



Prediction of Chronic and Infectious Diseases using Machine Learning Classifiers- A Systematic Approach

N. Komal Kumar^{1*} K. Thirunadana Sikamani¹

¹*Department of Computer Science and Engineering,
St. Peter's Institute of Higher Education and Research, Avadi, Chennai, India*
* Corresponding author's Email: komalkumarnapa@gmail.com

Abstract: Infectious and chronic diseases devastate millions of people across the world each year. Nonetheless, each type of disease substantiates differently. According to the National Centre for Health Statistics, USA, Infectious diseases or communicable diseases are the ones based on the cause, which spreads from person to person or animal to person caused by microorganisms such as bacteria or parasite and can be cured. Chronic diseases are based on the effect, which may have the origin of infectious disease, prolonged to three or more months, doesn't spread from one person to another and cannot be cured. Some chronic diseases such as cervical cancer and liver cancer have originated from infectious diseases such as human papillomavirus (HPV) and hepatitis B, C virus. This paper focuses on various machine learning classification techniques in predicting chronic diseases such as Cardio Vascular Disease (CVD), Chronic Kidney Disease (CKD), lung cancer, and infectious diseases such as hepatitis and dengue serotypes. In the analysis, ABC4.5 classifier outperformed with accuracy of 92.76 % than the other classifiers in predicting Chronic Kidney Disease (CKD), Random Forest classifier achieved an accuracy of 90.32% which is higher than Logistic regression of accuracy 83.87% in predicting hepatitis. Hoeffding classifier achieves an accuracy of 88.56% which is higher than the other classifier in predicting Cardio Vascular Disease. Multi swarm optimized Multilayer perceptron achieved an accuracy of 85.18% which is higher than the particle swarmed optimized multilayer perceptron in predicting dengue serotypes. Artificial Neural Network (ANN) classifier outperformed all the classifiers under analysis with an accuracy of 93.00 % in predicting lung cancer.

Keywords: Infectious, Diseases, Prediction, Classification, Data mining, Machine learning.

1. Introduction

Infectious diseases are feebleness caused by microorganisms such as viruses, bacteria, fungi or parasite [1]. They mostly reside in our body and they are harmless, but under certain conditions they cause disease. Infection can pass from one person to others; some can pass from animal to person [2] and may get by consuming contaminated water or food in the exposed environment. Symptoms and signs can vary depending on the causing organism but often include mild fever and cold. Infections such as measles and chickenpox can be averted by vaccines. Every infection has its sign and symptoms, common symptoms include fever, diarrhea, fatigue, muscle aches, and coughing. Bacteria are the one- cell

organisms which is responsible for disorders such as strep throat, tuberculosis, and urinary tract infections. Virus [3] causes a multitude of infectious disease ranging from the common cold to AIDS. Infectious disease such as ringworms and athlete foot are caused by fungi or parasite. Infectious disease can be caused in direct or indirect form; the direct form includes infection spread from person to person, animal to person or mother to an unborn child. These germs can spread through the fluid exchange from sexual [4] contact; indirect way can be caused by disease-causing organisms such as mosquitoes, fleas, lice or ticks. Chronic diseases are the ones which persist for a long time, mostly lasting 3 months or more as per the definitions of U.S. National Center for Health Statistics. Chronic disease cannot be prevented or cured by medication;

they include cardiovascular disease, arthritis, cancer, diabetes, epilepsy, obesity, and oral health problems. Use of tobacco, lack of physical activity and poor eating habits are the major causes of diseases. Cardiovascular diseases [5] are caused due to poor nutrition, lack of physical activity and tobacco use. Cancer [6] is a group of disease which involves the growth of abnormal cells and can spread to the different parts of the body. Diabetes [7] is the abnormal condition that results in high sugar level in the blood; their common type includes Type 1, Type 2, Prediabetes and Gestational diabetes, each having its sign and symptoms. The proposed methods can help the health authorities and medical practitioner to take precautionary measures to reduce deaths.

The work in prediction of infectious and chronic diseases is divided into sections; section 2 elaborates the related study in the field of predicting infectious and chronic diseases, section 3 overviews the datasets and its description, followed by section 4, where the methodologies is given, section 5 contains the experimental analysis and findings, finally concluded in section 6.

2. Related works

This section briefly discusses the related works carried on the prediction of chronic and infectious diseases using machine learning classifiers. Audrey Waits [8] conducted a systematic survey on disease transmission in the arctic region. The study compiled recent studies in the disease transmission field and compared with the previous results. The study concluded that increased temperature and precipitation were the causes of infectious diseases in the arctic region. The life cycle of the infectious diseases depends on the climatic factors; Lu [9] investigated the published evidence on the negative impact of climatic factors on infectious diseases, a region with high temperature were given more importance in the investigation. Xiaoxu [10] examined the scientific evidence on the impact of climate change on infectious diseases. In the analysis, three components such as infectious disease parameters, climatic factors, and selected infectious disease were considered for the analysis of establishing a correlation between infectious disease and climatic factors. The correlation between the climatic factors and infectious disease can be done by statistical methods and prediction can be done by machine learning algorithms. Komal Kumar [11] proposed a hybrid machine learning classifier in predicting Chronic Kidney Disease (CKD), the prediction follows a two-stage approach, in the first stage the classifier is subjected to a C4.5

