SUPPORT VECTOR REGRESSION FOR PREDICTING COVID-19 CASES

B Suvarna[1],Bhagya Lakshmi Nandipati[2], M Nirupama Bhat[3]

[1,2,3]Department of CSE, VFSTR Deemed to be University, Vadlamudi, Guntur, AP, India

ABSTRACT

This study uses the Machine Learning models to forecast the number of upcoming cases affected by COVID-19. Particularly two models for forecasting such as Support Vector Regression and Linear regression (LR)are used in this study to predict future number of cases of COVID-19. Number of upcoming covid-19 cases is going to be happen and deaths are predicted by using these two models for the next 365 days. The predictions are done for India and the worldwide. Between these two SVR is performing well compared with Linear Regression.

INTRODUCTION

Virus is an infectious agent. This virus resides in the living cells of the different life forms like humans, animals, plants, and bacteria. It is a small infectious agent inside the living host cells, where the host cell shall produce thousands of original copies [1-3]. This spread of viruses can be in many ways. Viral infections can cause disease like life-threatening infections in humans, animals, and plants.

Corona viruses are family of virus causes saviour illness in animals and humans. COVID-19 is the infectious disease caused by a corona virus, which is recently discovered. This virus mostly affects a person's respiratory system. It is a disease that was detected at the end of 2019 i.e on 31st December, 2019 after that the first case is confirmed in outside of the china i.e in Thailand on 13th January, 2020, in Japan it is confirmed on 15th January 2020, In USA it is confirmed on 21st January 2020 and announced a pandemic on March 11. Similarly the first death is confirmed in china is on 11th January 2020.In Wuhan, China, the first case was confirmed officially. Since then, there has been an exponential growth in the number of such cases around the globe.

As of June 12^{th,} 2020, the total reported cases reached 7,597,341 out of which 423,844 have died. How it spreads: This is mainly spreading through the droplets either through coughing or through sneezing, which is produced by the infected person. Many of the infections are coming through the direct contacts of a person, which causes one way of spreading. It may spread the disease by touching some contaminated surface or fabric and then touching one's mouth, nose, or eyes [4-7]. For people with the infection incubation period can be from one to fourteen days. And disease can also spread without any severe symptoms. So far there is no vaccine and medicine for preventing this virus but only we can limit the spread of virus by maintaining the distance between persons and maintaining hygiene. Some small work is already done in the analysis of COVID-19 [7-10].

Basically, it is a forecasting problem, several forecasting algorithms based on statistical theory [11] and ARIMA models are used to forecast future cases [12-13]. This paper attempts to forecast the future cases using Support Vector Regression and Linear Regression models.

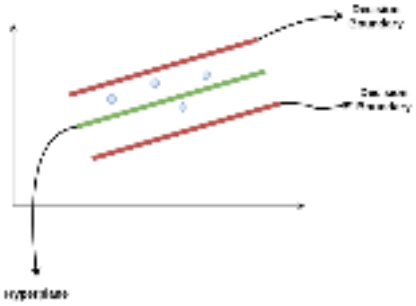
1. METHODOLOGIES

Two models are used in this study and those are Support Vector Regression and Linear Regression.

Support vector Regression

Support vector machines have been used widely for the classification problems in the supervised learning. Support Vector Regression uses the similar principle of Support Vector Machine, but this Support Vector Regression is used for the regression problems [14-16]. The model generated by classification of support vectors depends only on a subset of training data, as the cost function for building the model is not concerned with training points beyond the margin. Similarly, the model generated by Support Vector Regression depends solely on a subset of training data, since the cost function ignores samples whose prediction is close to their target.

Assume that there are two decision boundaries and one hyper plane. And all the points are within the decision boundary. Objective of SVR is to find the best fit line which is hyper plane which has maximum number of data points [17-20].



Assume that the equation of hyper plane is

$y=wx+b$

The equations of decision boundaries will be

$wx+b=+a$ and $wx+b=-a$

Thus the equation of hyper plane that should satisfy SVR as

$-a<y-wx+b<+a$

2. Linear Regression

The problem of Regression is to find the best fit line based on the training data where the input data contains the real numbers. Linear regression is a supervised learning algorithm, but the outcome of regression is real number rather than a class label. This can be used for finding the relationship between two continuous variables [21-23].

