International Journal of Advanced Science and Technology
Vol. 29, No. 12s, (2020), pp. 250-258

Performance Analysis Of Machine Learning Algorithms To Foresee The Psychological Wellbeing Of An Individual

Syama phanisankar K[1], Rohan Patel[2] L Venkata Saikanth[3], Dr Manikandan K[4]
[1,2,3]B.Tech in Computer Science and Engineering,[4]Associate professor
Vellore Institute of Technology,Vellore[1,2,3,4]
syamphanisankar@gmail.com[1],patelrohan160@gmail.com[2]
venkatasaikanth@gmail.com[3],kmanikandan@vit.ac.in[4]

Abstract

In Healthcare systems, for prediction psychological health several attributes have been developed. These psychological attributes integrate the study of outcomes of attributers such as stress, anger, fatigue and depression. The existing models are built up on multiple regression method and theories give a reason to effect model of the psychological health prediction. Though, this raised algorithm depends on the outcomes of the survey. The association among these inspection replies and psychological attributes are not linear due unevenness of data. For inconsistency in the data, the dependability of the algorithm cuts and finally maintenance cost of algorithm revision rises. Similarly, when new attributes are measured, the whole algorithm should be rebuilt from the beginning. Our effort is to observes the probability of machine learning algorithms to foresee the psychological attributes grounded on the reconstructed responses. The paper compares 4 machine learning algorithms i.e. Multi-layer perceptron, Principal component analysis, Random forest, and k nearest neighbour regression, are related and the experimental results are obtained.
Keywords: *Multi-layer perceptron (MLP), Principal component analysis (PCA), Random forest (RF), and k nearest neighbour regression (kNN)*

1. INTRODUCTION

Comparing machine learning algorithms for the healthcare service to prediction of patient's psychological health. The main group of customers for psychological wellness care services are ordinary people with mental illness. In such a situation, data inequity occurs, their outcomes on psychological illness attributes trend to be altered. Thus, Healthcare schemes should take classification algorithms, which causes data unevenness problematic as well. Additional task is to adapt the model for the collected data. With the new attributes or items after taking into count, the whole model again should be reconstructed. This paper, we compare the feasibility of different types of machine learning algorithms to foresee the psychological wellness attributes built up on the newly reconstructed responses and machine learning algorithms.

2. LITERATURE SURVEY

2.1. Reference [1]:

This paper is related the current disagreements of the order of replicability for behavioural study to be analysed by means of inference for statistically analysing the forecast of prediction of psychological health conditions in researches. Here the authors claim the in addition to the systematic workflow of the psychological workflow to the research on psychological experiments based on Machine learning algorithms will reflect in increasing the accuracy and at the same time minimise replicability of subjects. In comparison to numerical interference machine learning experimental analysis of data is uncertainty model and mainly spotlights to foresee the prediction instead of inference. Author add the possible drawbacks of the result of the ML based algorithms depending on the experimental output. Disoriented usage paths towards high-optimised precision prediction is very likely to be observed by using statistical inference. To avoid such defects the advanced machine learning technique are used to cross validate each model. Machine Learning models are normally referred black boxes and different algorithms are compared for much clear the forecast of output.

2.2. Reference [2]:

Multiplayer Online Battle Games are the foremost category of gaming mode played worldwide. The player to player interaction makes these games more attractive and addictive. Previous researches show there are more intense psychological disorders in the players who engage them too much to online multi-player games. Online gaming dis-order is linked to the parameters like impulsivity, anxiety and Attention deficit hyperactivity disorder. In this paper the authors propose the approach to the PUBG and the player stats which are related to the self-esteem measure of the game to foresee whether the player suffers from the Psychological gaming disorder or Attention deficit hyperactivity disorder or Generalized anxiety disorder. This paper abstracts the game and gamers statistics in the game form the global PUBG players to foresee the Psychological gaming disorder or Attention deficit hyperactivity disorder or Generalized anxiety disorder. The experimental output gives the prediction with accuracy of the following measures Psychological gaming disorder, Attention deficit hyperactivity disorder and Generalized anxiety disorder 93.18%, 81.81% and 84.9% respectively. The numerical give a confident correlation with Internet Gaming Disorder and Attention Deficit Hyperactivity Disorder representing the dangerous effects of Multiplayer Online Battle Arena.