classifier, in the second stage, the weak learners from the first stage were optimized in prediction. Komal Kumar [12] proposed a multi swarm optimized multilayer perceptron for predicting DENV serotypes, the work contained two stages in prediction, stage one deals with optimizing inputs by a multi swarm optimization technique followed by a multilayer perceptron. The study in [13] also deals with establishing a relationship between climatic factors and infectious disease. It examined the climatic factors, socio-economic factors leading to increased infection in various parts of Sweden. The relationship between the temperature and infectious disease followed a “U” shaped curve. Komal Kumar [14], proposed a machine learning classifier technique in predicting non-small lung cancer, machine learning classifiers such as support vector machine, k-NN, random forest, j48, and Artificial Neural Network were involved. In the analysis, Artificial Neural Network outperformed all the classifiers. In the biological infection, the defence against the virus/pathogens can be classified as two major components called tolerance and resistance. The work in [15] investigated the relationship between cancer and infectious disease, the study proposed an integrated approach where the disease tolerance is incorporated into cancer studies, where new therapeutic measures can be done which increases the patient’s survival and the quality of living. Komal Kumar [16], proposed an optimized random forest classifier in predicting diabetes mellitus, where a genetically optimized random forest tree classifier was used in the prediction of patients with diabetes mellitus. The interplay between bacteria and virus is presented in [17], where a model for micro biome is contributed in the development to viral infections and virally associated cancers. Machine learning tree classifiers were used in predicting diabetes mellitus, the work in [18] focused in predicting diabetes mellitus among patients using random forest, REP tree random tree and Logistic Model Tree (LMT). Komal Kumar [19] compared the performance of the machine learning classification algorithms such as logistic regression, random forest, decision tree, C4.5 and multi layer perceptron in predicting hepatitis infectious disease. A survey to classify protein structure for drug design is presented in [20], which provides the review of protein structure prediction algorithms in a distributed environment. Mareeswari [21] reviewed a machine learning technique to predict diabetes mellitus in patients using K Nearest Neighbour classification technique. The work in [22] reviews the dengue and Dengue Hemorrhagic Fever (DHF) internationally and also

characterized the risk factors, climatic factors and fatality in the endemic regions. Two cloud models were discussed in [23] in constructing cloud architectures for health care systems. The work in [24] proposed an integrating scheme for monitoring yoga activities using Internet of Things (IoT), sensors were attached to the body and monitored for physical changes. Sudhakar [25] proposed an automatic detection of brain tumour using image processing techniques such as grey level co occurrence matrix and neural networks. Cresenta [26] reviewed on the health start-ups in India and the factors affecting to its development and progression. Maheshwari [27] analyzed the occurrence of gestational diabetes from the doctor's prescription; a study was carried out to collect the data from the hospital in Chennai region. Recent developments in technology have enabled the users to identify, analyze and visualize various diseases using tweets [28] and discovery of drugs using machine learning techniques [29].

3. Dataset description

The work involves in the prediction of infectious diseases and chronic diseases such as hepatitis, DENV serotypes, chronic kidney disease, cardiovascular disease, and non small cell lung cancer. The hepatitis dataset contained 20 attributes such as sex, age, antiviral, steroid, malaise, fatigue, liver big, anorexia, liver firm, spleen_palapable, spiders, ascites, varices, bilirubin, alk_phosphate, sgot, albumin, protime, histology, and class with 155 instances obtained from University of California at Irvine repository. The dengue dataset contained two data, primary and secondary. Primary data was obtained by a questionnaires method and secondary data has been obtained from the medical centers in areas of Chennai Corporation from February 2017 to August 2017. The dataset contained 340 data observations and 28 attributes such as ID, Age, Male/Female, Temperature, Pulse, Acute fever, Platelet count, Rashes, Vomit, Abdominal pain, peeling of skin, Body ache, cold, Persistent Vomiting, Tiredness, bleeding of gums, Headache, Weakness, Tourniquet test, Bleeding from gastrointestinal tract, Fatigue, Rapid Breathing, Dengue Antigen (NS1), IgM, Deficiency of platelets, IgG, Dengue NS1 Antigen (Elisa), Narrow pulse pressure (<20mmHg). The chronic kidney diseases dataset [11] was collected from a local hospital in Chennai, Tamil Nadu, which contained 11 numeric and 14 nominal attributes with 400 instances with two class variables ckd and notckd. The cardiovascular diseases dataset was obtained from

UCI repository which contained 13 attributes with 345 observations classified into two classes "present" and "absent". The lung cancer dataset was also obtained from the UCI repository [30] which contained 15 attributes and 203 observations.

4. Methodologies

The aim of this paper is to predict infectious and chronic diseases such as hepatitis, DENV serotypes, chronic kidney disease, cardiovascular disease (CVD), and non small cell lung cancer using machine learning classifiers such as ABC4.5, Logistic Regression, Random Forest, Decision Tree, C4.5, Multilayer Perceptron, Support Vector Machine (SVM), K- nearest neighbor, Artificial Neural Network (ANN), Hoeffding tree, Logistic Model Trees (LMT), REPTree and MSO-MLP machine learning classifiers. The overall methodology is shown in Fig. 1.

The main steps in predicting infectious and chronic diseases using machine learning classifiers are as follows:

Step 1. In this step, data pre-processing is done to convert the raw data into a tidy data for analysis.

Step 2. The tidy data is subjected to 10-cross-validation and a percentage split of 60%.