Let X is a numerical which is independent variable and Y is a numerical vector which is based on the values of X. Then the relationship between two can be found by using the equation $Y=a_0+a_1X$

2. EXPERIMENTATION AND RESULTS

3.1 Flow of Execution:

For the experimentation, the data set is taken from Kaggle. **Flow of execution**: For the prediction, take the dataset as input and then pre-process the data [24-26]. Once the dataset is ready then apply the model on dataset and make the predictions by using test dataset and flow of execution is given in the figure1.

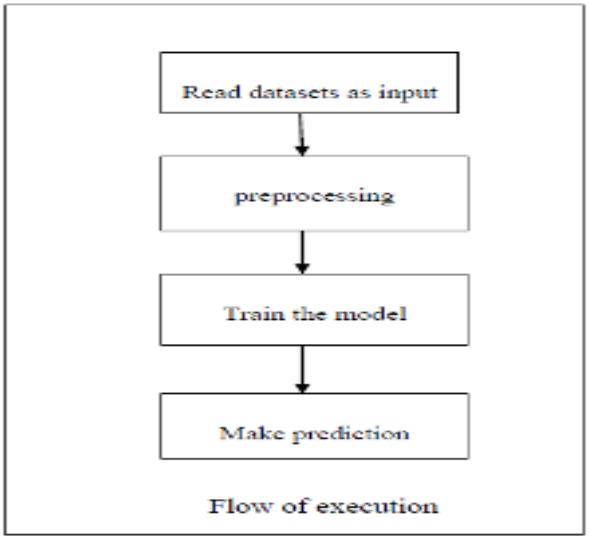


Figure 1: Flow of execution for prediction

We have extracted the data for India and taken the worldwide to get the confirmed cases and deaths. 85% of the data is taken as training and 15% of the data is used for testing. Description of the datasets is given in the tables from table1to table6. By using these models, we have predicted for the cases to the number of days i.e. for the next 60, 90,120,150,180 and 365 days.

Dataset for worldwide confirmed cases

Date	31-01-2019	01-01-2020	---	16-06-2020
Number of confirmed cases	17	27	--	8000847

Dataset for worldwide deaths

Date	11-01-2020	12-01-2020	---	19-06-2020
Number of deaths	01	01	--	4,55,777

Dataset for India confirmed cases

Date	12-03-2020	13-03-2020	---	14-06-2020
Number of confirmed cases	74	75	--	320922

Dataset for India deaths

Date	13-03-2020	01-01-2020	---	14-06-2020
Number of deaths	01	02	--	9195

3.2 Evaluation metrics

The performance of Support Vector Regression and Linear Regression is measured in terms of Mean Absolute Error, Mean Squared Error, Root Mean Squared Error and R^2 [27-31]. The formulas for each metric can be given as in the following equation.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right|$$

