International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET2073120

An Approach for improving the Prediction of Chronic Kidney Disease using Machine learning

Pooja Sharma[1], Prof. Saket J Swarndeep[2]

[1]Department of computer Engineering, L.J Institute of Engineering & Technology (Gujarat Technology University), Ahmedabad, Gujarat, India

[2]Professor, L.J Institute of Engineering & Technology (Gujarat Technology University), Ahmedabad, Gujarat, India

ABSTRACT

According the 2010 global burden of disease study, Chronic Kidney Diseases (CKD) was ranked 18th in the list of causes of total no. of deaths worldwide. 10% of the population worldwide is affected by CKD. The prediction of CKD can become a boon for the population to predict the health. Various method and techniques are undergoing the research phase for developing the most accurate CKD prediction system. Using Machine learning techniques is the most promising one in this area due to its computing function and Machine learning rules. Existing Systems are working well in predicting the accurate result but still more attributes of data and complicity of health parameter make the root layer for the innovation of new approaches. This study focuses on a novel approach for improving the prediction of CKD. In our proposed system we will implement the deep learning algorithms like Deep Neural Network. Chronic kidney disease detection system using deep network is shown here. This system of deep network accepts disease-symptoms as input and it is trained according to various training algorithms. After the network is trained, this trained network system is used for detection of kidney disease in the human body.

Keywords : Chronic Kidney Diseases, prediction, Deep Learning, Deep Neural Network

I. INTRODUCTION

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.

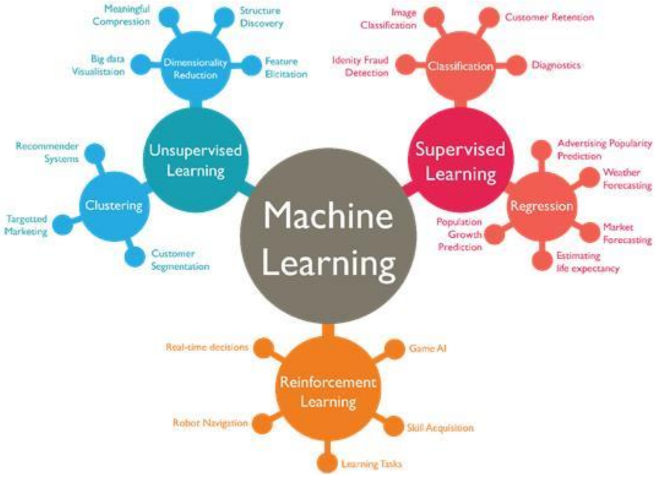


Figure 1. Machine learning methods[2]

Some machine learning methods:

Supervised learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans.

You can use supervised learning when the output data is known. The algorithm will predict new data.

Unsupervised learning

In unsupervised learning, an algorithm explores input data without being given an explicit output variable (e.g., explores customer demographic data to identify patterns).

You can use it when you do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you.

Reinforcement learning

It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.

Application of Machine Learning are in every field, few of them are here listed below.

DEEP NEURAL NETWORK

The neural network needs to learn all the time to solve tasks in a more qualified manner or even to use various methods to provide a better result. When it gets new information in the system, it learns how to act accordingly to a new situation.

Learning becomes deeper when tasks you solve get harder. Deep neural network represents the type of machine learning when the system uses many layers of nodes to derive high-level functions from input information. It means transforming the data into a more creative and abstract component

Deep Neural Network (DNN) is a multilayered Artificial Neural Network (ANN), where the multiple layers between the input and output layers. The DNN finds the correct mathematical manipulation between the input and output, whether it be a linear or non-linear relationship.

DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is represented as primitive layered composition. The extra layers allow for the composition of features from lower layers, theoretically modeling complex data with fewer units than a shallow network that performs similarly. DNNs are typically feed forward networks in which data flows from the input layer to the output layer without looping back.