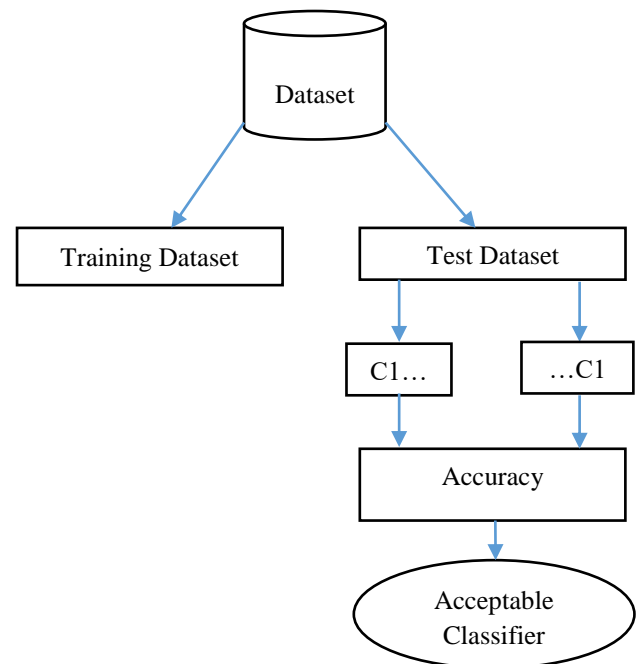


Figure. 1 Methodology

Step 3. By the percentage split, the tidy dataset is divided into test and training dataset. The training data is using to build the classifier model whereas the test dataset is used to test the classifier.

Step 4. The accuracy of the classifiers was computed using the confusion matrix represented below.

Confusion matrix [31] is represented as

		Confirmed by Observation	
		Yes	No
Predicted by test	Yes	TP (Success)	FP (Type-1 Error)
	No	FN (Type-II Error)	TN (True Negative or rejection)

TP- True Positive, FP- False Positive (Type I Error), FN- False Negative (Type II Error), TN- True Negative.

True Positive Rate (TPR), False Positive Rate (FPR), Precision and Recall can be calculated as mentioned in Eqs. (1)-(6).

1. True Positive Rate (TPR)

$$TPR = TP / (TP + FN) \tag{1}$$

2. False Positive Rate (FPR)

$$FPR = FP / (FP + TN) \tag{2}$$

3. Precision

$$Precision = TP / (TP + FP) \tag{3}$$

4. Recall

$$Recall = TN / (TN + FN) \tag{4}$$

5. Receiver Operating Characteristics (ROC)

It is a trade-off curve drawn between True Positive Rate (TPR) and False Positive Rate (FPR).

6. Accuracy (%) =

$$((TP + TN) \setminus (TP + TN + FP + FN)) \times 100 \tag{5}$$

7. ICI (%) = Accuracy (%) - 100 \tag{6}

Step 5. The final outcome represents the acceptable classifier for predicting chronic and infectious diseases.

5. Experimental analysis and findings

This section contains the experimental analysis findings for the twelve classifiers such as ABC4.5, Logistic Regression, Random Forest, Decision Tree, C4.5, Multilayer Perceptron, Support Vector Machine (SVM), K- nearest neighbor, Artificial Neural Network (ANN), Hoeffding tree, Logistic Model Trees (LMT), REPTree and MSO-MLP in predicting infectious and chronic diseases. The values obtained from the confusion matrix were used to calculate the performance metrics such as True Positive Rate (TPR), False Positive Rate (FPR), Accuracy, Precision, Recall, and Receiver Operating Characteristics (ROC).

The confusion matrix of Support Vector Machine (SVM), C4.5, PSO-MLP, Decision Tree and Adaptive boosted C4.5 for Chronic Kidney Disease dataset is represented in Fig. 2. With the confusion matrix, True Positive Rate (TPR), False Positive Rate (FPR), Accuracy, Precision, Recall, and Receiver Operating Characteristics (ROC) were obtained and represented in Figs. 3 and 4. The time taken to build the appropriate classifier is shown in Fig. 5. In the analysis, Support Vector Machine (SVM) achieved True Positive Rate (TPR) of 0.752, which is higher than the Decision tree classifier which achieved TPR of 0.744; PSO-MLP achieved TPR of 0.739 which is lower than the C4.5 classifier which achieved 0.744. C4.5 classifier achieved an accuracy of 75.32% which is higher than the Support Vector Machine (SVM) of accuracy 64.50. ABC4.5 classifier outperformed all the classifiers with an accuracy of 92.76%. The execution time of Decision Tree is 0.06 sec to build the classifier model which outperformed all the classifiers since it involves only one stage in building the model.