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2$$

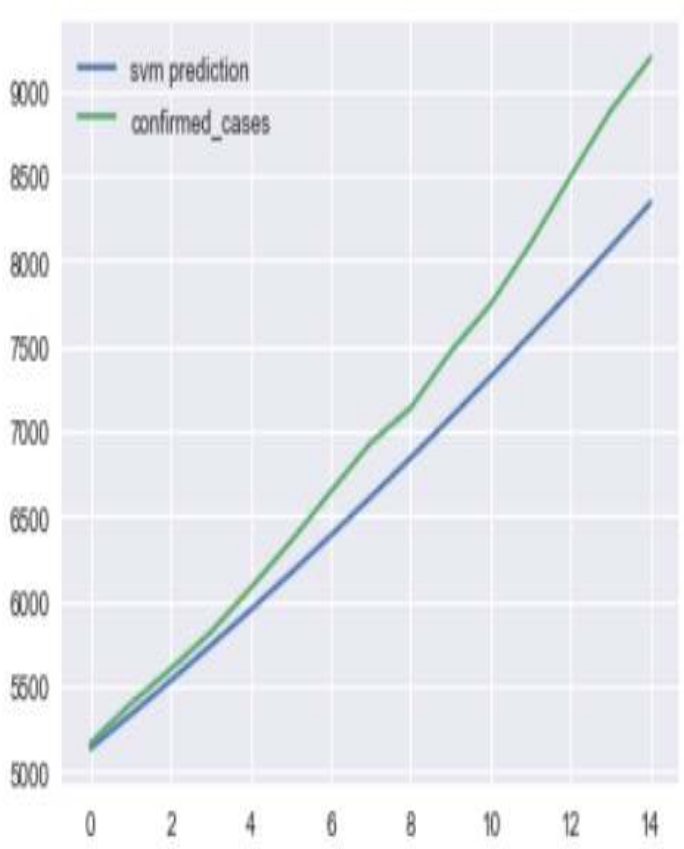
$$RMSE = \sqrt{MSE}$$

$$R^2 = \frac{\text{variance explained by model}}{\text{total variance}}$$

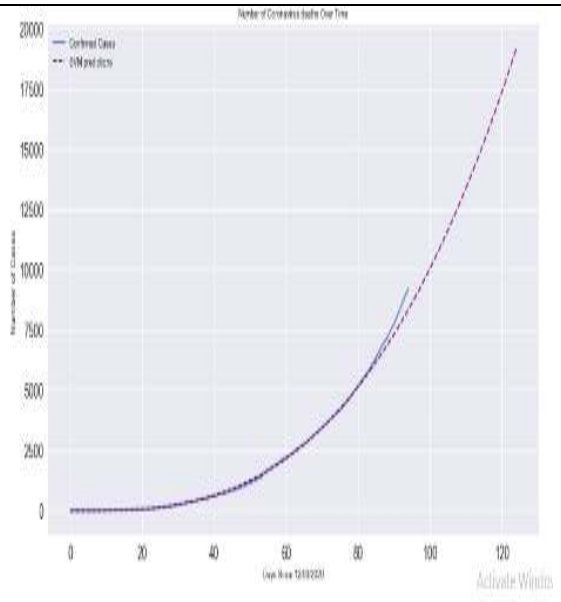
3.3 Results

By using these models, we have predicted for the cases to the number of days i.e. for the next 90,120,150,180 and 365 days. Results obtained by using the models are given in the form of tables from table7 to table13 and also in the form of graphs for the clear understanding by looking at visually. The graphs can be viewed from figure2 to figure7.

Support Vector Regression for 30 days



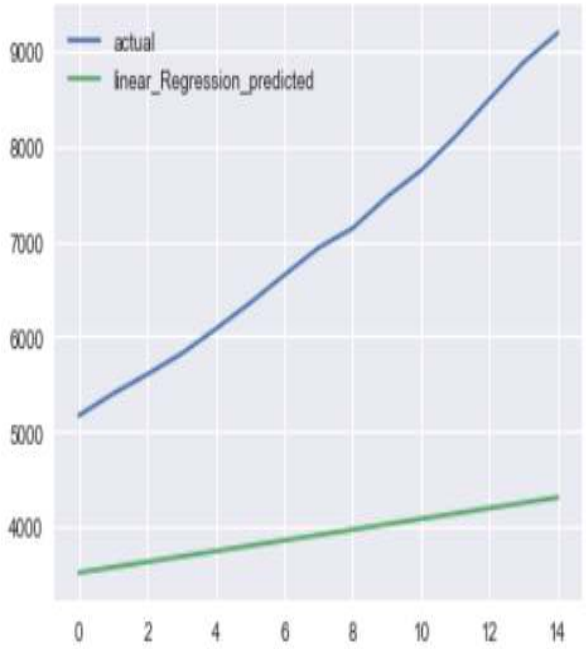
From the above graph we can observe that how the SVR is perfroming on testing data. Data is confirmed deaths for India.



From the above graph we can observe that how the SVR is predicting deaths for the next 30 days from on testing data. Data is confirmed deaths for India.