2.3. Reference [3]:

Number of attributes are related to different courses of depression. These attributes are built up on the contrasting groups with unknown values. This study estimates that the predicted values of lot of psychological, clinical and biological characteristics to foresee the depression and is targeted to acknowledge the best set of attributes for the machine learning algorithms. 804 unipolar psychologically ill patients are studied on a set of eighty-one demographic and other attributes are used for assessment of patient's condition and are followed up for another two years. Attributes are grouped by order (i) the occurrence of depression analysis and followed up for another two years. (ii) Disease trajectory clusters groups (fast reduction, n = 356, slow development n = 273, and long-lasting n = 175) recognized by dominant increase in growth analysis. An orderly logistic regression, post ordered by well controlled over fault of the system are used to foresee the course of depression and for evaluation of the prognostic numerical based on independent attributes. Considering the inventory of depressive symptomatology, the prediction can be rapid remission of depression with a precision of 62% and the presence of mental disability is measured with an accuracy of 66%. Other medical, mental, or biological variables did not advance the forecasting of psychological status. Among the humongous set of attributes taken, only IDS has given the solid outputs, even though this study mirrors the one of the likely methodological method the precision is good enough and further improvement is needed for the deployment of the machine learning algorithms.

2.4. Reference [4]:

To reduce the risk of jeopardising the Mental and psychological health of the people and Help the people to make correct choice to decrease the mental or social tensions and increase the awareness among the general public. The efficient application of project at different levels can be persuaded if relativity and the dynamic nature of the algorithms can be executed on psychological health of people. The author designs a form to collect the data set on Psychological wellness of youth. Even though the final conclusion will be given by a certified psychologist, who are not constantly accessible in rural areas. This project mainly concerns the rural areas with minimalistic mental and psychological health care. In these kinds of situations, the approaches of different machine learning technique and data mining are used to create an interface and the results are used to assess the psychological status of the person. This research also handles the group method of data handling (GMDH) is used and the double multi layered poly neural network with active neurons are used to highlight the classifiers for four classes of wellbeing of the adolescence. The analysed data contains the above two hundred youths ageing between twelve years to seventeen years from different schools in rural areas. The outcomes of the research indicate the percentage of correct classification for the two cases of wellbeing and not well-being which are always accurate up to more than 90% for an individual.

2.5. Reference[5]:

Now-a-days stress is a major issue among the students particularly Mental Stress. In the past this age group is considered as most care free but now tables are turned as they are facing lot of stress. Depression, suicide, heart attack, and stroke are generally caused due to stress. This research concludes the mental stress of students before the exam and during the use of internet. The main moto of this research is to calculate the mental stress of students during different points in their life. This is most underrated problem among students which is often ignored. An analysis was performed using four machine learning algorithms on how these measures have effect on the mind of a student and will likewise correlate this stress with the time spent on the internet. The collected data was from JIIT and has a sample set of two hundred and six pupils' info. The 4 different algorithms are use LR, NB, RF and SVM are used and compassion, specificity, and precision are used as parameter for a presentation. Presentation and precision of information are additional modified through ten-Fold cross authentication. Maximum possible precision is given by SVM - 85%.

3. PSYCHOLOGICAL WELLNESS DATA:

For this research, we measure different mental health conditions like anxiety, depression, stress, outrage, natural estimations e.g.: temperature, activity, and humidity etc. These measures also include some basic profiles like age, sex, occupation and many more. These measures are gathered different literature surveys on psychology and validated by several experts (like professors and post-doctoral researchers). Survey led by proficient survey agency and at last hundreds of valid responses are utilized. Above natural estimations or environmental measurement (e.g., temperature, luminance, activity, humidity, and noise level in the room) are measured using specific sensors.

