International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Advanced Coherent System For Predicting Cardiac Risks using Data Mining Techniques

L.Arthi, S. Sujeetha, J.Thirunavukkarasu, S. Kalaiarasi Karunya

Abstract: ***Considering health care and medical industry related data there are millions or tons of data which contains numerous hidden information. This information can be mined through which we can make effective decisions in their related industry. There are numerous far advanced methods and techniques in mining and determining the useful decisions using the retrieved useful information. Such an effective system called Coherent cardiac risk prediction system (CCRPS) is developed using neural networks in early detection or prediction of various risk level in cardiac disease. This work employs a multilayer perception neural network with back propagation as the training algorithm. This system aims in predicting the likelihood of patients getting disease related to cardiac such as CHD, a prior heart attack, uncontrolled hypertension, abnormal heart valves, congenital heart disease (heart defects present at birth) and heart muscle disease. The system uses a total of twenty-one medical related parameters such as age, sex, chest pain type, resting blood pressure (in mm Hg on admission to the hospital), serum cholesterol in mg/dl, Smoking, stress etc for prediction purpose. It enables or activated the important knowledge such as how the medical factors related to cardiac disease and patterns and the relationship to be established. Through this system we obtain effective results that have crafted its own diagnostic method or way to predict the risk level measurement of cardiac disease.***

Keywords: ***data mining, mining tools, classification, neural networks, multilayer perception neural network, back propagation, risk diagnosis.***

I. INTRODUCTION

Data mining or Knowledge discovery is the way of extracting meaningful information from a huge data. The data which obtains from various sources are collected as huge data sets which may or may not be in an orderly manner. These data contain so many secret information hidden within them. Many organisations may not be aware of such information and hence retrieving such useful information are not possible. The solution for retrieving such information with-in a span of time is done with the help of some tools and algorithms.

There are many Data mining techniques which helps in analysing the data and make much better decisions in the organizations. In this paper we focus onto the medical sciences more specifically about cardiac disease, where more patterns or some hidden data can be retrieved and treated through the above said techniques. This helps in diagnosing or predicting near to accuracy and treating the medical cases in much efficient manner. This also promotes automation in early diagnosing and treating phases. Cardiac disease is considered to be a dreadful disease which lead to sudden death or severe disability with psychological impact and affects the economic standards of a family. As per the survey reports of WHO, more than seventeen million people across the globe are dying every year because of Coronary Artery Disease. There are numerous heart disease, some are Coronary artery disease, Heart valve disease, Angina, Heart Arrhythmias, Endocarditis, Rheumatic heart disease, Cardiomyopathy, Congenital Heart Disease which occurs by numerous factors. Menopause in later stages in women, complications during pregnancy can also be a reason for heart disease and heart attacks. As a tremendous growth in healthcare industries as well as new diseases occurring day by day the healthcare data centres and so huge and they get millions and millions of data each second. Hence data mining and machine learning algorithms plays the most fundamental part of extracting meaningful information. Even some advanced machine learning techniques are used so that some basic automations are made. Prediction automation in most of the cases are always a good practise in healthcare industry for instance when a patient enters with a heart pain the specialist wont just predict with a touch examination, but allowed to a diagnostic centre after an emergency treatment. The prediction with an automation may have more parameters such that the results are almost accurate decision which helps the medical practitioners to treat the patients well.

A. Weka

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. Weka makes learning applied machine learning I recommend Weka to beginners in machine learning because it lets them focus on learning the process of applied machine learning rather than getting bogged down by the mathematics and the programming — those can come later.easy, efficient, and fun.

Manuscript published on 30 September 2019.
*****Correspondence Author(s)
L.Arthi , Assistant professor, Department of Information technology, Sri Sairam Engineering College, Chennai, Tamilnadu, India.
S. Sujeetha, Assistant professor , Department of Information Technology, Sri Sairam Institute of Technology, Chennai, Tamilnadu, India.
J.Thirunavukkarasu, Assistant professor, Department of computer science and Engineering, Sri venkateswaraa college of Technology, Sriperumbudur, Tamilnadu, India.
S. Kalaiarasi Karunya, Assistant professor, Department of computer science and Engineering, Sri venkateswaraa college of Technology, Sriperumbudur, Tamilnadu, India.