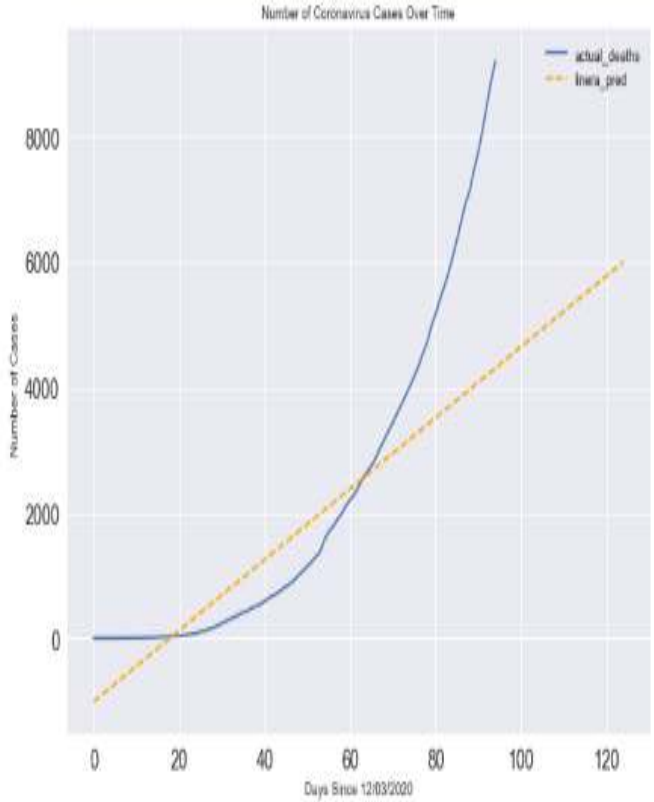
Linear Regression

Linear Regression for next 30 days



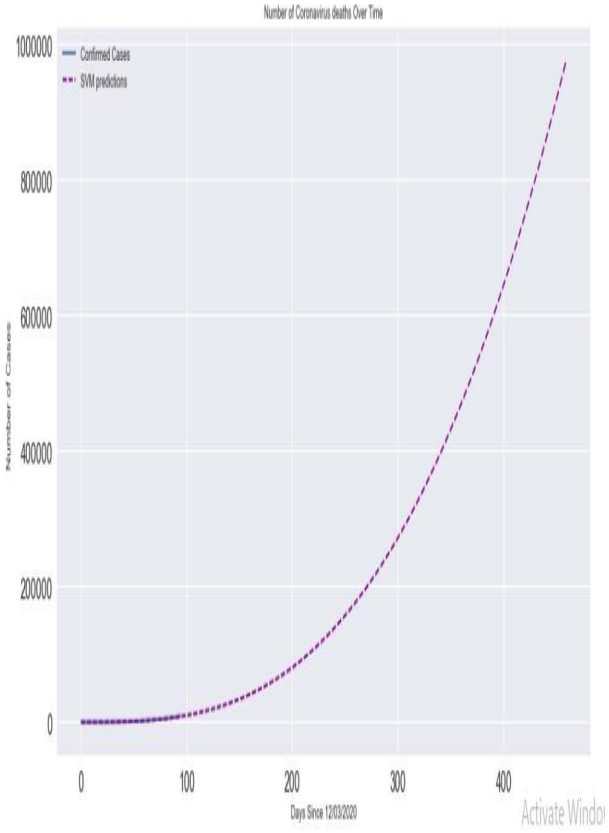
From the above graph we can observe that how the linear regression is perfroming on testing data. Data is confirmed deaths for India.

From the below graph we can observe that how the SVR is predicting deaths for the next 30 days from on testing data. Data is confirmed deaths for India.