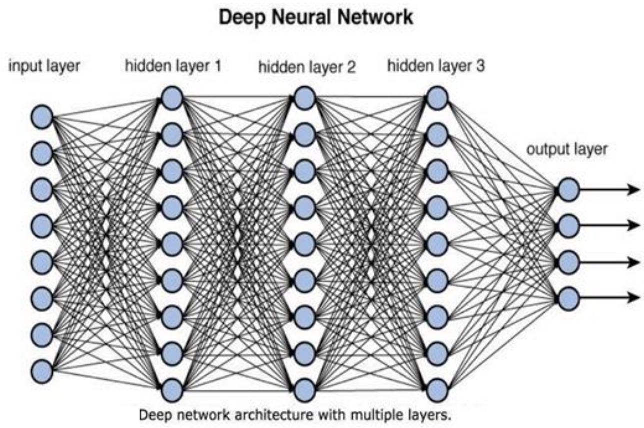


Figure 2: Deep Neural Network[3]

II. LITERATURE REVIEW

To manage the high-risk condition of CKD affecting millions of people around the world. This paper tried to analyze chronic kidney disease dataset using two important types of data analytics, statistical and predictive analytics in order to create a 100% accurate model based on machine learning algorithm. The results obtained can be a key for gaining insights from the dataset, forecasting the CKD status of the new patients and adopting good strategies for improving the safety, efficiency, and quality of the care processes toward CKD disease. [4]

This paper has presented an IoT with Cloud based CDSS framework and implemented for the prediction of CKD with its level of severity. This paper provides a systematic scheme for the CKD, and the relevant healthcare data is created by the use of UCI Repository dataset. In addition to that, the medical sensors are utilized to gather data from the CKD affected patients and maintained as the patient records. We employed an ML algorithm using DNN to carry out the learning tasks of mapping data into two classes like 'Normal' and the 'Abnormal'. The use of PSO based feature selection significantly enhances the classifier results with the classification accuracy of 98.25 whereas the accuracy is only 99.25 prior to the feature selection process. The experimental results validate the performance of the proposed model in terms of different classification measures over other classifier models. [5]

Chronic Kidney Disease (CKD) detection system using neural network is successfully implemented here. This prediction system has found high accuracy and can be alternative method for doctors, it can also be used by normal people to find probability of having CKD. This prediction system is capable of detecting chronic kidney disease with new set of inputs. Earlier Rough dataset is converted into highly preprocessed data by filling most probable values for all missing values. All four learning algorithms are tested on same dataset and all learning algorithms are giving good performance in case of highly preprocessed data. Levenberg Marquardt is found best algorithms for kidney dataset based on prediction accuracy. Based on training time, scaled conjugate gradient and resilient back propagation are found more efficient than Levenberg and Bayesian regularization. This research work gives an efficient and economical solution to CKD detection problem by using neural network. [6]

In this paper, three algorithms are used, namely, Artificial Neural Network along with Gravitational Search Algorithm (GSA+ANN), Genetic Algorithm along with ANN (GA+ANN), and K-Nearest Neighbor (KNN). They are used to assess the training performance of classification accuracy. The Gravitational Search Algorithm enhances the classification accuracy of neural networks by skipping the local minima and converging to global minima. The main goal to estimate the performance of the algorithm over kidney disease dataset is Accuracy. Comparing these three algorithms, (GSA+ANN) gives better accuracy, sensitivity, and with equal specificity which are very important metrics in a medical report. [7]

In this article, a genetic algorithm trained neural network has been proposed to efficiently detect chronic kidney disease (CKD) at an earlier stage. The local search-based learning algorithms may be trapped in local optima, the problem has been overcome using GA to train the neural network which actually tries to minimize the root-mean-squared error involved during the training phase. The performance of NN-GA-based model has been compared with NN, MLP-FFN, and Random Forest classifiers in terms of accuracy, precision, recall, and F-Measure. The results have suggested that NN-GA has outperformed other existing classifiers and is able to detect CKD with more efficiency. The future research may be focused on studying other such optimization techniques to train NNs to effectively improve the performance of NNs in real-life applications[8]