It is a GUI tool that allows you to load datasets, run algorithms and design and run experiments with results statistically robust enough to publish. Weka provides a number of small common machine learning datasets that you can use to practice on.

B. Rapid Miner

RapidMiner is mainly used for data and business understanding which helps to access, load and analyze any type of data – both structured and unstructured data like text, images, and media. It can also extract information from these types of data and transform unstructured data into structured. It helps the analyst to explore the data by immediately understanding and creating a plan automatically and extracts key information and creating an optimal data set for predictive analytics. It can blend structured with unstructured data and then leverage all the data for predictive analysis. It is equipped with an unparalleled set of modeling capabilities and machine learning algorithms for supervised and unsupervised learning. RapidMiner is flexible, robust and allow you to focus on building the best possible models for any use case that allows you to score and validate the plotted data. Unlike many other predictive analytics tools, RapidMiner covers even the trickiest data science use cases without the need to program. Beyond all the great functionality for preparing data and building models, RapidMiner has a set of utility-like process control operations that lets you build processes that behave like a program to repeat and loop over tasks, branch flows and call on system resources. RapidMiner also supports a variety of scripting languages.

C. Orange

In Data mining we call Orange as "Fruitful & Fun" which is an Open source ML tool and data visualization for a beginner as well as expertise. It has Interactive data analysis workflows with a large toolbox. Orange helps in making visual programming and data analytics with no earlier programming experience. It also performs Interactive Data Visualization with simple data analysis with visualization. The Graphic user interface in orange allows exploratory data analysis instead of coding, which results in prototyping of data analysis. It almost seems like Placing widgets on the canvas and connecting them like plug and play and evaluating them. It uses various add-ons to mine external data and eben performs natural language processing and text mining etc.It has latest features is its add-on for single-cell data analytics.

D. Theoretical Background

As far as mining is concerned, it is a process where a series of steps are carried out in an effective manner in such a way that, series of steps are carried out iteratively as a process of finding valuable information from a huge set of data. The process is an interactive process which incorporates multidisciplinary works with the following steps:

Selection Step : Generally database contains more data where all data of are not necessary for the word which has to be carried by us. Selection is a process which is used to select certain amount of data retrieved from the data bases for the the work of furter analysis of work

Pre-processing Step: This is the most essential step carried out before doing any analysis where some of the inconsistent data and certain noise which s present the given selected data are eliminated and ready further treatment

Data cleaning Step: Data cleaning is one of the most important and fundamental process which give solution for inconsistency in data han and cleans the error present in the above preprocessed data bases, in such a way that repairs the missing values and wrong values etc. Transformation Step: Transformation is the major step carried out for preparing the database which is is ready for the mining purpose in such a way that the data is aggregated and certain normalisation is done Data mining Step: The data mining follows intelligent strategic algorithms which incorporates clustering and classification kind of algorithms to extract useful patterns in order to evaluate the usefulness Interpretation/ Evaluation step: These are termed as block models which visualizes some best fitting patterns from the pre examined resultant pattern by certain predictive algorithms and evaluation patterns. This step also eliminated redundant patterns and focuses on the uniqueness of pattern.

II. LITERATURE SURVEY

☆M.S.Amin, et al(2019) Evaluated 7 classification algorithms and selected significant features with extraction techniques and came with classification model for predicting heart diseases. This resultant model is termed as a predictive model which involved hybrid mining techniques achieving 87.4% accuracy in cardiac disease prediction.

☆C.B.C Latha and S.C.Jeeva (2019) used a new ensemble classification method by combining various classifiers so that the weaker classification algorithms after combining becomes a hybrid method which inturn gives higher accuracy rates of increase of 7% and detecting the illness in early stages.

☆Purushottam, et al(2016) contributes to make accurate decisions by non-specialized physician.In this system rules are prioritized as Original, Duplicate, Pruned, Sorted ,without duplicate and executed which results in best potential in anticipating the coronary illness risk level in heart rates more precisely.

