BBRC
Bioscience Biotechnology
Research Communications

Improved Clustering Using Deep Learning Model on Water Resource Engineering

Sreelatha. P[1], J. Faritha Banu[2], T. Ch. Anil Kumar[3], D. Sugumar[4],

Shailendra Kumar Rawat[5], Ahmad Jawad Niazi[6]

[1]Department of Biomedical Engineering, KPR Institute of
Engineering and Technology, Arasur, Tamil Nadu, India.

[2]Department of Computer Science and Engineering, SRM Institute of
Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.

[3]Department of Mechanical Engineering, Vignan's Foundation
for Science Technology and Research, Vadlamudi, Guntur, Andhra Pradesh, India.

[4]Department of ECE, Signal Processing Lab, Karunya Institute of Technology
and Sciences, Karunya Nagar, Coimbatore, Tamil Nadu, India.

[5]Department of Computer Science, Maharishi Institute of
Information Technology, Lucknow, Uttar Pradesh, India.

[6]Department of Architecture, Kabul Polytechnic
University, Kabul City, Afghanistan. Corresponding author email: sreelathaselvaraj@gmail.com

ABSTRACT

In recent decade, there has been an abundant generation of high-dimensional data in larger volumes while measuring the water resource environment. This environment consists of large spatial areas with high resolution and high temporal data, where such collection needs to be summarized. In this paper, we develop a deep learning model to recognize the meaningful relationship between the objects in data during decision-making process. Appropriate extraction of meaningful relationship using deep learning reduces the redundant information that provides inter-variable relationships and characteristic patterns. The deep learning framework recognizes the patterns via clustering the data that provides better understanding of objects. The experimental results are conducted with various training options and parameter selection on the information extracted. The results of simulation shows that the proposed deep learning model achieves improved clustering than other methods.

KEY WORDS: CLUSTERING, DEEP LEARNING, WATER RESOURCE ENGINEERING, PATTERNS, DATA.

INTRODUCTION

Measurements of water resources, typically gathered at high time resolution in wide spatial regions, provide huge volumes of high-dimensional information that needs to be structured and summarized to extract useful signals for decision-making (Choubin B et. al. (2017)). Appropriate data organization and summary will minimize the information accessible to a few of feature patterns and intervariable relations (Narmatha et. al. (2017)).

Deep Learning (DL) is a sophisticated, artificial neural network for drawing patterns in multivariate large data sets and identifying clusters (Polomčić (2017), Jia (2018), Brentan (2018)). DL has an intuitive implementation, resistance to missing and noisy data, the capacity to integrate data into retail time and a simple visual overview of the system and intervariable interactions, making it very popular in water resources. Hydrologic variables can also

be investigated for physical, chemical, social or economic systems without requiring expert knowledge from each of the different fields, since the DL organizes data without a requirement to explicitly understand the complex underlying systems which could generate the data (Brentan (2018), Feng (2019).

In data collection and clustering the data items around these patterns, the DL extracts the most often available patterns and organises the findings into an intuitive and small-scale visualization or map (Hoover et. al. (2018). A large number of high-dimensional observations are displayed with a much smaller number of low-dimensional vectors that form centers for clusters of the original data (Wu (2021), Yang (2018)). This helps with the examination of the attributes of the data set by analyzing the ordered structure and members of the low-dimensional cluster (Cesario et. al. (2019).

The DL offers different advantages compared to other techniques for environmental data investigation. There is no need to provide the number of clusters contained in the information. The resultant clusters will be sorted in the output view, highlighting similarities and differences, which may be analyzed further if requested. DL creation does not require an explicit comprehension of a system that produces the data; analyses are performed based solely on the data given to the system; processes and relations are complicated and possibly indestructible (Malik (2020), Lin (2020), Xing (2020), Fathi (2018), Fu (2017)).

