardiac disease. Heart disease can be diagnosed using a number of

Grenze ID: 01.GIJET.10.2.168

procedures, with angiography being a technique that is often utilized worldwide. But there are drawbacks to the angiography method. An early diagnosis of cardiac abnormalities is essential for prompt intervention, which can stop further consequences like heart attacks and slow the course of the condition [2]. By using patient records to identify early indicators of the disease, machine learning algorithms can develop preventive measures and improve patient care. There are many techniques for detecting diseases early that are being researched. The Random Forest (RF) approach, Artificial Neural Networks (ANN), and Support Vector Machine (SVM), which frequently produce higher accuracy rates than Multi-linear Regression, are just a few of the tools and methods used to achieve these needs.
These risk factors, which are shown in Table I and II below, are thought to be manageable:

TABLE I: RISK FACTORS NOT UNDER CONTROL

Attributes	Description
Gender	Biological variations
Age	Age-related alterations in the heart and blood vessels
Family History	Genetic influences

TABLE II: FREQUENTLY OCCURRING RISK FACTORS FOR HEART DISEASE

Attributes	Description
Diabetes	High blood glucose levels cause damage to blood vessels, which causes heart problems.
High Cholesterol levels	It is challenging to get enough blood to circulate through the arteries because blood vessel fat deposits have increased.
Hypertension	Heart attacks are caused by the force of blood flow against the arterial walls.
Obesity	Alters the heart's capacity to beat.
Smoking	The heart's functionality is altered by the chemicals found in cigarettes.
Physical Inactivity	Probability of developing high cholesterol.

These methods and technologies are essential in the effort to identify cardiac disease early, which paves the way for better patient treatment and more efficient preventative measures.

II. LITERATURE REVIEW

Numerous researchers have worked on predicting heart illness using different machine learning algorithms by taking different heart-related variables. According to L.Ali et al [1], several machine learning-based models were created to help enhance the way cardiologists make diagnoses and to predict heart failure (HF) before it occurs. Give an example of an expert system that uses the stacking of two support vector machine (SVM) models to accurately predict HF. The initial SVM model is linear and L1 regularized. It can eliminate features that provide no purpose by setting their coefficients to zero. Using L2 regularization, the second SVM model is constructed. It is used as a prediction model. We offer a hybrid grid search algorithm (HGSA) that is capable of simultaneously optimizing the two models. The experimental findings demonstrate that the technique improves the performance of a conventional SVM model by 3.3%. In addition, the suggested strategy out performs the ten previously proposed methods, which had accuracy ranges between 57.85% and 91.83%.
The aim of a study by Drod et al. (2022) [2] was to identify the most important risk factors for cardiovascular disease (CVD) in patients with metabolic-associated fatty liver disease (MAFLD) using machine learning (ML) techniques. 191 MAFLD patients had their blood biochemically analyzed, and subclinical atherosclerosis was evaluated. ML techniques, such as multiple logistic regression classifier, univariate feature ranking, and principal component analysis (PCA), were used to construct a model to determine people who had the highest risk of CVD. The most important clinical traits, according to the study, were hypercholesterolemia, plaque scores, and length of diabetes. With an AUC of 0.87, the ML method worked well, correctly classifying 114/144 (79.17%) low-risk patients and 40/47 (85.11%) high-risk patients
Decision Trees were used to implement Purushottam et al.'s[3] effective approach for predicting heart disease, which provided 87.3% accuracy during training and 86.3% accuracy during testing.
Lerina Aversano[4] utilizing machine learning techniques, this study aims to predict cardiac disease by examining a wide range of clinical data made available by electronic patient records. The study investigates the use of several

machine learning and deep learning methods to forecast the occurrence of heart disease using feature model-generated data. The study also explores hyper-parameter optimization to improve the performance of the chosen machine learning classifiers. The evaluation is carried out using actual data, and the results show good predictive performance. In conclusion, this study uses extensive clinical data from electronic patient records to predict cardiac disease using machine learning and deep learning techniques. The techniques such as KNN, NB, and CB are the less performing for predicting heart disease. The accuracy of KNN is 70%, NB accuracy is 84 and CB accuracy is 87%.