☆Ritika Chadha, et al(2016) brings a methodology of techniques like ANN, Decision Tree and Naive Bayes together for the same experimental results where, the observations reveals that ANN outperformed Naive Bayes and Decision Tree.

☆Mirpouya Mirmozaffari, et al(2017) worked in different classification algorithms on heart disease databases of more than 200 instances and eight attributes which gave good accurate results. Further to increase the accuracy he did preprocessing with different supervied and unsupervised algorithms which resulted in accuracy of about 98%

☆Abhishek Taneja(2013) designed a predictive model for heart disease detection and enhanced the reliability of heart disease diagnosis using echocardiography which used three different supervised machine learning algorithms such as J48 Classifier, Naïve Bayes and Multilayer Perceptron

Retrieval Number: K25260981119/19©BEIESP
DOI: 10.35940/ijitee.K2526.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

. Among the classifiers J48 classifier appears to be the accurate of nearly 96%.

☆Ilayaraja M and Meyyappan T (2015) devised a novice method to predict heart risk levels through frequent itemsets which are generated based on the chosen symptoms and minimum support value.

III. METHODS

The experiment was carried out on an open source database resource for heart disease. Where the dataset present is of multivariable characteristics with 75 attributes and 303 records or instances. The datasets are divided as test and train sets in 6:4 ratio afor further process. A data mining tool WEKA 3.8.2- latest version for windows was used for the experiment..

IV. PROPOSED ALGORITHM RESULTS

A new algorithm is proposed which provides much better results and an improvement over already existing methods. In this section, here elaborate the whole framework of our new approach as shown in figure 2.

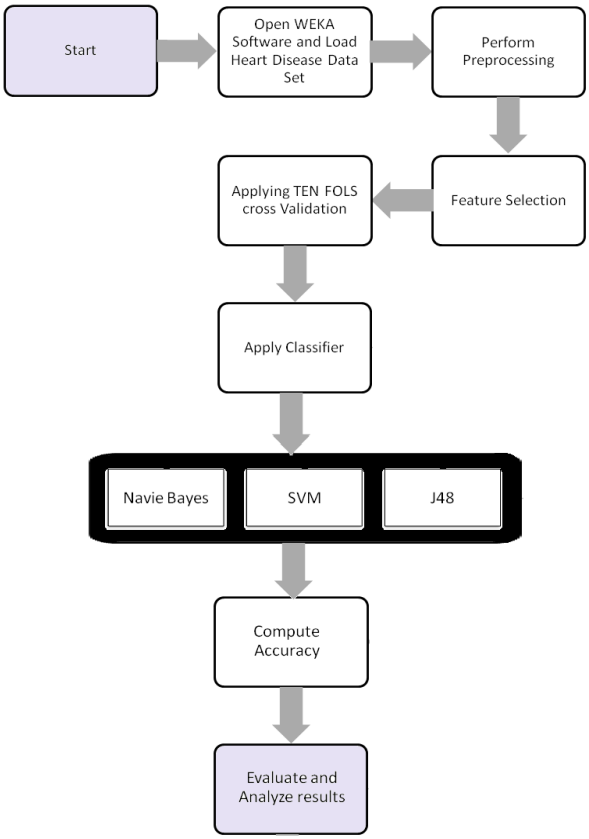

Figure 2: whole framework of our new approach

A. Input and read the dataset

Firstly, dataset selected is heart disease dataset [13]. There are 568 classified instances and 13 attributes in this dataset, which are used to evaluate performance or to predict the heart disease.

B. Select tool

After selecting the dataset, it is loaded into the WEKA tool for analyse[14].

C. Select algorithm

When the dataset is loaded, then subsequent step is to apply clustering algorithm heart disease prediction and diagnosis.

D. Cluster mode

In order to have a good measure of the performance of the clustering algorithms, the cluster mode is repeatedly performed in the WEKA tool.