III. METHODS AND MATERIALS

3.1 PROPOSED SYSTEM

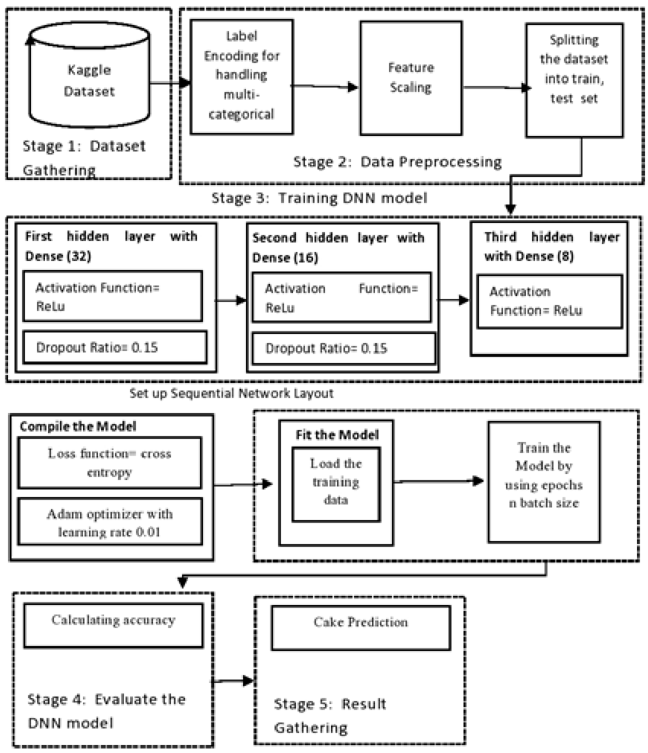


Fig 3: Flow chart of Proposed System

Following are the stages of the above proposed System:

Stage 1: collection of Dataset.

Dataset is taken form Kaggle community.

Stage 2: Data pre-processing.

In Deep Learning models if you feed the garbage data, then you should expect garbage result with high probability. So data pre-processing step is necessary for cleansing of data which is required for the given model. Data pre-processing can be done by normalization, dealing with the missing data.

Stage 3: Training of Neural Network.

In this stage the training data will feed to the input layer of DNN and then using the input layer nodes, it is forwarded to the hidden layer, which process some measures for predicting the disease and output will be forwarded to the output layer where we obtain the result or not.

Stage 4: Testing the Model.

After training the model it is required to test the model for whether the output given by it is accurate or not

Stage 5: Result

The result will show whether the patient suffering from chronic kidney disease or not.

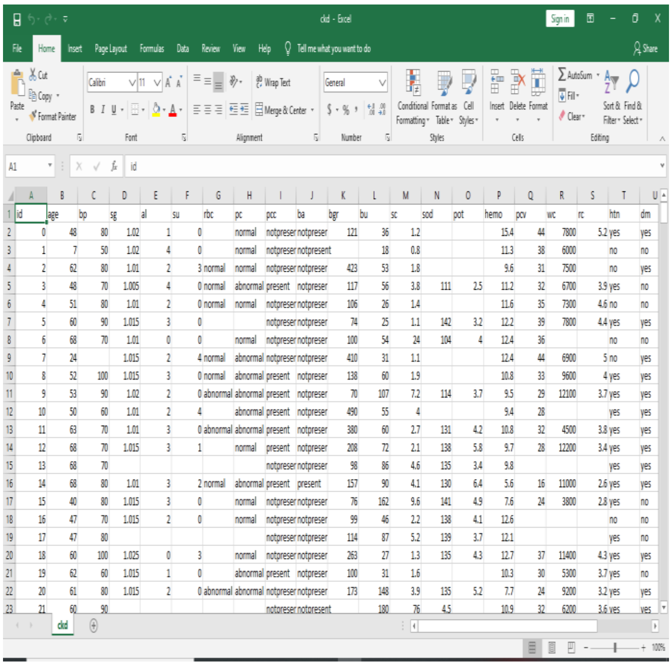
.

3.2 IMPLIMENTATION

In this Section we are showing how to implement the proposed work.

3.1.1. Data Preprocessing:

The Dataset is taken from Kaggle community, the dataset contains 26 columns in which 11 columns are of decimal n string value and 3 of Boolean.