Problem statement:

The main factor that increases the utility is a valuable service provided to the customers by the administration. By these services customer's happiness or well-being status will boost up. The present universal healthcare systems gather the psychological characteristics data to improve the wellbeing status but it does have few limitations. Likewise, past research works vigorously depend on complex review things. Well-being is generally classified into physical wellness and psychological wellness. In this research we focus on psychological wellness. Psychological wellness effect and causes model is developed using multiple regression method. The reliability of the model decreases by the inconsistency in the data that is collected from user survey. Non-linear nature of these indices and variables are not captured by this multiple regression method. To endure data inconsistency problem well universal healthcare system should use classification algorithms. Adaptability of the model is also one of the major challenges. whole model ought to be built once more, when new attributes or measures or items are considered. In this research, we look at the possibility of machine learning algorithms to for reconstructed responses to estimate or predict the psychological wellness. If any machine learning approach is well designed then the realism of the psychological wellness measures will be increased [12-20].
By far to our knowledge, applying machine learning algorithms to psychological wellness data lacks in previous literature serves. Machine learning algorithms perform differently for different datasets.
One which provides better prediction for any type of datasets (like seen data or unseen data) such type of algorithms should be chosen. Four best algorithms are considered in this research paper to predict the psychological wellness each one has its own advantages and disadvantages. Algorithms like, k nearest neighbour (KNN), Multi-layer perceptron (MLP), Random Forest, PCA are used in this research. For the remainder of this paper, we depict the obtaining process and data features in section 2.we briefly summarize the machine learning algorithms KNN,MLP,RF,PCA we consider for this research in section 3.Outputs obtained by this algorithms are revealed and examined in section4.We conclude this research paper with future works in section 5.

3.1. K nearest neighbour regression:

The K Nearest Neighbour (KNN) is a very simple and basic algorithm. It is a supervised machine learning algorithm which is easy-to-implement and can be used to tackle both regression and classification problem. Classification is the primary area where KNN is used. Classifies new points based on a similarity measure. The KNN algorithm predicts that the same kind of things are nearer to each other. KNN's major drawback is it gets significantly slow as the data volume expands and for which it will be impractical option in situations where predictions need to be made quickly. There are many applications of KNN in different fields like medicine, economic forecasting, Business purposes like (data compression) etc. It is used in classification, used to get missing values and measure document similarities, used in pattern recognition and gene expression and also used to get 3D structure of protein and has many more uses or applications.
KNN predicts the input data's function value f(x) using K closest neighbouring data in the training data set TD={x_1,x_2,...., x_n} as follows:

$$\hat{y} = \hat{f}(x) = \sum_{x' \in N(x)} \frac{f(x') \cdot e^{-D_j^2/\sigma}}{\sum_{x' \in N(x)} e^{-D_j^2/\sigma}}$$

where P(x) is the set of k nearest data to x in the training dataset (Figure n). P(x) should contain always odd values. Note that the predicted value f (x) is a weighted average where the weights are proportional to the Euclidean distance D2 I. This is one type of kNN regression and other weight functions can also be used.