Dr. Chintan Bhatt [5] In order to lower the fatality rate brought on by cardiovascular disorders, our research creates a model that can accurately forecast these conditions. The approach of k-modes clustering with Huang beginning that is suggested in this paper can increase classification precision. Models like the decision tree classifier (DT), multilayer perceptron (MP), random forest (RF), and XGBoost (XGB) are employed. On a real-world dataset of 70,000 cases from Kaggle, the suggested model is tested. This underpinning research has shown that multilayer perceptron with cross-validation has surpassed all other algorithms in terms of accuracy, which is the conclusion that can be derived from it. The best accuracy was attained, 87.28%.

R. Deepika, [6] This approach is made to continually monitor patients and gather diagnostic information from them at regular intervals in order to enable prompt identification of potential problems. This study's main goal is to improve the predictive system's performance by making sure it can accurately recognize and record all the crucial characteristics linked to heart problems. The main objective of the study is to enhance the accuracy and comprehensiveness of the system's recognition and analysis of the many indications and patterns connected to heart-related issues for more accurate and early detection.

III. Comparative Study Of Literature Survey

TABLE III: SHOWS COMPARATIVE STUDY OF SURVEY

	Techniques used	Accuracy
L.Ali et al	Decision tree, hybrid grid search algorithm (HGSA)	accuracy ranges between 57.85% and 91.83%.
Drod et al.	MAFLD	85.11%
Purushottam et al.'s	Decision Tree	87.3% accuracy during training and 86.3% accuracy during testing.
Lerina Aversano	KNN, NB, CB	KNN-70%, NB-84% and CB-87%.
Dr. Chintan Bhatt	Decision tree classifier(DT), multilayer perceptron (MP), random forest (RF), and XGBoost (XGB)	87%

Performance Variability: The considerable performance variability of methods such as Decision Tree with HGSA emphasizes the significance of parameter optimization techniques.

Robustness: MAFLD performs consistently, indicating that it might be more resilient to changes in tasks or datasets.

Overfitting: Regularization techniques could be used to address the possible overfitting of the Decision Tree, as indicated by the minor decrease in accuracy from training to testing.

Algorithm Selection: Different algorithms such as Decision Trees, Random Forest, and XGBoost may provide equivalent performance depending on the particular job and dataset features, making the selection of an approach based on computational complexity or interpretability easier.

IV. Methodology

A classification model is created using RF, a machine learning technique, by building a number of decision trees on training data sets. When dealing with large data sets, this approach provides excellent accuracy by choosing a tree based on the majority of selections. In order to create a more potent ensemble model, this approach combines the two feature selection techniques of bagging and random selection. The risk of overfitting and the amount of training time are reduced by using multiple trees with the RF technique. Additionally, it provides estimates for important classification factors and missing data, all of which boost accuracy.

1. Data collection

Before model training to predict early heart disease, we need a dataset with pertinent patient data, including things like age, gender, blood pressure, cholesterol levels, family history of heart disease, smoking habits, and other clinical factors. The dataset should also state whether each patient has or has not developed cardiac disease.

2. Data Preprocessing

Data preprocessing is essential to ensure that the input data is clean and suitable for training the neural network. This includes handling missing data, normalizing or standardizing features, and encoding categorical variables. Encoding categorical variables, Normalization or standardizing features, and handling missing data.

3. Model Architecture

Creating the architecture of the neural network is a crucial stage. An input layer with neurons corresponding to the features, one or more hidden layers, and an output layer with a single neuron representing the binary classification (heart disease or no heart disease) might make up a common architecture for heart disease prediction.

4 Model Training

Model training is crucial part of machine learning. In this stage model tries to learn the pattern from dataset and on that basis new data point or unknown data point predicted. Before model training we need to split dataset into training and testing part. The need of splitting in the machine learning is an important task.

5 Testing & Validation

To make sure the model generalizes effectively to new data, the model's performance is assessed using a separate validation dataset. The model's accuracy, precision, recall, and F1-score for heart disease prediction are further tested on a held-out test dataset.

