Indonesian Journal of Electrical Engineering and Computer Science
Vol. 29, No. 3, March 2023, pp. 1551~1559
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v29.i3.pp1551-1559

Study and innovation of effective classification of XML documents using an advanced deep learning approach

Sahunthala Sanmugam[1], Angelina Geetha[1], Latha Parthiban[2]
[1]Department of Computer Science, Hindustan Institute of Technology and Science, Kelambakkam, India
[2]Department of Computer Science, Pondicherry University, India

Article Info

Article history:

Received Sep 7, 2022
Revised Oct 18, 2022
Accepted Oct 24, 2022

Keywords:

Archimed improved numerical
Optimization deep learning
Classification
Optimization
XML data

ABSTRACT

In the digital world, classifying real sensed data in huge volumes derived from numerical problems is a challenging task due to the computational complexity of the metaheuristic searching process. The deep learning approach includes convolutional neural network (CNN), long short-term memory (LSTM), and Bidirectional (BI)-LSTM, suitable for an optimistic processing time of analyzing XML datasets (i.e., social media, trade center, and surveillance data exchanged in the internet world). However, it faces process deviation when datasets extend their range beyond the expected volume. This paper proposes a novel deep learning formwork referred to as archimed improved numerical optimization deep learning (AINODL) to improve the classification of XML datasets. The proposed AINODL framework first extracts feature from XML documents using the vector space model. Secondly, it classifies the XML data using the inbuilt function of the AINODL framework. The experiments demonstrate that the performance parameters accuracy (90%), sensitivity (93%), and specificity (94%) of the proposed AINODL framework are significantly enhanced compared with the existing approaches CNN, LSTM, and BI-LSTM.

Corresponding Author:

Sahunthala Sanmugam
Department of Computer Science, Hindustan Institute of Technology and Science
Padur, Kelambakkam, Tamil Nadu, India
Email: sahukalai2011@gmail.com

1. INTRODUCTION

Nowadays, the data is analyzed with different aspects to improve business growth. The numerical problems are implemented in massive data to explore the various features of the data set. The XML data plays an essential role in a lot of applications in business intelligence. The machine learning technique is the program that takes the data given by the user and analyzes it automatically based on the model built in the program. The open challenge in the business is generating complete accuracy in the domain of data mining. It extracts the user-required data from colossal data volume and enhances the process by using an innovative technique that benefits the business layers. Most of the web applications are based on the XML model. It follows a hierarchical base structure format for analyzing the given data [1]. In this, the data classification process is more complex than in the other data models. The query is processed among more than one XML data set, forming the integrity operation among the data set [2]. Later, it was done based on the semantics information. The deep learning process is the machine learning technique explicit about performing multiple interconnections with more than one hidden layer (i.e., the input of each hidden layer is the previous layer's output) to achieve the required process. Therefore, it needs optimization to retain the business value by protecting the application layer from external attacks. Such attacks are classified based on the vulnerabilities [3], [4]. The Spatio-temporal XML documents give the spatial information, which is further updated, inserted, and deleted based on fuzzy spatial-temporal fuzzy structure [5]. The XML data-based various web applications are classified in [6], [7]. Subsequently, the semantics of the XML data is analyzed, and disambiguous data is generated sphere neighborhood [8] methods.

Similarly, XML data is mapped in the real-time application with suitable matches [9], [10]. The XML data is classified based on the concept and context based on the information [11]-[13]. The basic structure of deep learning is used in network intrusion detection with the various weight parameters between hidden layers. More than once, it has undergone real-time application [14], [15]. The traditional machine learning approach for classification and clustering is explained in [16], [17]. The efficient classification based on without features is discussed in [18]-[20]. The classification technique kernel principal component analysis-kernel extreme learning machine (KPCA-KELM) technique is proposed to fuzzy model XML data. The accuracy and training time is evaluated and compared with the existing technique Extreme Learning Machine (ELM). The various classification technique is introduced [21]-[26] in XML data. The contribution of this paper is given as shown in:

- To provide easy access to different XML datasets (social media, trade center, and surveillance data) to extract silent features points using vector space model and effectively classify XML datasets into other classes.
- To identify the intensity of feature depth and correctly classify by using the proposed archimed improved numerical optimization deep learning (AINODL) framework. It is continuously operated until sufficient knowledge base information is acquired from the given dataset.
- To improve the performance of the training phase by knowledge gathering silent fractures by effectively calculating the adaptive parameter such as F1score, precision, specificity, sensitivity, FPR, and accuracy.