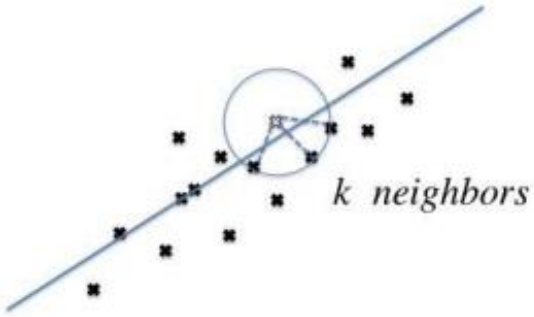

Fig-1 n-KNN Scheme

3.2. Multi-Layered Perceptron

The Multi-layered perceptron is a feedforward neural network, it consists of three layers namely input layer, hidden layers and output layer. Input layer-takes the input, hidden layer-performs all operations to the input, output layer-displays the output. It can be used for identification of images, detection of spam, prediction of voting etc. It is a learning algorithm which learns by training on a dataset. It transfers information from one to other through interconnected neurons. Values are assigned to each neuron in the network. Weights are assigned to connections between the layers and a technique called Backpropagation used to optimize the weights by using outputs as an input. An error will be occurred if there is output is changed from original output and Backpropagation returns this error back through network so that weights are readjusted and errors will be minimized. This process will be repeated till we get the correct output. The neural networks work fine by the final weights. The main advantages of Multi-layer Perceptron are: It has a capability to learn non-linear models and in real-time using partial fit.

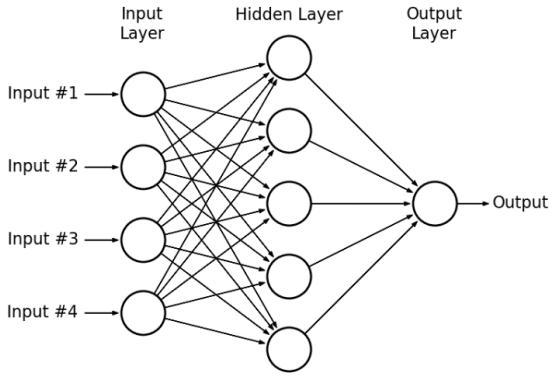

Fig- 2 Multilayer perceptron model of accuracy and loss plots

3.3. Principal Component Analysis (Pca)

PCA is a statistical system that permits to outline the data content in huge information tables by methods for an arrangement of summary indices that can be all the more effortlessly envisioned and analysed. Working: PCA calculates the covariance of matrix of data points from the data set. It calculates eigen vectors and rearranges the resultant eigen values in the descending order. It chooses k dimensions from first k eigen vectors and Changes the first n dimensional data points into k dimensions. The goal of PCA is that those recently discovered dimensions ought to permit us to anticipate/reconstruct the first dimensions. The remaking/projection error are to be limited. Applications: PCA is prevalently utilized as a dimensionality reduction procedure in things like facial acknowledgment, PC vision and image compression. It is likewise utilized for discovering designs in information of high measurement in the field of finance, data mining, bioinformatics, psychology wellness and so forth.

In 2-D dataset have only two dimensions like (count, mental health consequence). In two-dimensional coordinate system this 2-D dataset will be marked. But if we need to obtain variation, PCA finds another arranged framework in which each point has another point's (x, y) value. The axes don't signify like anything physical; they're combinations of count and mental health consequence called "principal components" that are picked to give one axes lots of variation. The extensional type of the PCA is known as multi-dimensional Principal component Analysis. PCA is increasingly helpful to expel the dimensions in the multi variate information analysis for the multi-dimensional information. singular value decomposition (SVD) is predominantly usable to gauge the PCA due to its numerical stability. And higher-order SVD (HOSVD) calculates the multi-dimensional PCA. In this paper, we use multi-dimensional PCA to the multi-dimensional data of psychology wellness and depict the aftereffects of the experimental analysis.

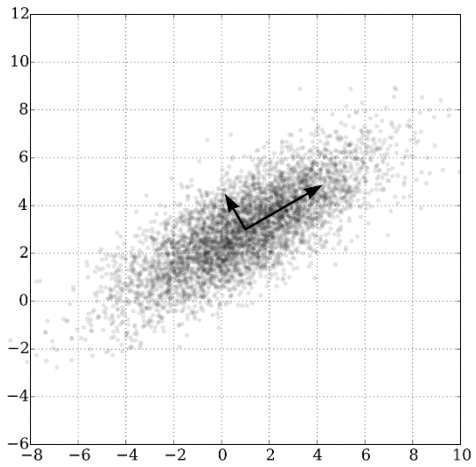

Fig-3

3.4. Random forest:

Random Forest algorithm came from the decision trees. In decision tree question asked about the data depending on the features. And each question will have either true or false solutions that divided the nodes according to that. The data points will be moved down the tree depending on the solutions of the questions. Leaf nodes won't have questions in it as final predictions will be made here. So, to find a new point just move down to the tree. But decision tree will be overfitted if we didn't limit the number of levels because it'll keep on going till it has exact one leaf node for every observation. This kind of classifications won't be completely true always. So, we can combine more decision trees to single model which is called Random Forest.