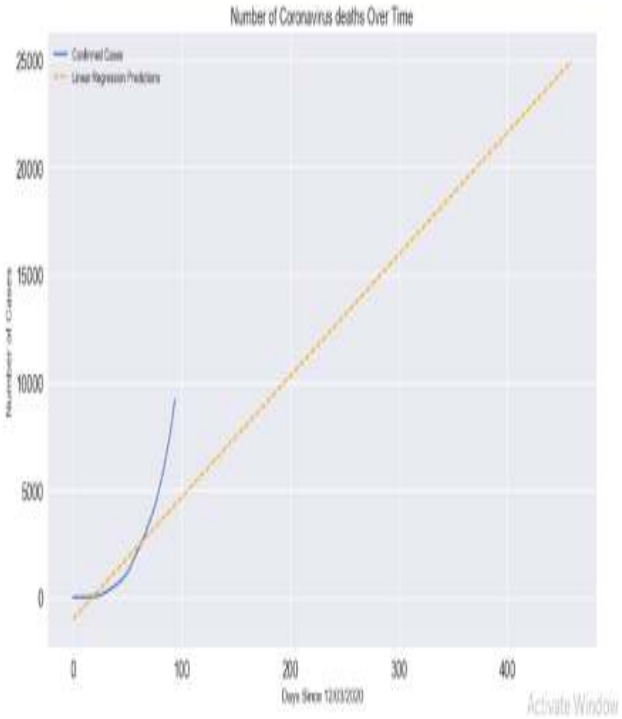


Deaths for India for 365 days

From the below graph we can observe that how the SVR is predicting deaths for the next 365 days from on testing data. Data is confirmed deaths for India.



From the below graph we can observe that how the Linear Regression is predicting deaths for the next 365 days from on testing data. Data is confirmed deaths for India.



Results obtained usig different models are given in the form of tables

Table 4. SVR predictions

Days	Date	RMSE	MAE	Predicted cases
90 days	05/06/21	35760	34060	2019604
120 days	05/07/21	36029	34330	3173988
150 days	04/08/21	35760	34060	4711281
180 days	03/09/21	35757	34057	6672126
365 days	09/03/21	35760	34060	31369437

Table 5. Linear Regression predictions

Days	Date	RMSE	MAE	Predicted cases
90 days	05/06/21	124488	119545	309557

120 days	05/07/21	124488	119545	365573
150 days	04/08/21	124488	119545	421589
180 days	03/09/21	124488	119545	477605
365 days	09/03/21	124488	119545	823035

Table 6. Death prediction using SVR

Days	Date	RMSE	MAE	R-square value	Predicted deaths
90 days	05/06/21	429	340	0.80	62737
120 days	05/07/21	429	340	0.80	98719
150 days	04/08/21	429	340	0.80	146341
180 days	03/09/21	429	340	0.80	207226
365 days	03/07/22	432	343	0.80	974251

Table 7. Death prediction using LR

Days	Date	RMSE	MAE	R-square	Predicted deaths
90 days	05/06/21	429	340	-176	62737
120 days	05/07/21	429	340	-176	98719
150 days	04/08/21	429	340	-176	12779
180 days	03/09/21	429	340	-176	14475
365 days	03/07/22	432	343	-176	24929

Table 8. Worldwide confirmed cases using SVR

Days	Date	RMSE	MAE	Predicted cases
90 days	17/09/20	9638199	2226417	44979589
120 days	17/10/20	9638199	2226417	56977842
150 days	17/11/20	9638199	2226417	67979852
180 days	17/11/20	9638199	2226417	88979852
365 days	16/06/21	9638199	2226417	274979589

Table 9.Worldwide confirmed cases using LR

Days	Date	RMSE	MAE	Predicted cases
90 days	17/09/20	10391079	4527384	15,21,11741
120 days	17/10/20	10391079	4527384	16,31,44541
150 days	17/11/20	10391079	4527384	17,62,89401
180 days	17/11/20	10391079	4527384	19,22,34451
365 days	16/06/21	10391079	4527384	24,22,34451

Table 10. Worldwide death dates SVR

Days	Date	RMSE	MAE	Predicted deaths
90 days	17/09/20	114774	106436	2451503
120 days	17/10/20	114386	106037	3444381
150 days	17/11/20	114386	106037	4644381
180 days	17/11/20	114386	106037	5644381
365 days	19/06/21	114386	106037	22739697

Table 11. Worldwide death data usingLR

Days	Date	RMSE	MAE	Predicted cases
90 days	17/09/20	108943	108125	554668
120 days	17/10/20	108943	108125	631694
150	17/11/20	108943	108125	687694

days				
180 days	17/11/20	108943	108125	785747
365 days	19/06/21	108943	108125	1260742

3. CONCLUSION

In this study, for predicting the next numbers of cases that are going to be happening in the upcoming weeks or months, algorithms like Support Vector Regression and Linear Regression are used. This type of prediction is helpful for the governments to take the necessary arrangemnets in hospitals. By using these algorithms we have predicted the no of cases for the next 365 days and these algorithms can also be used for predicting next year's also. From the obtained results Support vector Regression is predicting very well baecuase its R-square value is high compared with other algorithms.