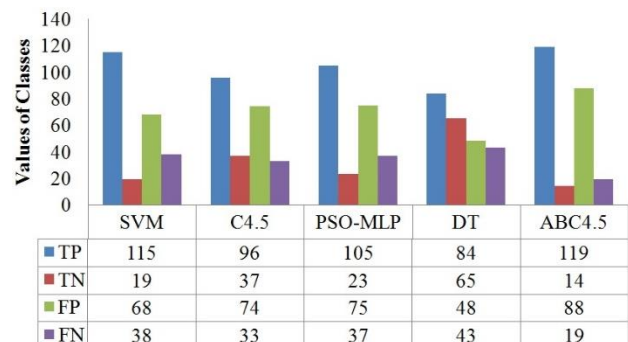


Figure. 2 Confusion matrix for chronic kidney disease dataset

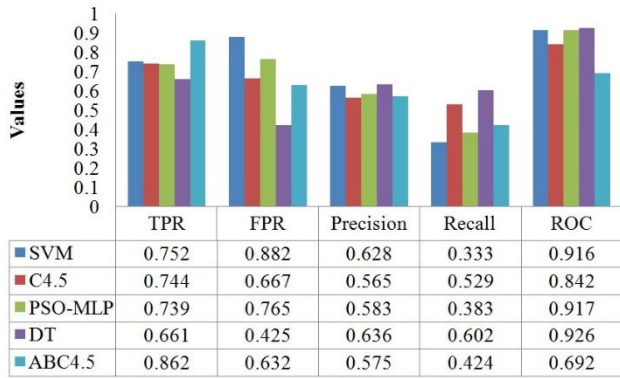


Figure. 3 Performance parameters for chronic kidney disease dataset

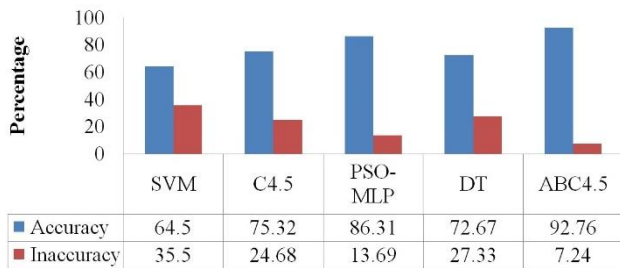


Figure. 4 Accuracy and inaccuracy for chronic kidney disease dataset

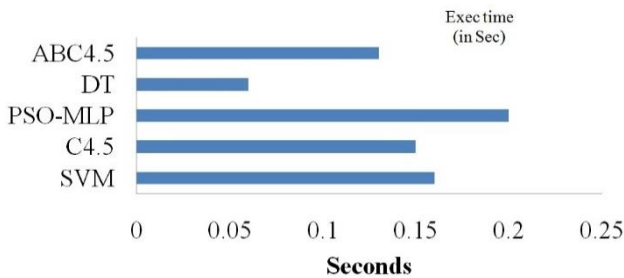


Figure. 5 Execution time for chronic kidney disease dataset

Table 1. Confusion matrix of logistic regression for hepatitis dataset

a	b	← Classified as
8	2	a=DIE
8	44	b=LIVE

Table 2. Confusion matrix of random forest for hepatitis dataset

a	b	← Classified as
8	2	a=DIE
4	48	b=LIVE

Table 3. Confusion matrix of decision tree for hepatitis dataset

a	b	← Classified as
7	3	a=DIE
5	47	b=LIVE

Table 4. Confusion matrix of C4.5 for hepatitis dataset

a	b	← Classified as
3	7	a=DIE
3	49	b=LIVE

Table 5. Confusion matrix of multi-layer perceptron for hepatitis dataset

a	b	← Classified as
7	3	a=DIE
12	40	b=LIVE

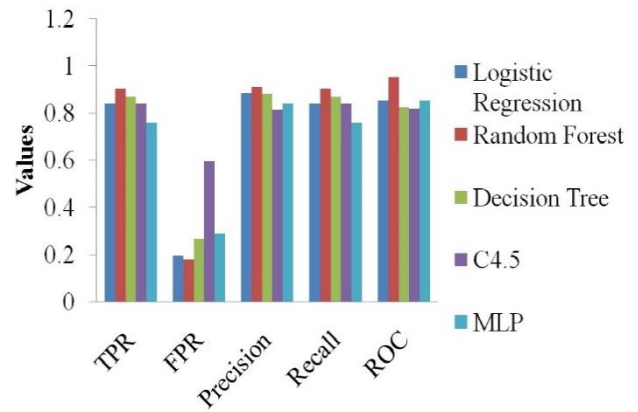


Figure. 6 Performance parameters for hepatitis dataset

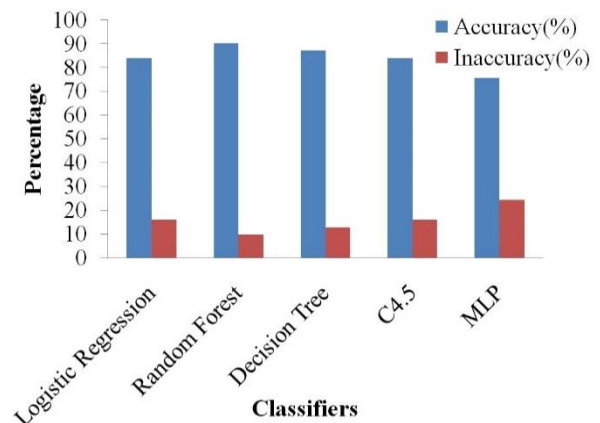


Figure. 7 Accuracy and inaccuracy for hepatitis dataset

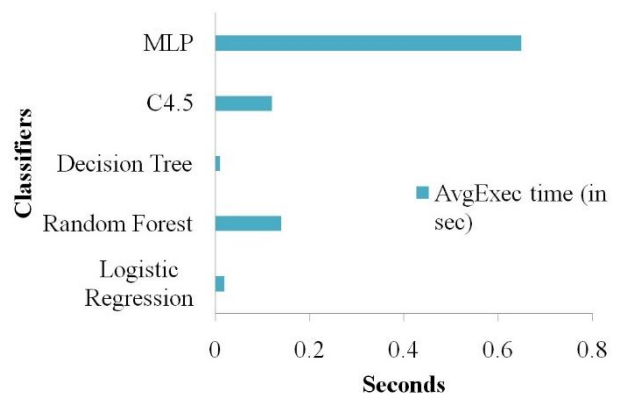


Figure. 8 Execution time for hepatitis dataset

Tables 1-5 represent the confusion matrix of Logistic Regression, Random Forest, Decision tree, C4.5 and Multilayer Perceptron classifiers for hepatitis dataset. Fig. 6 shows the performance metrics, Fig. 7 shows the accuracy and inaccuracy and Fig. 8 shows the execution time to build the model of the classifiers. In the analysis, Multilayer Perceptron (MLP) achieved a TPR of 0.758 which lower than the Logistic Regression classifier of TPR 0.8245. Random Forest classifier achieved TPR of 0.9245 which outperformed all the classifiers under analysis. Random Forest classifier achieved an accuracy of 90.32% which is higher than Logistic regression of accuracy 83.87%. The execution time of Decision tree is 0.01 sec which is lower than all the classifiers since it involves only one stage in building the classifier.