1	id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm
2	0	48	80	1.02	1	0		normal	notpreser	notpreser	121	36	1.2			15.4	44	7800	5.2	yes	yes
3	1	7	50	1.02	4	0		normal	notpreser	notpresent		18	0.8			11.3	38	6000		no	no
4	2	62	80	1.01	2	3	normal	normal	notpreser	notpreser	423	53	1.8			9.6	31	7500		no	yes
5	3	48	70	1.005	4	0	normal	abnormal	present	notpreser	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no
6	4	51	80	1.01	2	0	normal	normal	notpreser	notpreser	106	26	1.4			11.6	35	7300	4.6	no	no
7	5	60	90	1.015	3	0			notpreser	notpreser	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes
8	6	68	70	1.01	0	0		normal	notpreser	notpreser	100	54	24	104	4	12.4	36			no	no
9	7	24		1.015	2	4	normal	abnormal	notpreser	notpreser	410	31	1.1			12.4	44	6900	5	no	yes
10	8	52	100	1.015	3	0	normal	abnormal	present	notpreser	138	60	1.9			10.8	33	9600	4	yes	yes
11	9	53	90	1.02	2	0	abnormal	abnormal	present	notpreser	70	107	7.2	114	3.7	9.5	29	12100	3.7	yes	yes
12	10	50	60	1.01	2	4		abnormal	present	notpreser	490	55	4			9.4	28			yes	yes
13	11	63	70	1.01	3	0	abnormal	abnormal	present	notpreser	380	60	2.7	131	4.2	10.8	32	4500	3.8	yes	yes
14	12	68	70	1.015	3	1		normal	present	notpreser	208	72	2.1	138	5.8	9.7	28	12200	3.4	yes	yes
15	13	68	70						notpreser	notpreser	98	86	4.6	135	3.4	9.8				yes	yes
16	14	68	80	1.01	3	2	normal	abnormal	present	present	157	90	4.1	130	6.4	5.6	16	11000	2.6	yes	yes
17	15	40	80	1.015	3	0		normal	notpreser	notpreser	76	162	9.6	141	4.9	7.6	24	3800	2.8	yes	no
18	16	47	70	1.015	2	0		normal	notpreser	notpreser	99	46	2.2	138	4.1	12.6				no	no
19	17	47	80						notpreser	notpreser	114	87	5.2	139	3.7	12.1				yes	no
20	18	60	100	1.025	0	3		normal	notpreser	notpreser	263	27	1.3	135	4.3	12.7	37	11400	4.3	yes	yes
21	19	62	60	1.015	1	0		abnormal	present	notpreser	100	31	1.6			10.3	30	5300	3.7	yes	no
22	20	61	80	1.015	2	0	abnormal	abnormal	notpreser	notpreser	173	148	3.9	135	5.2	7.7	24	9200	3.2	yes	yes
23	21	60	90						notpreser	notpresent		180	76	4.5		10.9	32	6200	3.6	yes	yes

Figure 3.1: Dataset Sample

As we can see in the above screenshot we have ingredients as the value of the field "Ingredients" in each cosmetic item of the dataset, but when we train our model and user request the query that time it is not appropriate for them to specify which ingredient they want, so this field is of no use. During this process, another problem was how the machine will understand the data because most of our data is in textual format. So for that, we have applied some preprocessing steps.

Step 1: Label Encoding for handling the textual data

```
from sklearn.preprocessing import LabelEncoder

for i in ['rbc','pc','pcc','ba','htn','dm','cad','appet','pe','ane','classification']:
    df[i] = LabelEncoder().fit_transform(df[i])
```

Figure 3.2: Label Encoding for handling of textual data

Step 2: Transform or scaling the features

```
from sklearn.preprocessing import MinMaxScaler

for i in df.columns:
    df[i] = MinMaxScaler().fit_transform(df[i].astype(float).values.reshape(-1, 1))
```

Figure 3.3: Feature scaling for making data into range

Step 3: Import the packages for partioning the data and for model

```
from keras.models import Sequential
from keras.layers import Dense, Activation
from keras.layers import Dropout
from sklearn.model_selection import train_test_split
```

