www.revistageintec.net
ISSN: 2237-0722
revista GEINTEC Gestão, Inovação e Tecnologias

Metro Water Fraudulent Prediction in Houses Using Convolutional Neural Network and Recurrent Neural Network

D. Sreekanth[1]; Dr.K. Thinakaran[2*]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.

[1]dabbarasrikanth@gmail.com

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India.

[2*]thinakarank.sse@saveetha.com

Abstract

Aim: *The main aim of the study is to predict metro water fraud accurately by Recurrent Neural Network Algorithms and compare the prediction accuracy with Convolutional Neural Network.* ***Materials and Methods:*** *In the existing system Convolutional Neural Network algorithm is used and in the proposed system Recurrent Neural Network algorithm is used. CNN with sample size =20 and RNN with sample size =20 was iterated forty times for predicting the accuracy. The algorithms have been implemented and tested over a dataset which consists of 8002 records.* ***Result:*** *After performing the experiment we get mean accuracy of 94.5210 by using Recurrent Neural Network algorithm and we get accuracy of 93.4950 by using Convolutional Neural Network algorithm for metro water fraudulent prediction. There is a statistical significant difference in accuracy for two algorithms is $p<0.05$ by performing independent samples t-tests.* ***Conclusion:*** *The comparison results show that the Recurrent Neural Network algorithm appears to be better performance than Convolutional Neural Network algorithms.*

Key-words: Metro Water, Convolutional Neural Network Algorithm, Recurrent Neural Network Algorithm, Machine Learning, Statistical Analysis, Novel Detection.

1. Introduction

Water may be the crucial element for the uses of households, industry and agriculture. By using machine learning we have to detect the fraud to improve water and sanitation services. It is important because water utilities can increase cost recovery by reducing administrative technical losses(Herath, 2017). The paper describes the data and provides the methods for the availability,

accessibility and sanitation services. It also presents details of six metro cities in India(Lee et al., 2021; Saroj et al., 2020). Here so many applications are used in the Recurrent Neural Network. The application of this research is to detect fraudulent behaviour in water consumption using machine learning (Hu et al., 2018). The Recurrent Neural Network describes it as a well based performer in making and finding accuracy. Using machine learning techniques the RNN has so many networks to display the performance of algorithms (Puleio, 2021).

The leakage defects of metro water in the metro cities and the defects is even more in the metro cities (Zhao & Huang, 2019). Metro water can hold the record of the metro water cities. Water is the most important to our life without water we cannot survive(Shruti et al., 2020). The wlr gives the scenario identification, vulnerability and flooding assessment in the coastal areas. It is based on the flooding risk analysis under possible scenarios (Y. Zhang et al., 2019). The article discussed the fraud in the metro cities for the metro water. Here the records should be clear for the maintenance of metro water records and it is easy to say whether the fraud has taken place or not (Kleijn & de Kleijn, n.d.). The consumption should be less than the daily needs so the water should not be wasted and the metro water fraudulent behaviour finding will predict the accuracy, so the water should not be contaminated(Lin et al., 2021). In predicting water fraud the regression model is used by using the different parameters, currently it predicts the water quality. It can also have different temperature methods (Nafi & Brans, 2018).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Gheena & Ezhilarasan, 2019; Jose et al., 2020; Ke et al., 2019; Krishnaswamy et al., 2020; Malli Sureshbabu et al., 2019; Mehta et al., 2019; Muthukrishnan et al., 2020; M. S. Samuel et al., 2019; S. R. Samuel et al., 2020; Sathish & Karthick, 2020; Sharma et al., 2019; Varghese et al., 2019; Venu, Raju, et al., 2019; Venu, Subramani, et al., 2019; Vignesh et al., 2019; Vijayakumar Jain et al., 2019). Now the growing trend in this area motivated us to pursue this project.

To detect the water fraudulent is important because water utilities can increase cost recovery by reducing administrative technical losses. This paper is intended to implement recent machine learning algorithms for metro water fraudulent prediction using Recurrent Neural Network and comparison of prediction accuracy with Convolutional Neural Network Algorithms.

2. Materials and Methods

The experiment was conducted in Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University. We use two supervised learning algorithms as

two groups, Convolutional neural network and Recurrent neural network. We perform two iterations on each group, one for metro water fraud training and other for metro water fraud testing using these two algorithms. We use a Pre-power test about 80%. The sample size for each group is 20 and the total sample size is 40.