Random Forest is a classified algorithm where it creates the forests randomly. The accuracy is directly proportional to the number of trees formed. There are two steps in the RF algorithm. Random forest creation is the first one and predicting by classifiers to the one which is created in the first step is the second step.

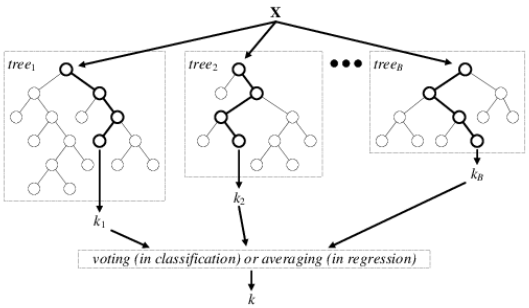

Fig-4

Pseudocode:
1. Takes the input and follows the rules set by the decision trees which are created randomly and it stores the predicted output.
2. It calculates the votes of the output that are predicted before.
3. Uses Random Forest Algorithm to find the final predicted output by calculating high voted output.

Applications:
Random Forest works efficiently in bank services, medical care, stocks and also in E-commerce. The main advantage of the Random Forest Algorithm is It avoids overfitting, identifies the most important things in the dataset which is trained.

4. CONCLUSION

From the data we used the following algorithms the accuracy of the prediction is as follows:

Principal component analysis – 63.91%

k Nearest Neighbour – 48.14%

Naive Bayes – 63.75%

Random Forest Classifier – 65.75%

So the best algorithm for use is Random forest. We suggest that the random forest is very accurate of all the above algorithms.