Thus, the Convolutional Neural Network (CNN) may be used to investigate large multidisciplinary or human-environmental data sets in which it is difficult to specifically define intervariable connections. In this situation, it would be hard and time-consuming to create an extensive model for the mathematical system. The capacity to include fresh data after creating the function map, which provides an online, real-time data analysis tool that is beneficial in the processing of sensor data in real-time, is also a clear advantage of the DL (Leitão (2019), Hu (2019), Yoon (2017), Ropero (2021), Predescu (2018)). The detection, clustering and data visualization of patterns reveals information which may not be observed in vast volumes of data otherwise. For each of these processes, other approaches exist independently, but the combination results in unique analytical capacities for the DLs. For instance, clustering algorithms that are not coupled with reduced dimensions cannot be easily displayed and are hence comprehended and transmitted less quickly.

Although the use of CNNs for environmental applications is widespread and has increased, uncertainty exists in the CNN technique owing to the two inherent competing objectives of the approximate data for the feature map nodes and the preservation of the topology of the dataset. This leads to difficulties in parameters as various options improve and perhaps hinder elements of the outcomes of CNN. Parameter selections affect the creation of CNN outputs, with the usage of separate parameter sets coming from various maps.

The analyst should thus comprehend and understand the various parameter possibilities. A dependence on the default software parameters can lead to maps which are not of the most adequate size and form for a particular collection of data and may not fit the study's unique goals by smoothing and generalizing the level of the feature map output. Interpretation of information based on maps of this kind may disclose less data than might be discovered in connection with the distribution of each variable and the intervariable relationships in the data set. In the light of the proportional advantages of each alternative, analysts must make an educated choice between the choices.

In this paper, CNN model is used to recognize the meaningful relationship between the objects in data during decision-making process. Appropriate extraction of meaningful relationship using deep learning reduces the redundant information that provides inter-variable relationships and characteristic patterns. The deep learning framework recognizes the patterns via clustering the data that provides better understanding of objects.
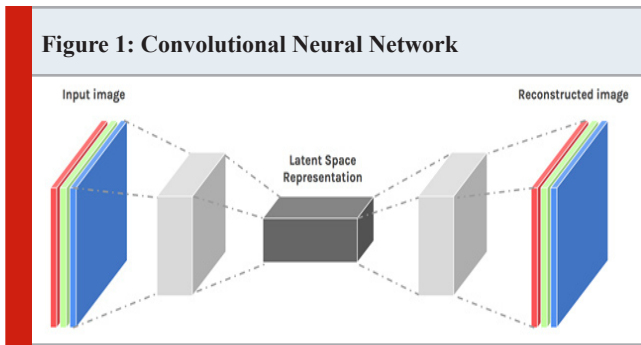
2. Related Works: A system for the prioritization of a real-time clustering system in a water distribution system has been suggested by Predescu, A., et al (Luo et. al. (2021)). For real-time situations that provide decision assistance for water distribution system management, data clustering methods have been updated. Lu, S. et al. (Huang et. al. (2018)) proposes, in the context of bidding projects, to examine the topic of multi-target group decision-making using the mathematical clustering approach. Bidder 3 has finished with the selection of a score matrix and index weight vector of the third bidder involved as the contracting partner most appropriate for building the project. The results show that the above-mentioned technique is more reasonable and trustworthy, and thus the choice to evaluate the bid is more scientific. Huang, M., et al. (Azimi S et. al. (2019)) propose to mimic the nonlinearity of water quality parameters and water quality forecasts, with a forecast model based on a fluctuating water wavelet neural network.

The number of fluffy rules was calculated by means of self-adjusted fluffy clustering. The network parameters were optimized with a hybrid learning method based on the genetic algorithm and the algorithm for gradient descent. Azimi, S. et al. (Villarin M. C. et. al. (2019)) helped evaluate declines in possible drinking water quality in the GIS context with the use of artificial neural networks and improved fuzzy clustering algorithms. The likelihood that the water quality of drinking water will alter adversely each year is assessed in this research. The model has been tested using genuine examples from the South East Aquifers, the Central IRAN areas, and, in particular, the important portions of the East Aquifers.