The dataset; which we used in the paper is training dataset and testing dataset. and The dataset contains 8001 rows and 8 columns. The dataset contains two files “dataset.csv” and “test.csv”. The file dataset.csv contains eight attributes named as “Condition”, “Case”, “Diameter”, “Year”, “Month”, “Reference”, “Consumption”, “Label” and test.csv file contains seven attributes named as “Condition”, “Case”, “Diameter”, “Year”, “Month”, “Reference”, “Consumption”. We need only the text attribute as dependent for analysis and classification.

Convolutional Neural Network Algorithm

Convolutional Neural Network algorithm is a supervised learning algorithm .It indicates that the network employs an operation called convolution. It will work by using tree layers, they are input layer, hidden layer, output layer. The principle of this convolutional neural network is to activate the function, pool layer and fully connected layer.

Pseudocode:

1.Input: d:dataset
2.Output: CNN trained model
3.let f be the matrix
4.For i in data do
5. Let f_i be the matrix of sample i
6. For j in i do
7. V ←$vector_{(j,w)}$
8. append v_j to f_i
9. append f_i to f
10. f_{train}, f_{test}, l_{train}, l_{test} ← split the datasets
11. M ← CNN(f_{train},l_{train})
12. Evaluate(i,l_{test},M,)
13. Return score

3. Recurrent Neural Network Algorithm

Recurrent Neural Network is a powerful network and robust type of neural networks. In the RNN the main part is internal memory which remembers the input. This novel technique used to analyse the fraudulent behaviour. It will use the internal memory as arbitrary inputs in the sequence of the inputs.

Pseudocode

1. Create def function for rnn
2. Declare variables rnn_acc, rnn_history
3. x.reshape( ) shape the dataset and remove the null values
4. lstm_model = sequential ( ) do the dropout values
5. Print the accuracy
6. Print the graph.

The platform used to evaluate the machine learning algorithms was jupyter lab. The hardware configurations were ryzen processors with a RAM size of 8 GB. The system type used was a 64-bit, OS, X64 based processor with HDD of 1 TB. The operating system used was windows and the tool used was jupyter lab with python programming language.

The dataset is trained and tested data is collected. Data preprocessing has to be done. After this it will remove the unnecessary attributes from the dataset and concatenating and shuffling also need to be done. Data exploration shows the contents present in the dataset. Convert the dataset that it contains only the data needed for the classifier. Split the dataset into a training set and testing set. Now implement the machine learning classifier and use the training dataset to train the classifier. After training the classifier uses a testing dataset to test the trained classifier to get the predicted accuracy from the classifier.

The analysis was done by IBM SPSS version 21. It is a statistical software tool used for data analysis. For both proposed and existing algorithms iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analysing accuracy. With value obtained from the iterations Independent Sample T-test was performed .

4. Results

From **Table-1**, we can observe the accuracy of the Recurrent Neural Network is approximately (94%) and Convolutional Neural Network is approximately (93%). The accuracy varies for different test sizes in decimals. The mean, standard deviation and significant difference of Recurrent Neural Network and Convolutional Neural Network is tabulated **(Table 2)** which shows there is a significant difference between two groups since $p<0.03$(Independent Samples Test). The mean, standard deviation and standard error mean of Recurrent Neural Network algorithm and Convolutional Neural Network is tabulated **(Table 3)** which shows that Recurrent Neural network has an accuracy mean of 94.52%,Std.Deviation of 0.19950 for the sample size of 20 where the Convolutional Neural Network has an accuracy mean of 93.49%, Std.Deviation of 0.20869 for the sample size of 20, based on the above results the statistical significance of Recurrent Neural Network is high. Bar graph is comparing the mean accuracy of Convolutional Neural Network and Recurrent Neural Network algorithms **(Fig1).**

Table 1 - Comparison of Prediction accuracy between Convolutional Neural Network and Recurrent Neural Network. The RNN attained accuracy 94% when compared to CNN accuracy 93%.