REFERENCES

[1] Qiang Li, Wei Feng: Trend and forecasting of the COVID-19 outbreak in China. Arxiv. 2002.05866, 10.1016/j.jinf.2020.02.014,2020.

[2] Bohdan M. Pavlyshenko: Machine Learning models for time series forecasting. MPDI, 2019, pp. 1-11. doi:10.3390/data401001,2018.

[3] Jagadish Kumar, K.P.S.S. Hembram: Epidemiological study of novel corona virus (COVID-19). arxiv.2003.11376,2020.

[4] Fattah J, Ezzine L, Aman Z, El Moussami H, Lachhab A: Forecasting of demand using ARIMA model. International Journal of Engineering Business Management; 10:1847979018808673, 2018.

[5] Pavlyshenko, B.M. Linear: Machine learning and probabilistic approaches for time series analysis. In Proceedings of the IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 23–27 August 2016; IEEE: Piscataway, NJ, USA, pp.377–381.2016

[6] Robson, Winston A. "The Math of Prophet— Breaking down the Equation behind Face book's open-source Time Series Forecasting procedure", *Medium*, Future Vision, June 17 2019, https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a

[7] Alakhsethy,Support Vector Regression Tutorial for Machine Learning..https://www.analyticsvidhya.com/blog/2020/03/support-vetor-regression-tutorial-for-machine-learning/

[8] FurqanRustam, Aizaz Ahmed Reshi, ArifMehmood, SaleemUllah, Byung-Wonon, Waqar Aslam and Gyu Sang Choi(2020): COVID-19 Future Forecasting Using Supervised Machine Learning Models. Doi: 10.1109/ACCESS.2020.2997311.IEEE Access, Volume 8, pp.101489-101499.2020

[9] M. B. Jamshidi , A. Lalbakhsh J. Talla1 , Z. Peroutka , F. Hadjilooei , P. Lalbakhsh , M. Jamshidi, L. La Spada , M. Mirmozafari , M. Dehghani , A. Sabet, S. Roshani, S. Roshani, N. Bayat-Makou, B. Mohamadzade , Z. Malek , A. Jamshidi, S. Kiani, H. Hashemi-Dezaki , W. Mohyuddin (2020): Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment. Doi:10.1109/ACCESS.2017.

[10] Nanning Zheng, Shaoyi Du, Jianji Wang , He Zhang, Wenting Cui , Zijian Kang, Tao Yang, Bin Lou, Yuting Chi, Hong Long, Mei Ma, Qi Yuan, Shupei Zhang, Dong Zhang, Feng Ye, and Jingmin Xin(2020): Predicting COVID-19 in China Using Hybrid AI Model. IEEE Transactions on Cybernetics. Vol. 50, pp.2891-2904.

[11] E ying(2012): Application of support vector regression algorithm in colleges recruiting student's prediction.ICCSEE,pp:173-176.
DOI 10.1109/ICCSEE.2012.456

[12] Lin Jia, Kewen Li, Yu Jiang, Xin Guo, Ting Zhao: Prediction and analysis of Coronavirus Disease 2019,arxiv.2003.05447,2020.

[13] Gaurav Gupta, NeeruRathee(2015): Performance comparison of Support Vector Regression and Relevance Vector Regression for facial expression recognition, (ICSCTI) **DOI:** 10.1109/ICSCTI.2015.7489548

[14] T.P.Latchoumi, Latha Parthiban, Advanced Overlap Community Detection by Associative Rule Mining and Multi-View Ant Clustering. International Journal of Engineering and Technology,7, pp. 21-29, 2018.