B. Support Vector Machine

Both linear and non-linear data can be classified using support vector machines. In order to categorize the data, it first performs a nonlinear mapping to move the initial training data into a higher dimension, and then it looks for a linear best separation hyperplane. It is situated on the mapping where a high hyperplane serves as the boundary between two data points.

For this we used secondary data from the Kaggle repository platform at https://www.kaggle.com/morality-rate-patient-hospital". We got features from this open-source platform. All attributes are fully described in Table 4, along with the range or total number of values for each attribute.

TABLE IV: HEART DISEASES MORALITY RATE PATIENT DATASET ATTRIBUTES DETAILED INFORMATION

	Age	Age.Group	Gender	Locality	Marital status	Life.Style	Sleep	Category	Depression	Hyperlipi	...	oldpeak	slope	ca	thal	num	SK	SK.React	Reaction	Mortality	Follow.Up
0	45	41-50	Female	RURAL	MARRIED	NO	NO	FREE	YES	YES	...	3.0	2	0	7	2	1	NO	0	0	60
1	51	51-60	Female	URBAN	MARRIED	NO	NO	FREE	YES	YES	...	1.2	2	0	7	2	1	NO	0	0	15
2	55	51-60	Female	RURAL	MARRIED	YES	YES	FREE	YES	YES	...	3.4	2	0	3	2	1	NO	0	0	6
3	55	51-60	Female	RURAL	MARRIED	YES	YES	FREE	YES	YES	...	2.0	2	1	7	3	1	NO	0	0	52
4	56	51-60	Female	RURAL	MARRIED	YES	NO	FREE	YES	YES	...	4.0	3	2	7	3	1	NO	0	0	34

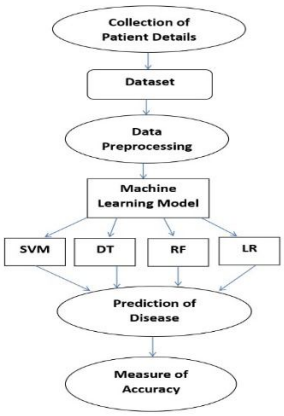

Figure 1. 4-layer Early Heart Disease Prediction

V. DISCUSSION OF RESULTS

The effectiveness of the Random Forest classifier and other classifiers like ANN, SVM was evaluated on the data. This evaluation approach has been applied to the Kaggle dataset using the Random For RF, ANN, SVM and shows the confusion matrix for all models after data cleaning and feature selection strategies have been applied. When compared to alternative methodologies and current classifiers, our proposed model achieves the best results in all performance metrics for the dataset studied. We achieved 94.44% accuracy in our proposed approach when compared to other current algorithms. We applied various models to the dataset before and after using these statistical metrics. Heart disease is the biggest cause of death worldwide and is easier to treat when discovered

early. The phrase "heart disease" covers a wide range of heart disorders. Objective is to examine databases to learn We have what number and gender of patients in these samples? how many people have high cholesterol and diabetes; how many people live a healthy lifestyle whether they smoke or not. The table 5 below displays the comparison results obtained from applying them to the dataset for predicting heart disease.

TABLE V: COMPARISON OF RESULT BY VARIOUS MODEL

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.77&	0.77	0.77	0.77
Logistic Regression	86.11%	0.86	0.86	0.86
Decision Tree	83.33%	0.83	0.83	0.83
Random Forest Classifier	94.44%	0.95	0.94	0.94

These days, it can be difficult to identify heart defects early. As a preventive precaution, we are attempting to identify heart irregularities by analyzing the dataset using Random Forest, ANN, and SVM approaches because silent attacks can have few or no symptoms. The outcome displays the highest accuracy rating i.e. 94.44%.

VI. CONCLUSION

These days, it can be difficult to identify heart defects early. As a preventive precaution, we are attempting to identify heart irregularities by analyzing the dataset using Random Forest, ANN, and SVM approaches because silent attacks can have few or no symptoms. The outcome displays the highest accuracy rating i.e. 94.44%.