In order to improve resolution and statistical power, Villarín, M.C. (Luo et. al. (2021)) has investigated water usage at the micro-scale level. Fifteen predictor variables were originally investigated, including sociodemographical factors and buildings. The average cadastral value and the number of inhabitants per home were shown as a preliminary category,

followed by an extensive, multivariate, linear regression analysis. Other critical elements include, for example, building height and residential density. A space study also showed that the use of water in the city center is a gradient, from higher values to lower peripheral values. Luo, D. et al. (Sumathy et. al. (2021)) developed a model of clustering the effects of grey clouds in order to evaluate the level of drought. The degree of gray impact between each index and ideal index is used for determining the index weight and, combined with the subjective mass, provides the full weight; the traditional feature of option is converted into a grey cloud feature using the maximum deviation and maximum entry principle which fully reflects the coexistence of several decisions.

3. Proposed Methodology: In this section, we discuss the modified convolutional neural network for the purpose of clustering the water resource image datasets with the application of preprocessing which is shown in Fig 1.

Figure 1: Convolutional Neural Network

Pre-processing: Naturally, environmental variables are made up of many measuring scales and data kinds. Standardization of the input data matrix columns before map training guarantees that variables with larger magnitudes or variances don't overlap with less dissimilar, smaller-magnitude, or unit-based variables. This equates roughly to each variable's contribution to the outcome. With regard to the preliminary procedure, risks lead to the major axis of the feature map being aligned with the larger variable, resulting in a feature map that represents this variable predominantly.

Common preprocessing processes include input matrix transformations which equal the variances or minimum and maximum of each dimension:

- Normalization of columns by scaling the variances to 1.
- A linear transformation equals each variable's minimum and maximum.

Another technique, while much less frequent, is to independently scale the difference between all variables to represent their perceived relative importance. The logarithmic modifications of highly skewed variable distributions may be useful.

Clustering: The work is aimed at clustering water resource images using CNN, in order to provide optimal compliance with devices by adjusting the weights of the connections. This modification is based mostly on the Hebbian rule,

which permits a greater link between two units when both units function simultaneously. The Hebbian rule is:

$$\Delta w_{ij} = g\left(a_j(t), t_j\right) h\left(o_i(t), w_{ij}\right) \tag{1}$$

Where
w_{ij} is the link weight between units
$\alpha_j(t)$ is the activation function.
t_j is the learning input
o_i - output for the preceding unit
$g(\ldots)$ - unit activation function
$h(\ldots)$ - current link weight function.

Two forms of training exercise occur when weight is updated. Online learning applies changes to the network after each training pattern. Offline or combat learning collects weight changes in all patterns in the workout file and all modifications are applied via the workout pattern file after a complete cycle.

Backpropagation learning: For the arbitrary error function E, the essential concept of the BPNN employs a chain rule in determining the weight of each weight:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial a_i} \frac{\partial a_i}{\partial net_i} \frac{\partial net_i}{\partial w_{ij}} \tag{2}$$

Where
w_{ij} - weight between the layers.
α_i - activation function.
net_i - weighted neurons.

Once the partial derivative of each weight is understood, the error function may be minimized by a simple gradient descendant:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}}(t) \tag{3}$$

Where
η - learning rate.