Tables 6-10 represent the confusion matrix of Random Tree, Logistic Model Tree (LMT), Hoeffding Tree, J48 and Random Forest classifiers for Cardio Vascular Disease (CVD) dataset. Fig. 9 shows the performance metrics, Fig. 10 shows the accuracy and inaccuracy and Fig. 11 shows the execution time to build the model of the classifiers. In the analysis, hoeffding tree classifier achieved TPR of 0.885, which is higher than the random forest classifier of TPR 0.798. Hoeffding classifier achieves an accuracy of 88.56% which is higher than the other classifier under analysis. Random Tree took only 0.06 sec to build the classifier model.

Table 6. Confusion matrix of random forest for CVD dataset

Classified as	Yes	No
Yes	TP=47	TN=4
No	FP=18	FN=39

Table 7. Confusion matrix of J48 for CVD dataset

Classified as	Yes	No
Yes	TP=48	TN=3
No	FP=14	FN=43

Table 8. Confusion matrix of hoeffding tree for CVD dataset

Classified as	Yes	No
Yes	TP=45	TN=6
No	FP=10	FN=47

Table 9. Confusion matrix of LMT for CVD dataset

Classified as	Yes	No
Yes	TP=45	TN=6
No	FP=12	FN=45

Table 10. Confusion matrix of random tree for CVD dataset

Classified as	Yes	No
Yes	TP=42	TN=9
No	FP=25	FN=32

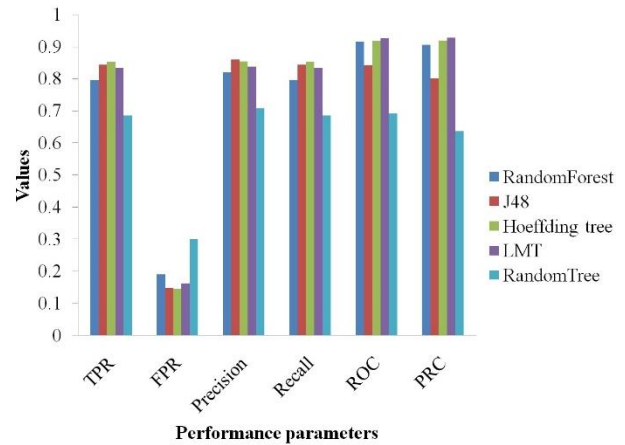


Figure. 9 Performance parameters for CVD dataset

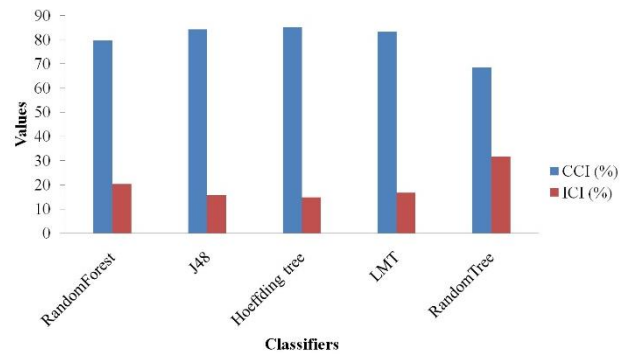


Figure. 10 Accuracy and inaccuracy for CVD dataset

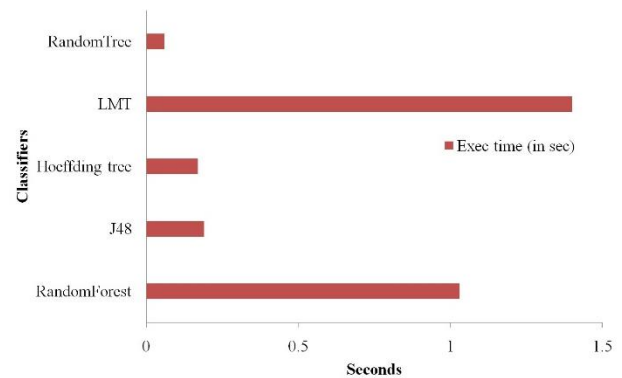


Figure. 11 Execution time for CVD dataset

Tables 11-13 show the confusion matrix of Decision Tree, Artificial Neural Network (ANN), MSO-MLP, PSO-MLP, and PSO. Figs. 12-14 represents the performance parameters, Fig. 15 represents the accuracy and Inaccuracy and Fig. 16 represents the execution time for building the model of the classifiers for dengue dataset. In the analysis, a multi swarm optimized Multilayer perceptron

Table 11. Confusion matrix for DENV-1 dengue dataset

Classifier	TP	TN	FP	FN
DT	47	4	18	39
ANN	48	3	14	43
MSO-MLP	45	6	10	47
PSO-ANN	45	6	12	45
PSO	42	9	25	32

Note: Confusion matrix obtained after percentage split of 70%

Table 12. Confusion matrix for DENV-2 dengue dataset

Classifier	TP	TN	FP	FN
DT	47	4	18	39
ANN	50	3	14	41
MSO-MLP	45	6	9	48
PSO-ANN	44	6	12	46
PSO	40	9	25	34

Table 13. Confusion matrix for DENV-3 dengue dataset

Classifier	TP	TN	FP	FN
DT	45	2	20	41
ANN	46	4	17	41
MSO-MLP	47	6	6	49
PSO-ANN	42	5	14	47
PSO	39	8	27	34