Figure 3.4: Packages for data splitting n model

Step 4: Splitting data into training and testing set.

```
X_train, X_test, y_train, y_test = train_test_split(df[features], df['classification'].values, test_size=0.30, random_state=42)
```

Figure 3.5: Code for data splitting n model

3.1.2. Training the DNN model:

DNN model contains three hidden layer ,in which first contain 32 nodes and second contain 16 nodes and third contains 8 nodes with Activation Function ReLU and 0.15 dropout ratio, then we have output layer containing 1 nodes for multi-class classification with Activation Function Sigmoid. We compiled the model using 'categorical cross-entropy' loss function, 'adam' optimizer with learning rate 0.01 and accuracy metrics.

```
In [31]: # Initialising the ANN
classifier = Sequential()

# Adding the input layer and the first hidden layer
classifier.add(Dense(units = 32, kernel_initializer = 'uniform', activation = 'relu', input_dim = 24))
classifier.add(Dropout(rate = 0.1))

# Adding the second hidden layer
classifier.add(Dense(units = 16, kernel_initializer = 'uniform', activation = 'relu'))
classifier.add(Dropout(rate = 0.1))

# Adding the third hidden layer
classifier.add(Dense(units = 8, kernel_initializer = 'uniform', activation = 'relu'))
classifier.add(Dropout(rate = 0.1))


# Adding the output layer
classifier.add(Dense(units = 1, kernel_initializer = 'uniform', activation = 'sigmoid'))

# Compiling the ANN
classifier.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])

classifier.summary()

Model: "sequential_1"
```

Figure 3.6: Proposed DNN model

3.1.3. Testing the proposed DNN model:

After training the DNN model with 50 epoch ,batch size 10 and using fit () function of Keras library we can test the result of the model which is measured by accuracy matrix.

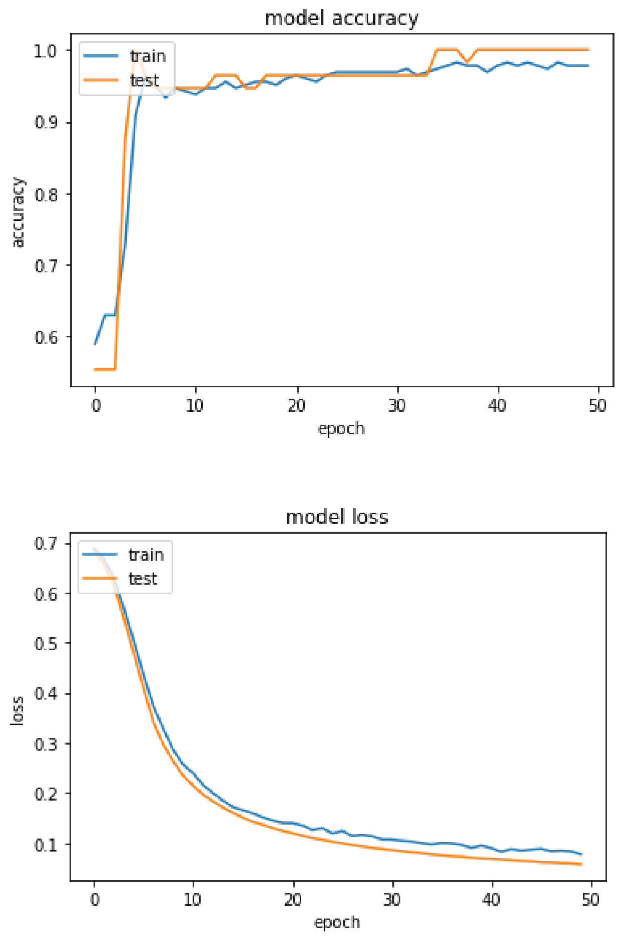


Figure 3.7 : Performance Matrix of DNN model

3.1.4 Result Gathering

As we now reached at last step of our execution process, here we will load our model which we have saved while training and testing and try to predict the class of each outfit which we have been formed using predict() function of Keras library.

```
In [ ]: y_prediction = classifier.predict_classes(X_test.values)

In [ ]: from sklearn.metrics import confusion_matrix
        cm = confusion_matrix(y_test, y_prediction)
        tn, fn, fp, tp=confusion_matrix(y_prediction,y_test).ravel()

        print(cm)

        accuracy = (cm[0][0]+cm[1][1])/(cm[0][0]+cm[0][1]+cm[1][0]+cm[1][1])
        print("Accuracy: "+ str(accuracy*100)+"%")
```