For the performance and time of the network integration, the user picks the parameter of the learning rate that can be inferred from Equation 6.4. For our experiments, the most frequent parameters are selected. A wide field of research is the investigation of advanced possibilities for neural network learning, which may thus be a focus of future testing. Online training in back propagation is generally much quicker than batch training, especially for large training groups with similar instances. The back propagation and updating workouts are strongly dependent on the correct selection of the parameter β after every model introduction. The rear propagation weight upgrade law, often known as the delta rule:

$$\Delta w_{ij} = \eta \delta_j o_i \tag{4}$$

$$\delta_j = \begin{cases} f'(net_j)(t_j - o_j) & \text{if unit } j \text{ is an output unit} \\ f'(net_j)\sum_k \delta_k w_{jk} & \text{if unit } j \text{ is an hidden unit} \end{cases} \tag{5}$$

Where

δ_j- error of unit j.

o_i- preceding output.

t_j- teaching input j.

i- predecessor index to unit j with w_{ij}

j - unit index.

k - successor index with weight w_{jk}

Learning Algorithm: BPNN is a system of local adaptive learning which offers supervised batch study in MLP. The choice of an educational rate for the back-propagation algorithm in equation 4 with a derivative scale impacts the time to reach convergence. Instead, a high education rate can cause a wobble that prevents an error from falling beneath its value. Too many steps are needed to find a suitable answer, when it is set too small. The assumption is that the detrimental influence of the partial derivative size on weight will be eradicated. Consequently, only the derivative's symbol is the road to weight update. The weight shift scale is uniquely computed.

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)} & \text{if } \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\[2mm] +\Delta_{ij}^{(t)} & \text{if } \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\[2mm] 0 & \text{else} \end{cases} \qquad (6)$$

where

$\dfrac{\partial E^{(t)}}{\partial w_{ij}}$ - Summed gradient information

The fundamental concept of the update conducted using the Rprop method to make it more acceptable was more information on the topology of the error function. During the learning process, each update value is based on the

E-error function. The second phase of RProp $\Delta_{ij}^{(t)}$ is therefore to be assessed on the basis of the following adjustment process:

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ * \Delta_{ij}^{(t-1)} & , \ \text{if } \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} * \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\[2mm] \eta^- * \Delta_{ij}^{(t-1)} & , \ \text{if } \dfrac{\partial E^{(t-1)}}{\partial w_{ij}^{(t-1)}} * \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \qquad 0 < \eta^- < 1 < \eta^+ \\[2mm] \Delta_{ij}^{(t-1)} & , \ \text{else} \end{cases} \qquad (7)$$

The weight update is also largely based on the size and influence of the partial extracts on the entire network. By altering their weight update value directly based on a component derivative sign alone, without respect to its dimensions, Rprop prevents this issue. Rprop employs the 'batch learning' or 'learning by epoch' theory when it attempts

to adapt the mechanism to the topology of the erroneous function. This means that weight updating and adaptation influence the gradient knowledge of the complete design collection. When this is suited to the learning process, it is very easy to select that value. To prevent weights from becoming too large, the maximum phase computed by the update value size is decreased. Convergence in this parameter is generally rather indifferent. However, certain issues only lead to very cautious intervention in order to prevent the algorithm from sticking to local minima.

Objective function: It is vital to know why CNN parameters are particularly hard before you begin to explore how parameters are selected. Objective functions are usually optimized in the parameter selection phase of the model construction to support the decision, given the ability to quantify the optimality of a set of parameters. It has nevertheless been demonstrated that there is no objective function of the CNN training process which is perfectly optimized. In a single mathematical statement, the CNN training technique is impossible to quantify but rather follows the gradient descent of a distinct set of energy functions for each node. Therefore, parameters should be chosen using alternative techniques, as mentioned below, such as quality measurements used on maps generated with various combinations of parameters.

Only by combining all weighted activations $f_{act}(x) = 1/(1+e^{-x})$ and then shutting down the result can this function determine the network input. The new activation time (t+1) is within the range [0.1]. The jth unit threshold is the vector. The CNN input is calculated as follows:

$$net_j(t) = \sum_i w_{ij} o_i(t) \qquad (8)$$

$$a_j(t+1) = \dfrac{1}{1 + e^{-\left(\sum_i w_{ij} o_i(t) - \theta_j\right)}} \qquad (9)$$

4. Performance Evaluation: This section provides the results of clustering between the proposed CNN and conventional deep learning models that includes Recurrent Neural Network (RNN), Multi Layered Perceptron (MLP), Deep Belief Network (DBN) and self-organizing map (SOM). The evaluation is conducted on a 2D dataset namely m Environment Canada's HYDAT database that aims to cluster the river basins. The proposed model is compared with existing methods in terms of various metrics that includes accuracy, variance, cohesion, recall and precision and recall rate.