In the remaining section of the paper, section 2 describes the classification of XML documents from related work. The proposed AINODL framework is discussed in section 3. The result and experimental setup are presented in section 4. Finally, the conclusion of the research is given in section 5.

2. RELATED WORK

In this section, we look at the work of several authors who use deep learning frameworks such as convolutional neural network (CNN) [27], extreme learning machine (ELM), and long short-term memory network (LSTM) to perform a hierarchical classification of XML data. The success of a new method for categorising XML data relies heavily on a well-designed optimization algorithm. This is due to the best features can be gleaned from the acquired XML database. For this reason, the XML data classification issue is tackled by considering the merits and drawbacks of a variety of optimization algorithms discussed in a hierarchical fashion as follows:

2.1. Archimedes optimization algorithm (AOA)

It is also referred to as the population-based technique, which follows Archimedes' principle [11]. It evaluates the statistical significance, convergence ability, exploitation-exploration ratio, and the diversity of analysis of alternatives (AOA) solutions. This technique uses neither the object fully nor partially depending on the classification feature. The object state is measured by (1).

$$G_a = W_k = DVA \quad (1)$$

Where $'G_a'$is the gravity of the object, and $'W_k'$what is the weight of the object. It may also be equivalent to likelihood neighbor's factors such as is the density, $'V'$is the volume, and is the acceleration of the object. This method, can improve the optimization performance on the highly non-linear problem in the volume of data. It supports the specific issue of optimal classification based on whether the object is located upward or downward depending on the current thing of the location. Thereby, each data point reaches the optimal location in the object. As a result, the numeric significance, convergence ability, exploitation-exploration ratio is generated by using the population. However, it failed to explore more fundamental problems in the world. Luan and Lin [12] the performance of the sensitivity parameter is analyzed of the given data. It involved various artificial intelligence techniques to evaluate more than sensitivity parameters that may occur in the actual data in the business application.

2.2. CNN and LSTM

Shone *et al*. [13] addressed the improvement of error backpropagation network through the variety of multi-layer neural networks. The features are extracted by high-level phrase sentences when applied to decoding facial recognition in which analyzing any type of documents, historical data collection, environmental information gathering, grey areas, advertising, and medical diagnosis problems. The non-linear activation function is applied to the operation of the convolution. The whole connection is used for the classification. The data features are extracted by the kernel function, which is specified based on the classification approach. Thereby, it tracks the location point's frame to follow the object's lower and upper movement in the specified location. Consequently, long short-term memory (LSTM) is a type of recurrent neural network (RNN). It follows the principle of long-term dependency of the object and updates the information of the object based on the current location of the cell in the state. This technique follows three stages. i) Forget stage: which decides the information to be deleted based on the previous updating of the object, ii) Second stage: the information to be used in the classification is generated, and iii) Third stage: output stage decides which information is to be updated.

$$fr_t = \sigma\left(W_{fr}[Ai_{t-1}, xr_t] + b_{fr}\right) \tag{2}$$

$$in_t = \sigma(W_{in}[Ai_{t-1}, xr_t] + b_{in}) \tag{3}$$

$$op_t = \sigma\left(W_{op}[Ai_{t-1}, xr_t] + b_{op}\right) \tag{4}$$

Where $fr_t$is the forgotten state of the object at the level t, $\sigma$is the sigmoid function, b_{fr} the bias of the forget state, in_t input of the current state, W_{in}weight of the input state, b_{in}discrimination of the input state, $op_t$the output state of the object at the state t, $W_{op}$is the weight of the object at the output stage t, $Ai_{t-1}$is the hidden value of the object at the previous level t-1, and $xr_t$is the current input of the classification. The performance parameters precision, recall, and F-score are evaluated, and the results are analyzed.