REFERENCES

[1] System L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert for the Effective Prediction of Heart Failure," IEEE Access, vol. 7, pp. 54007–54014, 2019, doi:10.1109/ACCESS.2019.2909969.
[2] Drożdż, K.; Nabrdalik, K.; Kwiendacz, H.; Hendel, M.; Olejarz, A.; Tomasik, A.; Bartman, W.; Nalepa, J.; Gumprecht, J.; Lip, G.Y.H. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol. 2022, 21, 240.
[3] Kanak Saxena, Richa Sharma, et al. Efficient heart disease prediction system. Procedia Computer Science, 85:962–969, 2016.
[4] Lerina Aversano "Using Machine Learning for early prediction of heart disease" 2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS) |978-1-6654-3706-6/22 ©2022 IEEE 10.1109/EAIS51927.2022.
[5] R. Deepika "Early Detection of Heart Disease Using Deep Learning Model" 978-1-6654-9761-9/22/31.00 2022 IEEE.
[6] Dr. Chintan Bhatt "Effective Heart Disease Prediction Using Machine Learning Techniques" Algorithms, EISSN 1999-4893, Published by MDPI, 1 February 2023.
[7] World Health Organization. Global status report on noncommunicable diseases 2014. 2014.WHO. Western Pasific regional action plan of concummunicable diseases. World Heal. Organ.; 2014.
[8] Wang Q, Fu AZ, Brenner S, Kalmus O, Banda HT, De Allegri M. Out-of-pocket expenditure on chronic non-communicable diseases in Sub-Saharan Africa: the case of rural Malawi. PLoS One 2015;10(1):1–15. https://doi.org/10.1371/journal. pone.0116897.
[9] Katarya R, Srinivas P. Predicting heart disease at early stages using machine learning: a survey. Proc Int Conf Electron Sustain Commun Syst ICESC 2020;2020 (Icesc):302–5. https://doi.org/10.1109/ICESC48915.2020.9155586.
[10] Rani P, Kumar R, Ahmed NMOS, Jain A. A decision support system for heart disease prediction based upon machine learning. J Reliab Intell Environ 2021;7(3): 263–75. https://doi.org/10.1007/s40860-021-00133-6.
[11] Nancy P, Swaminathan B, Navina K, Nandhine B, Lokesh P. Tuned Random Forest Algorithm for Improved Prediction of Cardiovascular Disease, 1; 2020. p. 1355–60. https://doi.org/10.35940/ijrte.A1599.059120.
[12] Almustafa KM. Prediction of heart disease and classifiers' sensitivity analysis. BMC Bioinf 2020;21(1):1–18. https://doi.org/10.1186/s12859-020-03626-y.
[13] Vinutha HP, Poornima B, Sagar BM. Detection of outliers using interquartile range technique from intrusion dataset, vol. 701. Springer Singapore; 2018. https://doi. org/10.1007/978-981-10-7563-6_53. heart failure diagnosis by using machine learning algorithms. Comput Math Methods Med 2021;2021.
[14] Thanh Noi P, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery.Sensors (Basel). 2017;18(no.https://doi.org/10.3390/s1801001
[15] Jabbar M, B. D.-J. of network and, and undefined 2016. Intelligent heart disease prediction system using random forest and evolutionary approach. Researchgate. Net 2016;4(2016):175–84.
[16] Borz I, Palade T, Puschita E, Dolea P, Pastrav A. Wireless sensor networks for healthcare monitoring. IFMBE Proc 2022;88(11):232–9. https://doi.org/10.1007/ 978-3-030-93564-1_26.
[17] Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl. 2017;9:1 16. doi.org/10.4236 / jilsa.2017.91001.