Cohesion: relates the closeness of the data in the cluster, which are related to each other clusters. The cohesion

between each cluster as:

$$p_i = \frac{\sum_{m \neq n} C_{mn}}{K * \frac{(K-1)}{2}}$$ (10)

Where,
m,n is regarded as the cluster data point,
K is regarded as the cluster data point,
C_{mn} is regarded as the data points distances.

Variance: show the separation of clusters, which is given by

$$V_i = \sum_{m \neq n} \frac{(p_i - C_{mn})^2}{\left(\frac{K(K-1)}{2}\right) - 1}$$ (11)

Where,
p_i - cluster cohesion value,
C_{mn} - data points distance,
K - data points in the clusters.

Precision is defined as a fraction of relevant retrieved datasets or positive predictive value. This is determined in terms of retrieval of true positive (TP) values and false positive (FP), which is given as:

Precision= TP/ (TP + FP)

Recall: is defined as the fraction of TP and the sum of retrieval TP and false positive (FP), which is given as:

Recall=TP/ (TP + FP)

Table 1. Accuracy rate

Clustering Algorithm	Accuracy
Modified CNN	0.871368
CNN	0.84166
RNN	0.76488
MLP	0.5952
DBN	0.69352
SOM	0.6808

Table 1 shows the results of accuracy rate between the proposed and existing methods. The results of simulation shows that the proposed CNN obtains improved rate of accuracy than other methods. Table 2 shows the results of Cohesion rate between the proposed and existing methods. The results of simulation shows that the proposed CNN obtains improved rate of cohesion rate than other methods. Table 3 shows the results of variance between the proposed and existing methods. The results of simulation shows that the proposed CNN obtains improved rate of variance

than other methods. Table 4 shows the results of precision between the proposed and existing methods. The results of simulation shows that the proposed CNN obtains improved rate of precision than other methods. Table 5 shows the results of Recall between the proposed and existing methods. The results of simulation shows that the proposed CNN obtains improved rate of recall than other methods.

Table 2. Cohesion rate

Clustering Algorithm	Cohesion
Modified CNN	0.8345
CNN	0.80924
RNN	0.78618
MLP	0.76878
DBN	0.75258
SOM	0.7403

Table 3. Variance

Clustering Algorithm	Variance
Modified CNN	0.75078
CNN	0.724798
RNN	0.7132
MLP	0.70304
DBN	0.69182
SOM	0.66864

Table 4. Precision

Clustering Algorithm	Precision
Modified CNN	0.77522
CNN	0.75756
RNN	0.7402
MLP	0.73372
DBN	0.72576
SOM	0.71562

Table 5. Recall

Clustering Algorithm	Recall
Modified CNN	0.80686
CNN	0.79276
RNN	0.77976
MLP	0.76868
DBN	0.75332
SOM	0.741928

In order for time changes in data sets to be indicated by the CNN visualizing processes, any one of the following techniques may be used to indicate: A CNN is trained by all the available data input, and changes in the mapping of data over time for each node are shown by time trends in the structure. The best matched map node for self-organizing feature-mapping is accessible based on available data item variables and the value of the missing variable adopted from the node vector to utilize CNNs for infilling missing data. For prediction, comparable value-added methods are employed based on proven intervariable relations. A collection of readily anticipated variables and then the unknown value of the variable adopted on the node vectors is the best matching map node. The integration of the new data into the trained and clustered feature map demonstrates how it is based on relations and clusters with existing data items. It is placed in the clusters and compared with the other members of the clusters.