Execution	CNN algorithm	RNN algorithm
1	93.00	94.00
2	93.26	93.96
3	93.06	93.67
4	92.96	94.33
5	92.67	93.87
6	92.45	93.98
7	93.23	94.34
8	93.45	93.68
9	93.11	94.22
10	93.54	94.11

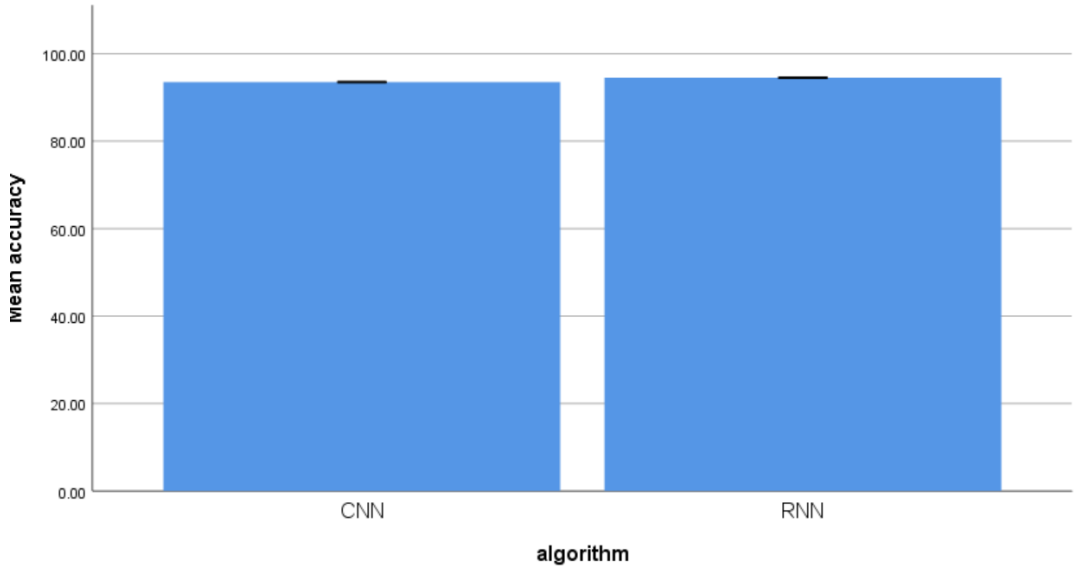
Table 2 - Group Statistics. Here the samples taken for the RNN algorithm is 20 and the mean of RNN algorithm is 94.5210 and CNN mean value is 93.4950. The below table will show the RNN attained Std. Deviation (0.19950) and Std error mean (0.04461).

Algorithm	N	Mean	Std. Deviation	Std. Mean Error
CNN	20	93.4950	0.20869	0.04667
RNN	20	94.5210	0.19950	0.04461

Table 3 - Independent Samples Test. The accuracy increases and the error rate decreases.

	Levene's Test		T-test for Equality of Means						
	F	**Sig.**	**t**	**df**	**Sig**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
Equal Variances assumed	0.104	0.748	-15.893	38	0.002	-1.02600	0.06456	1.15669	0.89531
Equal Variances not assumed			-15.893	37.923	0.002	-1.02600	0.06456	1.15670	0.89530

Fig 1 - Comparison of CNN and RNN algorithms In terms of Mean accuracy. The CNN algorithm has 93% accuracy and the RNN algorithm has 94% accuracy. The mean accuracy of RNN is better than CNN and the standard deviation of RNN is slightly better than CNN. X Axis: CNN vs RNN Algorithm, Y Axis: Mean accuracy of detection ± 1 SD.



5. Discussion

In this study of metro water fraudulent prediction, the Recurrent Neural Network algorithm has higher accuracy approximately (94%) in comparison to convolutional Neural Network algorithm approximately (93%). Recurrent Neural Network algorithms have better significance ($p<0.005$) which we get while using the SPSS tool for statistical calculations. Dong et al. have implemented Resnet50 and Recurrent Neural Network classifiers to detect the metro water fraudulent and obtained accuracy 89% (Dong et al., 2019). Zhang et al. Introduced the ensemble deep learning model and they used ensemble classifiers to detect metro water fraudulent and obtained accuracy 91%(Y.-F. Zhang et al., 2020). Metro water can hold the record of the metro water cities. Water is the most important to our life without water we cannot survive(Shruti et al., 2020). Kim et al. uses the convolutional neural network algorithm in ungauged watershed. Here they developed a discharge estimation model and CNN is best algorithm than RNN (Kim et al., 2020). The metro water should

not be contaminated because of the pollution to prevent water pollution from the various domains in the system mobility, communication, scalability in the data acquisition (Saravanan et al., 2018).