[18] Pahwa K, Kumar R. Prediction of heart disease using hybrid technique for selecting features. In: 2017 4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON). IEEE. p. 500 504. 8. Pouriyeh S, Vahid S, Sannino G, De Pietro G,
[19] Arabnia H, Gutierrez J. A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease 2017 IEEE symposium on computers and communications (ISCC). IEEE. p. 204 207.
[20] Swapna G, SomanKp, VinayakumarR,Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals,Procedia Computer Science,Volume 132,2018,Pages 1253-1262,ISSN 1877-0509, https://doi.org/10.1016/j.imu.2021
[21] Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, Wang L. Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. JMIR MhealthUhealth. 2019 Aug 2;7(8):e11966. doi: 10.2196/11966. PMID: 31376272; PMCID: PMC6696854.
[22] Fahd Saleh Alotaibi, "Implementation of Machine Learning Model to predict Heart Failure Disease," International Journal of Advanced Computer Science and Applications (IJACSA), vol.10, no.6, pp.261-268, 2019.
[23] Sanjay Kumar Sen, "Predicting and Diagnosing of Heart Disease using Machine Learning Algorithms," International Journal of Engineering and Computer Science (IJECS), vol.6, pp.21623-21631,2017.
[24] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020,PP. 302-305.doi:10.1109/ICESC48915.2020.9155586.
[25] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-4. doi: 10.1109/CCAA.2018.8777449.
[26] P. Sujatha and K. Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-7. doi: 10.1109/INOCON50539.2020.9298354
[27] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour,"An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," IEEE Access, vol. 7, pp. 180235–180243, 2019, doi: 10.1109/ACCESS.2019.2952107.
[28] M. Gjoreski, A. Gradisek, B. Budna, M. Gams, and G. Poglajen,"Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure from Heart Sounds," IEEE Access, vol. 8, pp. 20313–20324, 2020, doi: 10.1109/ACCESS.2020.2968900.
[29] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan,"An Automated Diagnostic System for Heart Disease Prediction Based on $\chi 2$ Statistical Model and Optimally Configured Deep Neural Network," IEEE Access, vol. 7, pp. 34938–34945, 2019,doi: 10.1109/ACCESS.2019.2904800.
[30] M. R. Ahmed, S. M. Hasan Mahmud, M. A. Hossin, H. Jahan, and S. R. Haider Noori, "A cloud based four-tier architecture for early detection of heart disease with machine learning algorithms," 2018 IEEE 4th Int. Conf. Comput. Commun. ICCC 2018, PP. 1951–1955,2018, doi:10.1109/CompComm.2018.8781022.
[31] N. H. Farhat, "Photonit neural networks and learning mathines the role of electron-trapping materials," IEEE Expert. Syst. Their Appl., vol. 7, no. 5, pp. 63–72, 1992, doi: 10.1109/64.163674.
[32] A. K. M Sazzadur Rahman, M. Mehedi Hasan, S. Asaduzzaman, M. Asaduzzaman, and S. Akhter Hossain, "An analysis of computational intelligence techniques for diabetes prediction Machine Learning View project An analysis of computational intelligence techniques for diabetes prediction," Int. J. Eng.&Technology, vol. 7, no. 4, pp. 6229–6232, 2018, doi:10.14419/ijet.v7i4.28245.
[33] Y. Karaca and C. Cattani, "7. Naive Bayesian classifier," Comput. Methods Data Anal., pp. 229–250, 2018, doi: 10.1515/9783110496369-007.
[34] G. H. Tang, A. B. M. Rabie, and U. Hägg, "Indian hedgehog: A mechanotransduction mediator in condylar cartilage," J. Dent. Res., vol. 83, no. 5, pp. 434–438, 2004, doi: 10.1177/154405910408300516.
[35] Y. Karaca and C. Cattani, "7. Naive Bayesian classifier,"Comput. Methods Data Anal., pp. 229–250, 2018, doi: 10.1515/9783110496369-007.
[36] Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," Procedia Comput. Sci., vol. 85, pp. 962–969,2016, doi: 10.1016/j.procs.2016.05.288.
[37] R. Bhave, P. G. Vyawahare and S. Shamkuwar, ""A approach for Hybrid Security Management in Health Care for Secure Cloud Storage: A Review"," 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2022, pp. 1-3, doi; 10.1109/ICNGIS54955.2022.10079740.
[38] P. Gogte, S. Purve, R. Bhave, D. Pethe, P. H.Chandankhede and B. P. Thakre, "Survey: An IOT Based Approach for Hybrid Security Management in Health Care for Secure Cloud Storage," 2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP), Nagpur, India, 2023, pp. 1-4, doi: 10.1109/ICETET-SIP58143.2023.10151578.
[39] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016, no. January 2016, pp. 381–386, 2017, doi: 10.1109/ICATCCT.2016.7912028.