E. Data Set

The heart disease dataset taken from UCI machine learning laboratory for this research paper. This database contains 13 attributes. Attribute Information included are:

Attribute 1	age
Attribute 2	sex
Attribute 3	chest pain type (4 values)
Attribute 4	resting blood pressure
Attribute 5	serum cholesterol in mg/dl
Attribute 6	fasting blood sugar > 120 mg/dl
Attribute 7	resting electrocardiographic results (values 0,1,2)
Attribute 8	maximum heart rate achieved
Attribute 9	exercise induced angina
Attribute 10	oldpeak = ST depression induced by exercise relative to rest
Attribute 11	slope = the slope of the peak exercise ST segment
Attribute 12	number of major vessels (0-3) colored by fluoroscopy
Attribute 13	thal: (3 = normal, 6 = fixed defect, 7 = reversible defect)

F. Performance analysis

The next step is to evaluate and compare the performance of different algorithms. The methodology of the study constitutes regarding collecting a set of free data mining and feature selection tools according to be tested, specifying the data sets to be used, and selecting a set of classification algorithm according to test the tools' performance. Demonstrates the overall methodology followed for fulfilling the goal of its research.

Table 1: Classification algorithm: Comparative analysis of Precision in Orange, Rapid Miner, and WEKA.

Classification Algorithms	Precision		
	Orange	Rapid Miner	WEKA
CCRPS	0.6780	0.689	0.801
Random Forest	0.779	0.784	0.818
Naïve Baize	0.824	0.845	0.837
SVM	0.817	0.587	0.845

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Figure 3: Precision Using Weka Tool , Orange Tool and RM

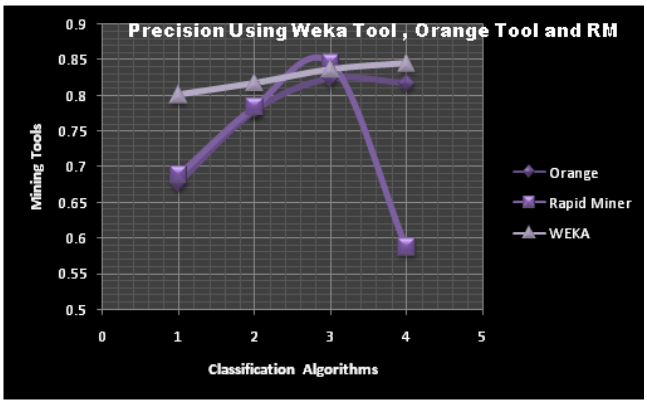

Table 2: Classification algorithm: Comparative analysis of Recall in Orange, Rapid Miner, and WEKA.

Classification Algorithms	Recall		
	Orange	Rapid Miner	WEKA
CCRPS	0.875	0.852	0.941
Random Forest	0.734	0.768	0.819
Naïve Baize	0.806	0.698	0.837
SVM	0.705	0.756	0.836

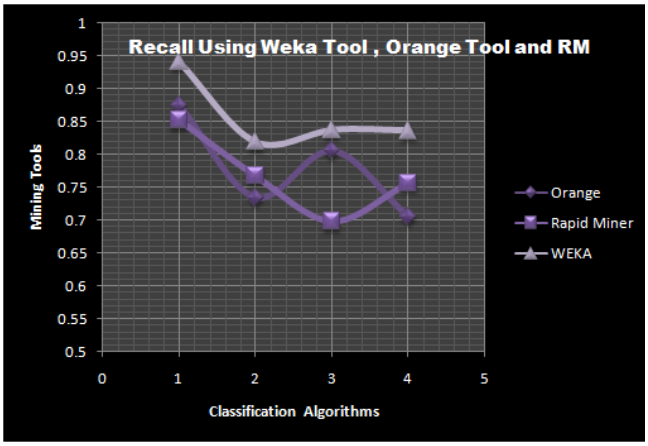

Figure 4: Recall Using Weka Tool , Orange Tool and RM

Table 3: Classification algorithm: Comparative analysis of Accuracy in Orange, Rapid Miner, and WEKA.