Factors affecting both the classifiers using metro water fraudulent behaviour is cost recovery, dataset size. The identification ability of the model is completely dependent on the RNN and its characteristics; small size datasets with a smaller number of class labels performs better convergence. Our research is aimed to develop the domain with the supervised learning methods (Jalilov, 2017). The data in the Recurrent Neural Network analyses the situation happening where the Convolutional Neural Network improves the classifier which is used in the metro water fraudulent prediction. The Convolutional Neural Network has so many disadvantages than the Recurrent Neural Network ("Gesture Recognition Using CNN and RNN," 2020).

Our institution is passionate about high quality evidence based research and has excelled in various fields ((Ezhilarasan et al., 2019; Mathew et al., 2020; Pc et al., 2018; Ramadurai et al., 2019; Ramesh et al., 2018; Sridharan et al., 2019; Vijayashree Priyadharsini, 2019). We hope this study adds to this rich legacy.

There are a lot of attributes which the dataset contains that are not helpful to predict accuracy , only very few to predict accuracy(%) for metro water fraudulent prediction in houses.The long sequence of data is not handled by activation function.

Many attributes can be considered for RNN in future so that it can work efficiently and improve the prediction accuracy. Attributes like profile, source, proofs can improve the accuracy.

6. Conclusion

The accuracy of metro water fraudulent prediction using Recurrent Neural Network algorithm has better accuracy 94% in comparison with Convolutional Neural Network accuracy of 93%. The discussion proves that the Recurrent Neural Network algorithm appears to be better than the Convolutional Neural Network algorithm.

Declarations

Conflict of Interests: No conflict of interest in this manuscript

Author Contribution

Author Sree is involved in data collection, data analysis, manuscript writing. Author K.Thi involved in data verification, conceptualization, data validation and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,(Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

7. Biozone Pvt.Ltd., Chennai-600032
8. Saveetha University
9. Saveetha Institute of Medical and Technical Sciences
10. Saveetha School of Engineering

References

Dong, Y., Wen, R., Li, Z., Zhang, K., & Zhang, L. (2019). Clu-RNN: A New RNN Based Approach to Diabetic Blood Glucose Prediction. In *2019 IEEE 7th International Conference on Bioinformatics and Computational Biology (ICBCB)*. https://doi.org/10.1109/icbcb.2019.8854670

Ezhilarasan, D., Apoorva, V.S., & Ashok Vardhan, N. (2019). Syzygium cumini extract induced reactive oxygen species-mediated apoptosis in human oral squamous carcinoma cells. *Journal of Oral Pathology & Medicine: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*, *48*(2), 115–121.

Gesture Recognition using CNN and RNN. (2020). In *Regular*, 9(2), 230–233. https://doi.org/10.35940/ijrte.b3417.079220

Gheena, S., & Ezhilarasan, D. (2019). Syringic acid triggers reactive oxygen species-mediated cytotoxicity in HepG2 cells. *Human & Experimental Toxicology*, *38*(6), 694–702.

Herath, S. (2017). Keynote address: Urban Water Management in Metro Colombo. In *2017 Moratuwa Engineering Research Conference (MERCon)*. https://doi.org/10.1109/mercon.2017.7980445

Hu, Y., Huang, S.Y., Hanner, R., Levin, J., & Lu, X. (2018). Study of fish products in Metro Vancouver using DNA barcoding methods reveals fraudulent labeling. In *Food Control* (Vol. 94, pp. 38–47). https://doi.org/10.1016/j.foodcont.2018.06.023

Jalilov, S.M. (2017). Value of Clean Water Resources: Estimating the Water Quality Improvement in Metro Manila, Philippines. In *Resources* 7(1), p. https://doi.org/10.3390/resources7010001

Jose, J., Ajitha, & Subbaiyan, H. (2020). Different treatment modalities followed by dental practitioners for Ellis class 2 fracture – A questionnaire-based survey. *The Open Dentistry Journal*, *14*(1), 59–65.

Ke, Y., Al Aboody, M.S., Alturaiki, W., Alsagaby, S.A., Alfaiz, F.A., Veeraraghavan, V.P., & Mickymaray, S. (2019). Photosynthesized gold nanoparticles from Catharanthus roseus induces caspase-mediated apoptosis in cervical cancer cells (HeLa). *Artificial Cells, Nanomedicine, and Biotechnology*, *47*(1), 1938–1946.