*Confusion matrix for all dataset obtained after percentage split of 60%

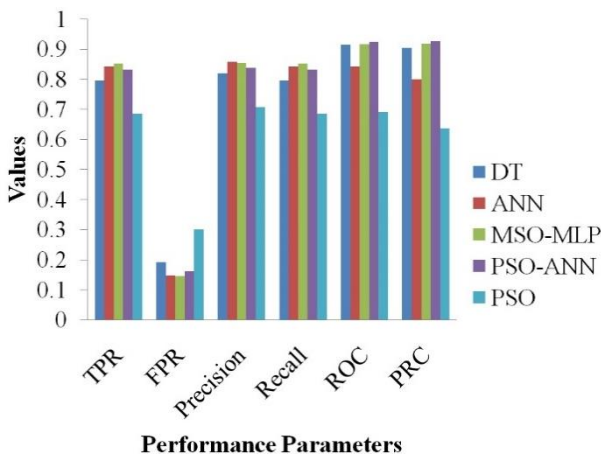


Figure. 12 Performance parameters for DENV-1 dengue dataset

achieved an accuracy of 85.18% which is higher than the particle swarmed optimized multilayer perceptron, which achieved high TRP of 0.865. It

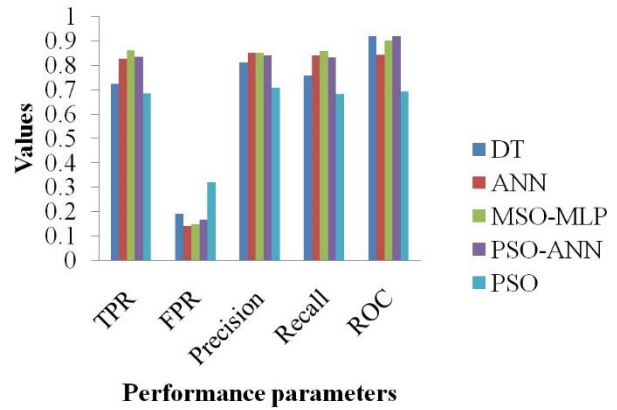


Figure. 13 Performance parameters for DENV-2 dengue dataset

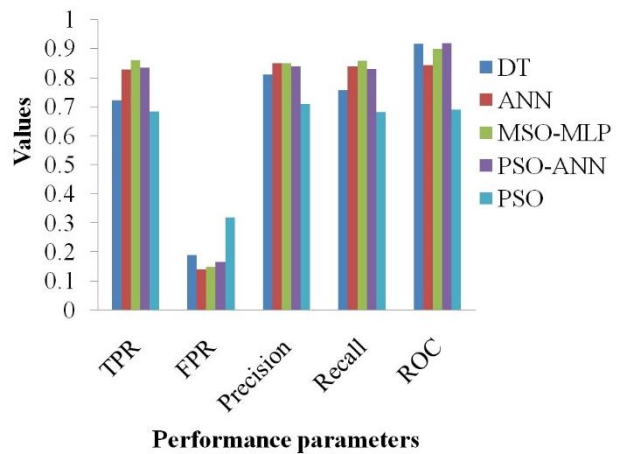


Figure. 14 Performance parameters for DENV-3 dengue dataset

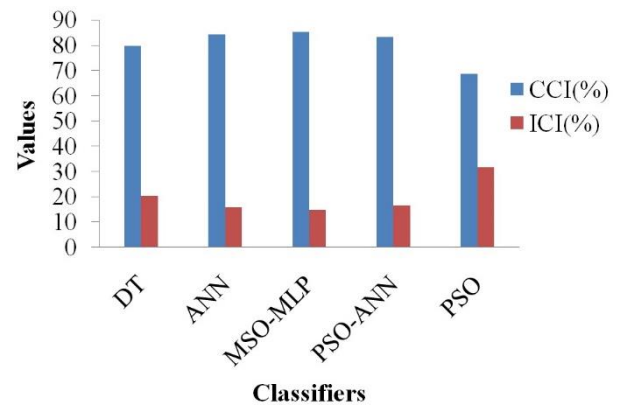


Figure. 15 Accuracy and Inaccuracy for dengue dataset

took 0.25 sec to build the classifier model which is less than the single swarm optimized multilayer perceptron.

The performance parameters of Support Vector Machine (SVM), k-NN, Random Forest, Artificial Neural Network and J48 is shown in Fig. 17. Fig. 18 represents the accuracy of the classifiers for lung cancer dataset. In the analysis, Artificial Neural Network (ANN) achieved a specificity of 0.942 which is higher than k-NN classifier which achieved

only 0.845. Random forest classifier achieved a TPR of 0.756 which has the lowest TPR. Support Vector Machine (SVM) classifier achieved an accuracy of 91.20% which is lower than the random forest classifier of accuracy 80.20%. ANN classifier outperformed all the classifiers under analysis with an accuracy of 93.40%.