[15] A.Vijay Vasanth, T.P.Latchoumis, B.Bavya, Sk.Sajida. Object Detection in Convolution Neural Networks Using Iterative Refinements, Revue d'IntelligenceArtificielle, IIETA 33(5), pp.367-372, 2019

[16] Ezhilarasi, T.P., Dilip, G., Latchoumi, T.P. and Balamurugan, K., 2020. UIP—A Smart Web Application to Manage Network Environments. In Proceedings of the Third International Conference on Computational Intelligence and Informatics (pp. 97-108). Springer, Singapore

[17] Vijay Vasanth A,Latchoumi T.P, Balamurugan Karnan,Yookesh T.L. Improving the Energy Efficiency in MANET using Learning-based Routing, Revue d'IntelligenceArtificielle, 34(3), pp 337-343, 2020

[18] T. P. Latchoumi, Latha Parthiban, Abnormality detection using weighed particle swarm optimization and smooth support vector machine Biomedical Research (SCIE), 28(11), pp.166-172, 2017.

[19] T.P. Latchoumi, T. P. Ezhilarasi, K. Balamurugan (2019), Bio-inspired Weighed Quantum Particle Swarm Optimization and Smooth Support Vector Machine ensembles for identification of abnormalities in medical data. SN Applied Sciences , 1137, pp.1-12, 2019 DOI: 10.1007/s42452-019-1179-8.

[20] T.P. Latchoumi, K. Balamurugan, K. Dinesh and T. P. Ezhilarasi. Particle swarm optimization approach for waterjet cavitation peening. Measurement , 141, pp.184-189, 2019.

[21] T.P. Latchoumi, Jayakumar, T. P. Ezhilarasi, Latha Parthiban, Mahalakshmi (2018), Comparison of Classification Techniques on Data Mining, International Journal of Pure and Applied Mathematics, 118(19), pp. 221-227, 2018.

[22] Vijay Vasanth, K.Venkatachalapathy, T.P.Latchoumi, Latha Parthiban, T.Sowmiya, V.Ohmprakash (2018), An Efficient Cache Refreshing Policy to Improve QoS in MANET Through RAMP. In Proceedings of the Second International Conference on Computational Intelligence and Informatics (pp. 369-381). Springer Lecture Series,Singapore.

[23] Kaushik Sekaran, R. Rajakumar, K. Dinesh, Y. Rajkumar, T. P. Latchoumi, SeifedineKadry, Sangsoon Lim. An Energy-Efficient Cluster Head Selection in Wireless Sensor Network using Grey Wolf Optimization algorithm, Telecommunication, Computing, Electronics and Control (TELKOMNIKA), 18(6), 2020. http://dx.doi.org/10.12928/telkomnika.v18i6.15199

[24] Balamurugan K, Uthayakumar M, Gowthaman S, Pandurangan R. A study on the compressive residual stress due to waterjet cavitation peening. Engineering Failure Analysis. 2018 Oct 1;92:268-77.

[25] B Battula, V Anusha, N Praveen, G Shankar, TP Latchoumi. Prediction of Vehicle Safety System Using Internet of Things, Journal of Green Engineering, 10(4), pp. 1786 – 1798, 2020.

[26] T. P. Latchoumi, Latha Parthiban, Retriving relevant data in the warehouse on cloud environment through MAS. International Journal of Applied Engineering Research, 10, pp.23-31, 2015.

[27] Latchoumi, T.P. and Parthiban, L. Secure Data Storage in Cloud Environment using MAS. Indian Journal of Science and Technology, 9, p.24-29, 2016

[28] Loganathan, J., Janakiraman, S., Latchoumi, T.P. and Shanthoshini, B. Dynamic Virtual Server For Optimized Web Service Interaction. International Journal of Pure and Applied Mathematics, 117(19), pp.371-377, 2017.

[29] Pruthviraju G, K.Balamurugan, T.P.Latchoumi, Ramakrishna M. A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC hybrid composite using Agglomerative Hierarchical Clustering and K-Means, Silicon, 2020, DOI: 10.1007/s12633-020-00447-9

[30] Garikipati, P. and Balamurugan, K., Abrasive Water Jet Machining Studies on AlSi 7+ 63% SiC Hybrid Composite. In Advances in Industrial Automation and Smart Manufacturing (pp. 743-751). Springer, Singapore.

[31] Ranjeeth, S., Latchoumi, T.P. and Paul, P.V. Role of gender on academic performance based on different parameters: Data from secondary school education. Data in brief, 29, p.105257, 2020.