Kim, D.Y., Da Ye, K., & Song, C.M. (2020). Developing a Discharge Estimation Model for Ungauged Watershed Using CNN and Hydrological Image. In *Water,* 12(12), 3534. https://doi.org/10.3390/w12123534

Kleijn, G. de, & de Kleijn, G. (n.d.). Fistulae and water fraud in late antique Constantinople. In *Fountains and Water Culture in Byzantium,* 55–67. https://doi.org/10.1017/cbo9781316226742.004

Krishnaswamy, H., Muthukrishnan, S., Thanikodi, S., Arockiaraj, G., Antony, & Venkatraman, V. (2020). Investigation of air conditioning temperature variation by modifying the structure of passenger car using computational fluid dynamics. *Thermal Science*, *24*(1 Part B), 495–498.

Lee, G., Kim, D., & Han, J.I. (2021). Gas-diffusion-electrode based direct electro-stripping system for gaseous ammonia recovery from livestock wastewater. *Water Research*, *196*, 117012.

Lin, D., Bai, L., Xu, D., Zhang, H., Guo, T., Li, G., & Liang, H. (2021). Effects of oxidation on humic-acid-enhanced gypsum scaling in different nanofiltration phases: Performance, mechanisms and prediction by differential log-transformed absorbance spectroscopy. *Water Research*, *195*, 116989.

Malli Sureshbabu, N., Selvarasu, K., V, J.K., Nandakumar, M., & Selvam, D. (2019). Concentrated Growth Factors as an Ingenious Biomaterial in Regeneration of Bony Defects after Periapical Surgery: A Report of Two Cases. *Case Reports in Dentistry*, *2019*, 7046203.

Mathew, M.G., Samuel, S.R., Soni, A.J., & Roopa, K.B. (2020). Evaluation of adhesion of Streptococcus mutans, plaque accumulation on zirconia and stainless steel crowns, and surrounding gingival inflammation in primary …. *Clinical Oral Investigations*. https://link.springer.com/article/10.1007/s00784-020-03204-9

Mehta, M., Deeksha, Tewari, D., Gupta, G., Awasthi, R., Singh, H., Pandey, P., Chellappan, D.K., Wadhwa, R., Collet, T., Hansbro, P.M., Kumar, S.R., Thangavelu, L., Negi, P., Dua, K., & Satija, S. (2019). Oligonucleotide therapy: An emerging focus area for drug delivery in chronic inflammatory respiratory diseases. *Chemico-Biological Interactions*, *308*, 206–215.

Muthukrishnan, S., Krishnaswamy, H., Thanikodi, S., Sundaresan, D., & Venkatraman, V. (2020). Support vector machine for modelling and simulation of heat exchangers. *Thermal Science*, *24*(1 Part B), 499–503.

Nafi, A., & Brans, J. (2018). Prediction of Water Utility Performance: The Case of the Water Efficiency Rate. In *Water* (Vol. 10, Issue 10, p. 1443). https://doi.org/10.3390/w10101443

Pc, J., Marimuthu, T., & Devadoss, P. (2018). Prevalence and measurement of anterior loop of the mandibular canal using CBCT: A cross sectional study. *Clinical Implant Dentistry and Related Research*. https://europepmc.org/article/med/29624863

Puleio, A. (2021). Recurrent neural network ensemble, a new instrument for the prediction of infectious diseases. *European Physical Journal plus*, *136*(3), 319.

Ramadurai, N., Gurunathan, D., Samuel, A.V., Subramanian, E., & Rodrigues, S.J.L. (2019). Effectiveness of 2% Articaine as an anesthetic agent in children: randomized controlled trial. *Clinical Oral Investigations*, *23*(9), 3543–3550.

Ramesh, A., Varghese, S., Jayakumar, N.D., & Malaiappan, S. (2018). Comparative estimation of sulfiredoxin levels between chronic periodontitis and healthy patients - A case-control study. *Journal of Periodontology*, *89*(10), 1241–1248.

Samuel, M.S., Bhattacharya, J., Raj, S., Santhanam, N., Singh, H., & Pradeep Singh, N.D. (2019). Efficient removal of Chromium(VI) from aqueous solution using chitosan grafted graphene oxide (CS-GO) nanocomposite. *International Journal of Biological Macromolecules*, *121*, 285–292.