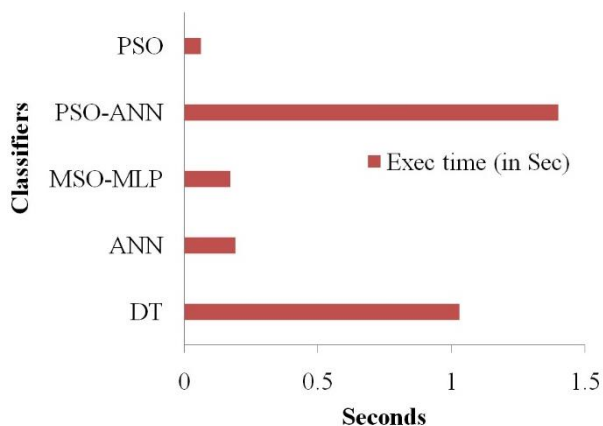


Figure. 16 Execution time for dengue dataset

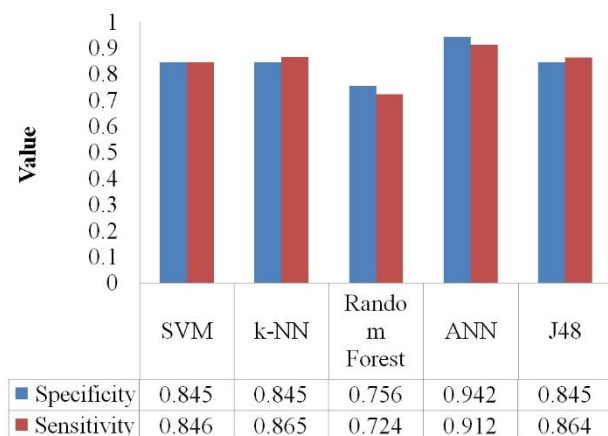


Figure. 17 Performance parameters for lung cancer dataset

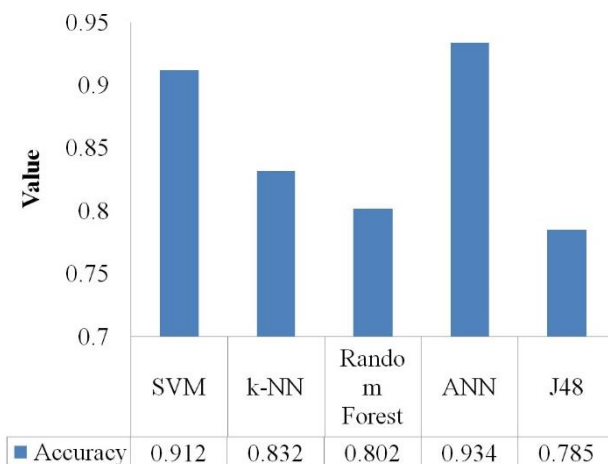


Figure. 18. Accuracy for lung cancer dataset

6. Conclusion and future direction

The performance comparison of machine learning classifiers such as ABC4.5, Logistic Regression, Random Forest, Decision Tree, C4.5, Multilayer Perceptron, Support Vector Machine (SVM), K- nearest neighbor, Artificial Neural Network (ANN), Hoeffding tree, Logistic Model Trees (LMT), REPTree and MSO-MLP in predicting infectious and chronic diseases such as hepatitis, DENV serotypes, chronic kidney disease, cardiovascular disease (CVD), and non small cell lung cancer were presented in this paper. In the analysis of chronic kidney disease, ABC4.5 hybrid classifier achieved higher True Positive Rate (TPR) of 0.862, accuracy of 92.76% which is higher than the single classifiers under investigation. ABC4.5 classifier took 0.12 sec to build the classifier model for its analysis which is lower than the C4.5 machine learning classifier. In the analysis of hepatitis disease, Random forest classifier achieved a True Positive Rate (TPR) of 0.9245, accuracy of 90.32% which is higher than the other classifiers. It took an execution time of 0.14 sec to build the classification model which is lower than the multi layer perceptron. Hoeffding tree classifier achieved a True Positive Rate (TPR) of 0.885, accuracy of 88.56% which is higher than the other classifiers in predicting the cardiovascular disease. Hoeffding tree took 0.17 sec to build the classifier model which is lower than the Logistic Model Tree (LMT) classifier. In case of dengue serotypes, MSO-MLP hybrid classifier achieved True Positive Rate (TPR) of 0.865, accuracy of 85.18% and execution time of 0.25 sec which is lower than the Decision Tree (DT) machine learning classifier. Hybrid machine learning algorithms can be developed for predicting infectious and chronic diseases in the future which can help the health authorities and medical practitioner to take precautionary measure in avoiding more deaths.

Acknowledgment

The authors are grateful to the management of St. Peter’s Institute of Higher Education and Research, Avadi, Chennai that greatly assists the research and development.

References

- [1] “India outranks US, China with world's highest net cropland area”, Retrieved 17 November 2018
- [2] E. M. Barrandeguy and M. Carossino, “Infectious Diseases in Donkeys and Mules: An