CONCLUSION

CNNs have large applications and numerous features that make them particularly suitable for data on water resources. When used intelligently, additional data about the linkages of the underlying systems will be revealed. Water resource engineers and scientists want to use CNNs to investigate high-dimensional non-linear data sets because they are a known available resource. This article helps researchers to develop and interpret a CNN appropriate for a given dataset, to provide a guide for understanding the important parameter selections and to interpret the CNN findings for exploratory data analysis. Improving each particular application of the CNN will lead to a better understanding of the intervariables of water resource systems.

REFERENCES

Azimi, S., Moghaddam, M.A. and Monfared, S.H., 2019. Prediction of annual drinking water quality reduction based on Groundwater Resource Index using the artificial neural network and fuzzy clustering. Journal of contaminant hydrology, 220, pp.6-17.

Brentan, B., Meirelles, G., Luvizotto Jr, E. and Izquierdo, J., 2018. Hybrid SOM+ k-Means clustering to improve planning, operation and management in water distribution systems. Environmental modelling & software, 106, pp.77-88.

Cesario, E., Vinci, A. and Zhu, X., 2019, June. Hierarchical Clustering of Spatial Urban Data. In International Conference on Numerical Computations: Theory and Algorithms (pp. 223-231). Springer, Cham.

Choubin, B., Solaimani, K., Roshan, M.H. and Malekian, A., 2017. Watershed classification by remote sensing indices: A fuzzy c-means clustering approach. Journal of Mountain Science, 14(10), pp.2053-2063.

Ezhilarasi, G.D., Latchoumi, T.P. and Balamurugan, K., 2020. UIP—A Smart Web Application to Manage Network Environments, Advances in Intelligent systems and computing book series.

Fathi, E., Zamani-Ahmadmahmoodi, R. and Zare-Bidaki, R., 2018. Water quality evaluation using water quality index and multivariate methods, Beheshtabad River, Iran. Applied Water Science, 8(7), pp.1-6.

Feng, Z.K., Niu, W.J., Zhang, R., Wang, S. and Cheng, C.T., 2019. Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization. Journal of Hydrology, 576, pp.229-238.

Fu, Q., Guo, J., Cui, S., Li, T. and Liu, D., 2017. An economic-efficiency analysis of water conservancy investment in Heilongjiang Province based on projection pursuit clustering and DEA models. Water Science and Technology: Water Supply, 17(6), pp.1579-1588.

Hoover, J.H., Coker, E., Barney, Y., Shuey, C. and Lewis, J., 2018. Spatial clustering of metal and metalloid mixtures in unregulated water sources on the Navajo Nation–Arizona, New Mexico, and Utah, USA. Science of The Total Environment, 633, pp.1667-1678.

Huang, M., Tian, D., Liu, H., Zhang, C., Yi, X., Cai, J., Ruan, J., Zhang, T., Kong, S. and Ying, G., 2018. A hybrid fuzzy wavelet neural network model with self-adapted fuzzy-means clustering and genetic algorithm for water quality prediction in rivers. Complexity, 2018.

Hu, G., Kaur, M., Hewage, K. and Sadiq, R., 2019. Fuzzy clustering analysis of hydraulic fracturing additives for environmental and human health risk mitigation. Clean Technologies and Environmental Policy, 21(1), pp.39-53.

Jia, Z., Cai, Y., Chen, Y. and Zeng, W., 2018. Regionalization of water environmental carrying capacity for supporting the sustainable water resources management and development in China. Resources, Conservation and Recycling, 134, pp.282-293.

Leitão, J., Simões, N., Sá Marques, J.A., Gil, P., Ribeiro, B. and Cardoso, A., 2019. Detecting urban water consumption patterns: a time-series clustering approach. Water Supply, 19(8), pp.2323-2329.