Classification Algorithms	Accuracy		
	Orange	Rapid Miner	WEKA
CCRPS	87.45	90.48	94.25
Random Forest	84.25	87.65	90.25
Naïve Baize	86.25	75.21	89.78
SVM	97.01	79.85	89.21

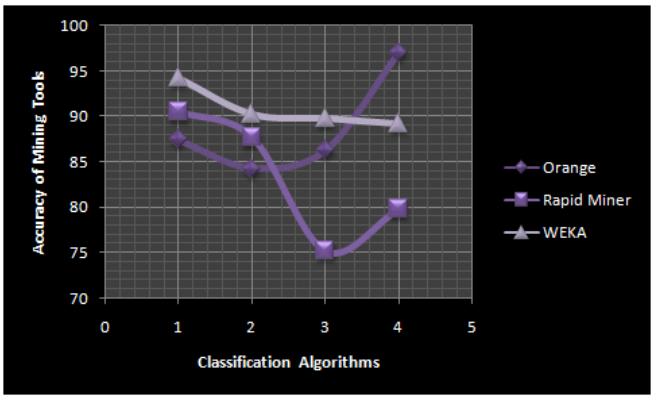

Figure 5: Accuracy Using Weka Tool , Orange Tool and RM

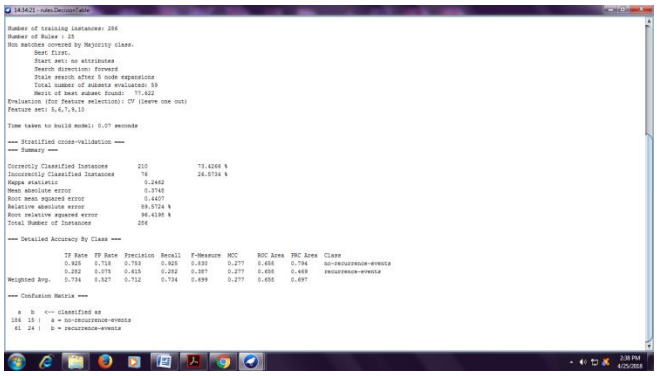

Figure 6: Analysis of Heart Disease dataset in the WEKA

Proposed methodology uses various algorithms and mainly Random Forest, Naïve Bayes and Support Vector Machine(SVM) algorithms to predict if any such cardiac risk levels exists for early automatic diagnosis with the help of above specifies instances and short retrieval time of results with a greater quality of services. This system guarrentes reduced costs in saving individuals life. One of the benefit of our work is to improve the existing methodology for better decision making by using different algorithms and feature selection methods.

V. CONCLUSION

In this paper, a Coherent cardiac risk prediction system (CCRPS) has been presented using various data mining techniques and advanced tool set for the algorithms. The CCRPS model suggests the better results and assists the medical diagnostic domain experts or even a novice in that field to plan for a better and early diagnosis of the risk level of the patient. This system performs realistically well even without retraining. Furthermore, the experimental results show that the system predicts heart disease near to 100% accuracy by using neural networks

REFERENCES

1. Amin, Hadia, Abita Devi, and Nida Ul Amin. "Predictive Analysis of Heart disease using K-Means and Apriori algorithms." (2019).
2. J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," 2016