Fig-5

The stats of the data we used for training and testing the algorithms

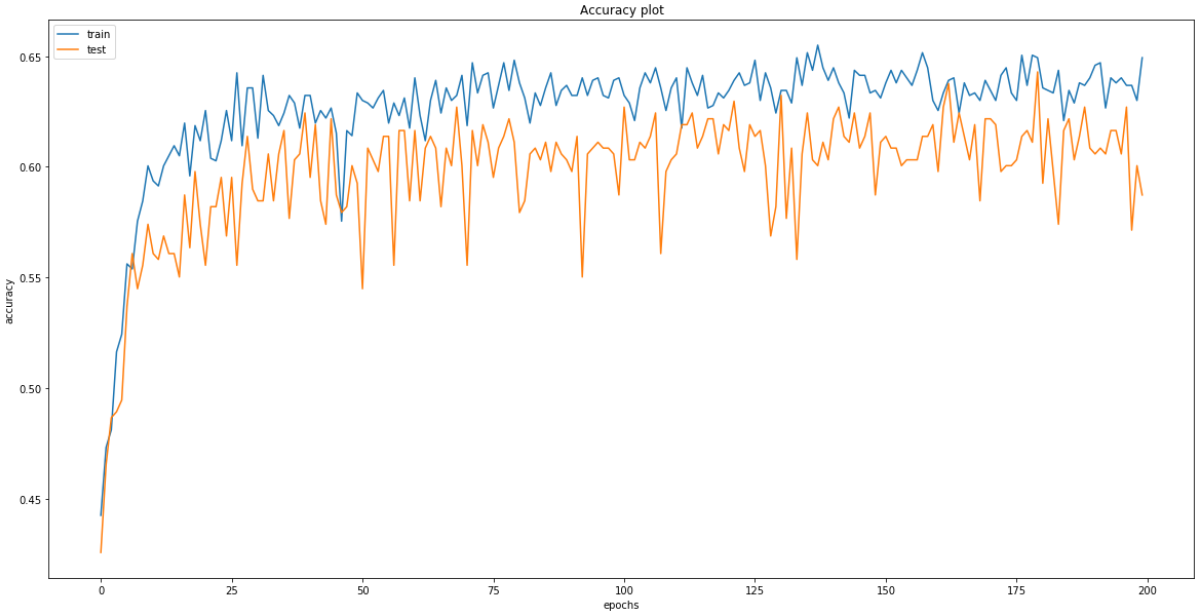

Fig-6
Accuracy plot

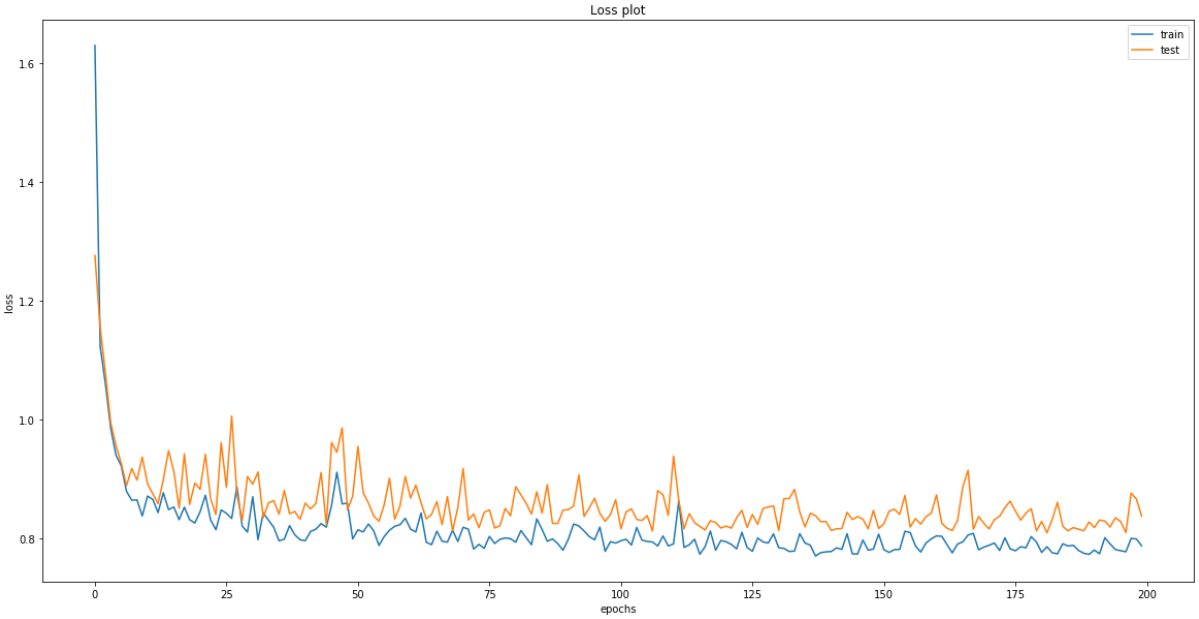

Fig-7
Loss plot

References

1. Graziella Orrù., Merylin Monaro., Ciro Conversano., Angelo Gemignani., Giuseppe Sartori., "Machine Learning in Psychometrics and Psychological Research", Frontiers in Psychology, Vol 10 (2020), Frontiers Media S.A., 2020.
2. Aggarwal, S. Saluja, S. Gambhir, V.Gupta, S.Satia, S.P.S. , "Predictinglikelihood of psychological disorders inPlayerUnknown'sBattlegrounds (PUBG) players from Asian countries usingsupervised machine learning", Addictive Behaviors, February 2020, 101.
3. Dinga R., Marquand AF., Veltman DJ., Beekman ATF., Schoevers RA., van Hemert AM., Penninx BWJH Schmaal L., "Predicting the naturalistic course ofdepression from a wide range of clinical,psychological, and biological data: a machinelearning approach." Translational