Lin, K., Chen, H., Xu, C.Y., Yan, P., Lan, T., Liu, Z. and Dong, C., 2020. Assessment of flash flood risk based on improved analytic hierarchy process method and integrated maximum likelihood clustering algorithm. Journal of Hydrology, 584, p.124696.

Lu, S., Shang, Y. and Li, Y., 2017. A research on the application of fuzzy iteration clustering in the water conservancy project. Journal of Cleaner Production, 151, pp.356-360.

Luo, D., Hu, Y. and Sun, D., 2021. Drought grade assessment method based on grey cloud incidence clustering model. Grey Systems: Theory and Application.

Malik, S., Gupta, K. and Singh, M., 2020. Resource Management in Fog Computing Using Clustering Techniques: A Systematic Study. Annals of the Romanian Society for Cell Biology, pp.77-92.

Mishra, P., Jimmy, L., Ogunmola, G.A., Phu, T.V., Jayanthiladevi, A. and Latchoumi, T.P., 2020, December.

Hydroponics cultivation using real time iot measurement system. In Journal of Physics: Conference Series (Vol. 1712, No. 1, p. 012040). IOP Publishing.

Narmatha, M., and R. Alageswaran. "Design of IoT based secure water supply application using zone based clustering." Advances in Natural and Applied Sciences 11, no. 4 (2017): 92-100.

Polomčić, Dušan, Zoran Gligorić, Dragoljub Bajić, and Čedomir Cvijović. "A hybrid model for forecasting groundwater levels based on fuzzy C-mean clustering and singular spectrum analysis." Water 9, no. 7 (2017): 541.

Prasath, S. (2020). Probabilistic Mceliece Public-Key Cryptography Based Identity Authentication For Secured Communication In Vanet. Solid State Technology, 63(6), 10167-10182.

Prasath, S. Validating Data Integrity in Steganographed Images using Embedded Checksum Technique. International Journal of Computer Applications, 975, 8887.

Predescu, A., Negru, C., Mocanu, M. and Lupu, C., 2018, May. Real-time clustering for priority evaluation in a water distribution system. In 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR) (pp. 1-6). IEEE.

Shukla, A., Kalnoor, G., Kumar, A., Yuvaraj, N., Manikandan, R. and Ramkumar, M., 2021. Improved recognition rate of different material category using convolutional neural networks. Materials Today: Proceedings.

Sumathy, S., Revathy, M. and Manikandan, R., 2021.

Improving the state of materials in cybersecurity attack detection in 5G wireless systems using machine learning. Materials Today: Proceedings.

Venkata Pavan, M., Karnan, B. and Latchoumi, T.P., 2021. PLA-Cu reinforced composite filament: Preparation and flexural property printed at different machining conditions. Advanced Composite Materials, https://doi.org/10.1080/09243046.2021, 1918608.

Villarín, M.C., 2019. Methodology based on fine spatial scale and preliminary clustering to improve multivariate linear regression analysis of domestic water consumption. Applied Geography, 103, pp.22-39.

Wu, L., Peng, Y., Fan, J., Wang, Y. and Huang, G., 2021. A novel kernel extreme learning machine model coupled with K-means clustering and firefly algorithm for estimating monthly reference evapotranspiration in parallel computation. Agricultural Water Management, 245, p.106624.

Xing, Y., Szabo, J., Magnuson, M. and Harper Jr, W.F., 2020. Clustering, morphology, and treatment resistance of Bacillus globigii spores recovered from a pilot-scale activated sludge system. Chemosphere, 260, p.127591.

Yang, A., Zhang, H., Stewart, R.A. and Nguyen, K.A., 2018. Water end use clustering using hybrid pattern recognition techniques—Artificial Bee Colony, Dynamic Time Warping and K-Medoids clustering. Int. J. Mach. Learn. Comput, 8, pp.483-487.

Yoon, Y., Hastak, M. and Cho, K., 2017. Preference clustering-based mediating group decision-making (PCM-GDM) method for infrastructure asset management. Expert Systems with Applications, 83, pp.206-214.