Figure 3.8 : Calculating the accuracy

IV. RESULT AND DISCUSSION

For the purpose of predicting the chronic kidney disease the Kaggle dataset had been used,deep neural network had been implemented to increase the accuracy.we got 99.1% accuracy by using the deep neural network .

V. CONCLUSION

This paper has presented the Deep neural network work to predicting Chronical Kidney Disease, This paper provides step-wise implementation of proposed work by using Kaggle dataset. The proposed system will classify whether the patient is suffering from chronic kidney disease or not and this proposed system will definitely help in improving the prediction of Chronic Kidney disease system by increasing its accuracy and prediction capability by reducing the error. In future work, we will study the application of other deep learning techniques and attempt to further improve the performance. The system can be further improved by extending the database for more better performance.

VI. REFERENCES

[1]. Madhuri Maru, "A Novel Approach for Improving Breast Cancer Risk Prediction using Machine Learning Algorithms : A Survey", IJSRSET,2019.

[2]. Matthew Mayo, KDnuggets; access on 15 November2019,https://www.kdnuggets.com/2019/04/poll-data-science-machine learning-methods-algorithms-use-2018-2019.html.

[3]. Glosser.ca; access on 17 November 2019, https://en.wikipedia.org/wiki/Artificial_neural_network#/media/File:Colored_neural_network.svg.

[4]. Safae Sossi Alaoui, Brahim Aksasse, and Yousef Farhaoui," Statistical and Predictive Analytics of Chronic Kidney Disease", Springer Nature Switzerland AG 2019.

[5]. Lakshmanaprabu S.K, Sachi Nandan Mohanty, Sheeba Rani S, Sujatha Krishnamoorthy, uthayakumar J, K.Shankar," Online clinical decision support system using optimal deep neural networks",ScienceDirect(2019).

[6]. Nilesh Borisagar, Dipa Barad and Priyanka Raval, “Chronic Kidney Disease Prediction using Back Propagation Neural Network Algorithm”, Springer Nature Singapore Pte Ltd. 2017

[7]. S. M. K. Chaitanya and P. Rajesh Kumar,” Detection of Chronic Kidney Disease by using Artificial Neural Networks and Gravitational Search Algorithm”, Springer Nature Singapore Pte Ltd. 2019

[8]. Sirshendu Hore, Sankhadeep Chatterjee, Rahul Kr.Shaw,Nilanjan Dey and Jitendra Virmani,” Detection of Chronic Kidney Disease:A NN-GA-Based Approch”, Springer Nature Singapore Pte Ltd. 2018.

[9]. Hanyu Zhang, Che-Lun Hung, William Cheng-Chung Chu, Ping-Fang Chiu, Chuan Yi Tang,”Chronic Kidney Disease Survival Prediction With Artificaial Neural Network”,IEEE 2018,

[10]. Akruti Dave, Prof. Gayatri Pandi,” Analysis Of Heart Disease Prediction System Using Artificial Neural Network”, 2018 JETIR.

[11]. EI-Houssainy A. Rady, Ayman S.Anwar,”Prediction Of Kidney Disease Stages Using Data Mining Algorithms”,Informatics in Medicine Unlocked(2019).

Cite this article as :

Pooja Sharma, Prof. Saket J Swarndeep, "An Approach for improving the Prediction of Chronic Kidney Disease using Machine learning ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 3, pp. 530-536, May-June 2020. Available at
doi : https://doi.org/10.32628/IJSRSET2073120
Journal URL : http://ijsrset.com/IJSRSET2073120