psychiatry[Transl Psychiatry] 2018 Nov 05; Vol. 8 (1), pp. 241. Date ofElectronic Publication: 2018 Nov 05.
4. Tyulyupo, S.V., Andrakhanov A.A., Dashieva B.A., Tyryshkin, A.V., "Adolescents Psychological Well-Being Estimation Based on a Data Mining Algorithm" 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 2018 IEEE 13th International Scientific and Technical Conference on. 1:475-478 Sep,2018.
5. Ahuja., Ravinder., Banga., Alisha., "Mental Stress Detection in University Students using Machine Learning Algorithms." Procedia Computer Science; 2019, Vol. 152, p349-353, 5p.
6. Diener E, "Assessing well-being: The collected works ofEd Diener", Social Indicators Research Series,Springer, TheNetherlands, 2009.
7. Diener E, Suh E, and Oishi S, "Recent findings onsubjective well-being", Indian Journal of ClinicalPsychology, Vol. 24, No. 1, 1997, pp. 25–41.
8. Drucker H, Burges CJ, Kaufman L, Smola A, and VapnikV, "Support vector regression machines", Advances inneural information processing systems, MORGANKAUFMANN PUBLISHERS, 1997, pp. 155–161.
9. Atkeson CG, Moore AW, and Schaal S, "Locallyweighted learning for control", Artificial IntelligenceReview, Vol. 11, No. 1-5, 1997, pp. 75–113.
10. Chang CC and Lin CJ, "LIBSVM: a library for supportvector machines", ACMTransactions on IntelligentSystems and Technology, Vol. 2, No. 3, 2011, pp. 1-2.
11. Specht DF, "A general regression neural network", IEEEtransactions on neural networks, Vol. 2, No. 6,1991,pp. 568–576.
12. S. Velliangiri, P. Karthikeyan & V. Vinoth Kumar (2020) Detection of distributed denial of service attack in cloud computing using the optimization-based deep networks, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2020.1744196
13. Praveen Sundar, P.V., Ranjith, D., Vinoth Kumar, V. et al. Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture. Int J Speech Technol (2020). https://doi.org/10.1007/s10772-020-09686-y
14. Vinoth Kumar V, Karthikeyan T, Praveen Sundar P V, Magesh G, Balajee J.M. (2020). A Quantum Approach in LiFi Security using Quantum Key Distribution. International Journal of Advanced Science and Technology, 29(6s), 2345-2354.
15. Umamaheswaran, S., Lakshmanan, R., Vinothkumar, V. et al. New and robust composite micro structure descriptor (CMSD) for CBIR. International Journal of Speech Technology (2019), doi:10.1007/s10772-019-09663-0
16. Karthikeyan, T., Sekaran, K., Ranjith, D., Vinoth kumar, V., Balajee, J.M. (2019) "Personalized Content Extraction and Text Classification Using Effective Web Scraping Techniques", International Journal of Web Portals (IJWP), 11(2), pp.41-52
17. Maithili, K , Vinothkumar, V, Latha, P (2018). "Analyzing the security mechanisms to prevent unauthorized access in cloud and network security" Journal of Computational and Theoretical Nanoscience, Vol.15, pp.2059-2063.
18. V.Vinoth Kumar, Ramamoorthy S (2017), "A Novel method of gateway selection to improve throughput performance in MANET", Journal of Advanced Research in Dynamical and Control Systems,9(Special Issue 16), pp. 420-432
19. Dhilip Kumar V, Vinoth Kumar V, Kandar D (2018), "Data Transmission Between Dedicated Short-Range Communication and WiMAX for Efficient Vehicular Communication" Journal of Computational and Theoretical Nanoscience, Vol.15, No.8, pp.2649-2654
20. Kouser, R.R., Manikandan, T., Kumar, V.V (2018), "Heart disease prediction system using artificial neural network, radial basis function and case based reasoning" Journal of Computational and Theoretical Nanoscience, 15, pp. 2810-2817