Samuel, S.R., Acharya, S., & Rao, J.C. (2020). School Interventions-based Prevention of Early-Childhood Caries among 3-5-year-old children from very low socioeconomic status: Two-year randomized trial. *Journal of Public Health Dentistry*, *80*(1), 51–60.

Saravanan, K., Anusuya, E., Kumar, R., & Son, L.H. (2018). Real-time water quality monitoring using Internet of Things in SCADA. *Environmental Monitoring and Assessment*, *190*(9), 556.

Saroj, S.K., Goli, S., Rana, M.J., & Choudhary, B.K. (2020). Data on water, sanitation, and hygiene in six select metro cities of India. *Data in Brief*, *29*, 105268.

Sathish, T., & Karthick, S. (2020). Wear behaviour analysis on aluminium alloy 7050 with reinforced SiC through taguchi approach. *Journal of Japan Research Institute for Advanced Copper-Base Materials and Technologies*, *9*(3), 3481–3487.

Sharma, P., Mehta, M., Dhanjal, D.S., Kaur, S., Gupta, G., Singh, H., Thangavelu, L., Rajeshkumar, S., Tambuwala, M., Bakshi, H.A., Chellappan, D.K., Dua, K., & Satija, S. (2019). Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. *Chemico-Biological Interactions*, *309*, 108720.

Shruti, V.C., Pérez-Guevara, F., & Kutralam-Muniasamy, G. (2020). Metro station free drinking water fountain- A potential "microplastics hotspot" for human consumption. *Environmental Pollution*, *261*, 114227.

Sridharan, G., Ramani, P., Patankar, S., & Vijayaraghavan, R. (2019). Evaluation of salivary metabolomics in oral leukoplakia and oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*, *48*(4), 299–306.

Varghese, S.S., Ramesh, A., & Veeraiyan, D.N. (2019). Blended Module-Based Teaching in Biostatistics and Research Methodology: A Retrospective Study with Postgraduate Dental Students. *Journal of Dental Education*, *83*(4), 445–450.

Venu, H., Raju, V.D., & Subramani, L. (2019). Combined effect of influence of nano additives, combustion chamber geometry and injection timing in a DI diesel engine fuelled with ternary (diesel-biodiesel-ethanol) blends. *Energy*, *174*, 386–406.

Venu, H., Subramani, L., & Raju, V.D. (2019). Emission reduction in a DI diesel engine using exhaust gas recirculation (EGR) of palm biodiesel blended with TiO2 nano additives. *Renewable Energy*, *140*, 245–263.

Vignesh, R., Sharmin, D., Rekha, C.V., Annamalai, S., & Baghkomeh, P.N. (2019). Management of Complicated Crown-Root Fracture by Extra-Oral Fragment Reattachment and Intentional Reimplantation with 2 Years Review. *Contemporary Clinical Dentistry*, *10*(2), 397–401.

Vijayakumar Jain, S., Muthusekhar, M.R., Baig, M.F., Senthilnathan, P., Loganathan, S., Abdul Wahab, P.U., Madhulakshmi, M., & Vohra, Y. (2019). Evaluation of Three-Dimensional Changes in Pharyngeal Airway Following Isolated Lefort One Osteotomy for the Correction of Vertical Maxillary Excess: A Prospective Study. *Journal of Maxillofacial and Oral Surgery*, *18*(1), 139–146.

Vijayashree Priyadharsini, J. (2019). In silico validation of the non-antibiotic drugs acetaminophen and ibuprofen as antibacterial agents against red complex pathogens. *Journal of Periodontology*, *90*(12), 1441–1448.

Zhang, Y., Ayyub, B.M., Zhang, D., Huang, H., & Saadat, Y. (2019). Impact of Water Level Rise on Urban Infrastructures: Washington, DC, and Shanghai as Case Studies. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *39*(12), 2718–2731.

Zhang, Y.F., Fitch, P., & Thorburn, P.J. (2020). Predicting the Trend of Dissolved Oxygen Based on the kPCA-RNN Model. In *Water* (Vol. 12, Issue 2, p. 585). https://doi.org/10.3390/w12020585

Zhao, S., & Huang, H. (2019). Deep learning-based instance segmentation for water leakage defects of Metro Shield Tunnel. In *Tunnels and Underground Cities: Engineering and Innovation meet Archaeology, Architecture and Art* (pp. 3475–3482). https://doi.org/10.1201/9780429424441-368