- Overview and Update”, *Journal of Equine Veterinary Science*, Vol. 65, pp. 98-105, 2018.
- [3] G. Bocharov A. Meyerhans, N. Bessonov, S. Trofimchuk, and V. Volpert, “Interplay between reaction and diffusion processes in governing the dynamics of virus infections”, *Journal of Theoretical Biology*, Vol. 457, pp. 221-236, 2018.
- [4] H. Ito, T. Yamamoto, and S. Morita, “Demography of sexually transmitted infections with vertical transmission”, *Applied Mathematics and Computation*, Vol. 348, pp. 363-370, 2019.
- [5] B. V. Reamy, P. M. Williams, and D. P. Kuckel, “Prevention of Cardiovascular Disease, Primary Care”, *Clinics in Office Practice*, Vol. 45, No. 1, pp. 25-44, 2018.
- [6] B. Bottazzi, E. Riboli, and A. Mantovani, “Aging, inflammation and cancer”, *Seminars in Immunology*, Vol. 40, pp. 74-82, 2018.
- [7] N. G. Vallianou, T. Stratigou, and S. Tsagarakis, “Microbiome and diabetes: Where are we now?”, *Diabetes Research and Clinical Practice*, Vol. 146, pp. 111-118, 2018.
- [8] A. Waits, A. Emelyanova, A. Oksanen, K. Abass, and A. Rautio, “Human infectious diseases and the changing climate in the Arctic”, *Environment International*, Vol. 121, No. 1, pp. 703-713, 2018.
- [9] L. Liang, P. Gong, “Climate change and human infectious diseases: A synthesis of research findings from global and spatio-temporal perspectives”, *Environment International*, Vol. 103, pp. 99-108, 2017.
- [10] X. Wu, Y. Lu, S. Zhou, L. Chen, and B. Xu, “Impact of climate change on human infectious diseases: Empirical evidence and human adaptation”, *Environment International*, Vol. 86, pp. 14-23, 2016.
- [11] N. Komal Kumar, R. Lakshmi Tulasi, and D. Vigneswari, “An Ensemble Multi-model Technique for Predicting Chronic Kidney Disease”, *International Journal of Electrical and Computer Engineering*, Vol. 9, No. 2, pp. 1321-1326, 2019.
- [12] N. Komal Kumar, D. Vigneswari, and B. A. S. Roopa Devi, “MSO – MLP diagnostic approach for detecting DENV serotypes”, *International Journal of Pure and Applied Mathematics*, Vol. 118, No. 5, pp. 1-6, 2018.
- [13] F. A. Mensah, G. Marbuah, and M. Mubanga, “Climate variability and infectious diseases nexus: Evidence from Sweden”, *Infectious Disease Modelling*, Vol. 2, No. 2, pp. 203-217, 2017.
- [14] N. K. Kumar, D. Vigneswari, M. Kavya, K. Ramya, and T. L. Druthi, “Predicting Non-Small Cell Lung Cancer: A Machine Learning Paradigm”, *Journal of Computational and Theoretical Nanoscience*, Vol. 15, No. 6/7, pp. 2055–2058, 2018.
- [15] S. Rao and J. S. Ayres, “Resistance and tolerance defenses in cancer: Lessons from infectious diseases”, *Seminars in Immunology*, Vol. 32, pp. 54-611, 2017.
- [16] N. Komal Kumar, D. Vigneswari, M. Vamsi Krishna, and G. V. Phanindra Reddy, “An Optimized Random Forest Classifier for Diabetes Mellitus”, *Abraham A., Dutta P., Mandal J., Bhattacharya A., Dutta S. (eds) Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing*, Vol 813. Pp. 765-773, 2019.
- [17] D. Vyshenska, K. C. Lam, N. Shulzhenko, and A. Morgun, “Interplay between viruses and bacterial microbiota in cancer development”, *Seminars in Immunology*, Vol. 32, pp. 14-24, 2017.
- [18] D. Vigneswari, N. K. Kumar, V. Ganesh Raj, A. Gugan, and S. R. Vikash, “Machine Learning Tree Classifiers in Predicting Diabetes Mellitus”, In: *Proc. of 2019 5th International Conference on Advanced Computing & Communication Systems*, pp. 84-87, 2019.
- [19] N. K. Kumar and D. Vigneswari, “Hepatitis-Infectious Disease Prediction using Classification Algorithms”, *Research J. Pharm. and Tech.*, Vol. 12, No. 8, pp. 3720-3725, 2019.
- [20] L. Dencelinx and T. Ramkumar, “Distributed Machine Learning Algorithms to classify Protein secondary structures for Drug Design – A Survey”, *Research J. Pharm. and Tech.*, Vol. 10. No. 9, pp. 3173-3180, 2017.
- [21] V. Mareeswari, R. Saranya, R. Mahalakshmi, and E. Preethi, “Prediction of Diabetes Using Data Mining Techniques”, *Research J. Pharm. and Tech.*, Vol. 10. No. 4, pp. 1098-1104, 2017.
- [22] Z. Thant, M. A. Tengku, H. Azmi, and K. Norizhar, “Dengue Virus Infections: Global Scenario, Classification, and Clinical Manifestations– A Review”, *Research J. Pharm. and Tech.*, Vol. 9, No. 1, pp. 83-90, 2016.
- [23] K. Santhi and R. Saravanan, “Performance Analysis of Cloud Computing Using Batch Queueing Models in Healthcare Systems”, *Research J. Pharm. and Tech.*, Vol. 10, No. 10, pp. 3331-3336, 2017.

- [24] M. Prasanna, T. Arunkumar, and S. Arunkumar, "A Mobile Application Based Smart System for Supporting Yoga Activities and Health Monitoring", *Research J. Pharm. and Tech.*, Vol. 10, No. 11, pp. 3863-3867, 2017.
- [25] T. Sudhakar, J. Bethanney Janney, D. Haritha, M. Juliet Sahaya, and V. Parvathy, "Automatic Detection and Classification of Brain Tumor using Image Processing Techniques", *Research J. Pharm and Tech.*, Vol. 10, No. 11, pp. 3692-3696, 2017.
- [26] L. Cresenta Shakila Motha, R. Nalini, R. Alamelu, R. Amudha, and V. Badrinath, "Health Startups in India- A Progression Towards Development", *Research J. Pharm. and Tech.*, Vol. 10, No. 12, pp. 4175-4177, 2017.
- [27] P. Maheshwari, B. Guna, M. R. Kumar, and S. R. Devi, "Drug Utilization and Evaluation of Gestational Diabetes Mellitus", *Research J. Pharm. and Tech.*, Vol. 10, No. 12, pp. 4206-4208, 2017.
- [28] S. Sivasankari, M. Kavitha, and G. Saranya, "Medical Analysis and Visualisation of Diseases using Tweet data", *Research J. Pharm. and Tech.*, Vol. 10, No. 12, pp. 4306-4312, 2017.
- [29] A. Chittora and A. Mary Mekala, "Discovery of Drug and Medicine Using Data Mining Techniques", *Research J. Pharm. and Tech.*, Vol. 10, No. 12, pp. 4147-4151, 2017.
- [30] <https://archive.ics.uci.edu/ml/index.php>
- [31] https://en.wikipedia.org/wiki/Confusion_matrix

x