International Conference on Circuit, Power and Computing Technologies (ICCPCT),Nagercoil,2016,pp.1-5. doi: 10.1109/ICCPCT.2016.7530265
3. Jain, Akansha, Manish Ahirwar, and Rajeev Pandey. "A Review on Intutive Prediction of Heart Disease Using Data Mining Techniques." (2019).
4. Latha, C. Beulah Christalin, and S. Carolin Jeeva. "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques." Informatics in Medicine Unlocked 16 (2019): 100203.
5. Han, Sang, Kyoung Kim, Eun Cha, Kyung Kim, and Ho Shon. "System framework for cardiovascular disease prediction based on big data technology." Symmetry 9, no. 12 (2017): 293.
6. Kumar, B. Senthil. "A Survey on Data Mining Techniques for Prediction of Heart Diseases." (2018).
7. Singh, Reetu, and E. Rajesh. "Prediction of Heart Disease by Clustering and Classification Techniques." (2019).
8. Jasim, Yaser Abdulaali, and Mustafa G. Saeed. "Developing a Software for Diagnosing Heart Disease via Data Mining Techniques." ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 7.3 (2018): 99-114.
9. Fukprapi, Chatnarong, Sotarat Thammaboosadee, and Jean Paolo Lacap. "COMBINATION OF ENVIRONMENTAL AND ECONOMIC FACTORS IN CORONARY HEART DISEASE PREVALENCE IDENTIFICATION." INTERNATIONAL JOURNAL OF APPLIED 12.1 (2019).
10. Momin, Mubeen, et al. "International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Smart Body Monitoring system using IoT and Machine Learning." (2019).
11. Rjeily, Carine Bou, et al. "Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field." Machine Learning Paradigms. Springer, Cham, 2019. 71-99.
12. Georga, Eleni I., et al. "Artificial Intelligence and Data Mining Methods for Cardiovascular Risk Prediction." Cardiovascular Computing-Methodologies and Clinical Applications. Springer, Singapore, 2019. 279-301.
13. Bohm, Allan. "Identification of KDD Problems from Medical Data." Artificial Intelligence Methods in Intelligent Algorithms(2019): 191.
14. Benhar, H., Ali Idri, and J. L. Fernández-Alemán. "A Systematic Mapping Study of Data Preparation in Heart Disease Knowledge Discovery." Journal of medical systems43.1 (2019): 17.
15. Adebayo, O. T. "Rule-based Naïve Bayes Classifier for Heart Disease Risk Prediction and Therapy Recommendation." Int J Clin Med Info 2.1 (2019): 51-59.
16. Goel, Pallavi M. "Comparative Analysis of Heart Disease Prediction." (2019).
17. Nagappan, M., et al. "Heart Disease Prediction Using Data Mining Technique." (2019).
18. Solanki, Yogita, and Sanjiv Sharma. "A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches." (2019).
19. Ma, Shaopeng, and Xiong Chen. "A data mining approach to predict risk of cardiovascular." AIP Conference Proceedings. Vol. 2058. No. 1. AIP Publishing, 2019.
20. Ayatollahi, Haleh, Leila Gholamhosseini, and Masoud Salehi. "Predicting coronary artery disease: a comparison between two data mining algorithms." BMC public health 19.1 (2019): 448.

AUTHORS PROFILE

Ms. L.Arthi, received the B.E., degree in 2010 and M.E., degree in 2012 from the Department of Computer science and engineering, Avinashilingam University, Coimbatore, Tamilnadu, India. Since 2012, she has been working as Assistant Professor in various engineering colleges (affiliated to Anna University, Chennai, Tamilnadu, India). His research interest includes IOT, Data Mining, Big Data and Data Analytics

Ms. S. Sujeetha received the B.Tech degree from the Department of Information Technology, Anna University, Chennai, Tamilnadu, India in 2005, the M.Tech degree from the Department of Information Technology , Sathyabama University, Chennai, Tamilnadu, India, in 2010. Since 2010, She has been working as Assistant Professor in Sri Sai Ram Institute of Technology (affiliated to Anna University, Chennai, Tamilnadu, India). Her research interest includes Big data ,Networking, Data Mining, cloud computing, Distributed Processing and Security. She has published more than 2 papers in international journal.

Mr.J.Thirunavukkarasu received the B.Tech degree from the Department of Information Technology, Bharath University, Chennai, Tamilnadu, India in 2009, the ME degree in Computer and Communication Engineering from the Department of Information and Communication Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, Tamilnadu, India, in 2012. Since 2012, he has been working as Assistant Professor in various engineering colleges (affiliated to Anna University, Chennai, Tamilnadu, India). His research interest includes Cloud Computing, Data Mining, Big Data, Distributed Processing and Wireless Technology. He has published more than 2 publication articles which includes 2 International journal paper publication.

Ms. S. Kalaiarasi Karunya received the B.E degree from the Department of Computer Science and Engineering from Sri Venkateswara College of Engineering and Technology affiliated by Anna University, Chennai, Tamilnadu, India in 2012, And the M.E degree from the Department of Computer Science and Engineering from DMI college of Engineering affiliated by Anna University, Chennai, Tamilnadu, India in 2015. Since 2016 she has been working as Assistant Professor (affiliated to Anna University, Chennai, Tamilnadu, India). Her interested areas in research are Cloud Computing, Big Data, Image Processing and Wireless Technology and have been published three journal paper in various publications.