2.3. ELM

It is a technique used in the earliest classification of XML documents [2]. The features are extracted using a vector space model. XML data is classified based on the features, and the given data set evaluates its classification performance. This technique follows a single hidden layer feed-forward neural network. The learning speed is fast in this network. The performance of the classification also significantly increases compared with the traditional neural network. Let there be M random samples $\left(x_i, t_j\right) \in r^{k*l}$. Therefore, the ELM model is represented by:

$$f\left(x\right) = \sum_{i=1}^{H} \beta_i G_i\left(x_j\right) \tag{5}$$

where $'H'$represents the number of hidden layer nodes, $G_i$the activation function, and the weight vector between the ith node and output nodes in the model. The main drawback of this technique is that matrix operations can calculate the output weights without iteratively tuning the weights.

3. PROPOSED AINODL FRAMEWORK

The proposed AINODL framework consists of three essential steps: i) feature extraction of the user data, ii) improved Bi-LSTM deep learning method for classification process, and iii) the performance evaluation module. The overall system design of the proposed AINODL framework is given in Figure 1.

The detailed description of the proposed AINODL framework is given as:

- Step 1: Initialize the position of all objects, which is used to find the optimal location of the thing. The primary objective function variable f_{obj} is mathematically computed by:

$$f_{obj} = (U - L) + b * rand * m \tag{6}$$

where, $'U'$is the upper boundary of the search, $'L'$is the lower boundary of the search agent, rand is the dimensional variable generated between 0 and 1, $'b'$is the bias value, and $'m'$is the number of features.

- Step 2: Compute the density, volume, and acceleration for every iteration and update all object values dynamically. Acceleration is the mathematical model to identify the best object in the set of entities. It is denoted by a:

$$a_j^{m+1} = \frac{a-(a)}{(a)-(a)} + rand * m \tag{7}$$

where $'a'$is accelerating the object to find the best movement of the object j. The value of density and dimension is defined between 0 and 1. In each iteration m the density, and dimension are varied linearly every iteration. Density and volume are computed by:

$$d_j^{m+1} = d_j^m + rand * \left(d_b - d_j^m\right) * b \quad (8)$$

$$v_j^{m+1} = v_j^m + rand * \left(v_b - v_j^m\right) * b \quad (9)$$

In the beginning the collision occurs in the object. After some time the thing tries to reach the stable state. The stable state of the object is defined by the term transfer object (TO).

$$TO = exp\ exp\left(\frac{m-i\,ter}{i\,ter}\right) * rand \quad (10)$$

Where, $'i\,ter'$ is the number of iterations of the proposed algorithm. The density value is decreased based on the value of$'i\,ter'$. Neighbors assist this value in the finding of best search optimal location is given by:

$$d^{m+1} = exp\ exp\left(\frac{m_{i\,ter}-m}{m_{i\,ter}}\right) - \left(\frac{m}{m_{i\,ter}}\right) * m * b \quad (11)$$

this leads to achieving an appropriate balance between exploration and exploitation.

- Step 3: Update the object moving direction. The following manipulation updates the object moving direction.

$$P_{position} = 4 * rand * d * b \quad (12)$$

Where $'d'$is the control variable that is set as 0.4.

- Step 4: Tuning the object using E-LSTM. In the proposed model the deep learning model E-LSTM is tuned to improve the performance of the classification process in the input of XML data. In general, the classification performance is based on parameters that are used in the implementation. The parameters of the E-LSTM are the number of layers (nl), number of hidden neurons of E-LSTM (nn), batch size (bs), epoch size (es), Learning rate (lr). The mentioned parameters are significantly utilized to improve the classification process.
- Step 5: Performance evaluation. The performance of the proposed algorithm is evaluated by accuracy, sensitivity, specificity, precision, FPR, and F1 Score. Accuracy is the percentage of the data which is predicted correctly in the algorithm. Sensitivity generates an actual positive recognition rate in the algorithm. This is calculated by total number of true positives and the total number of false negatives. Sensitivity gives a real negative recognition rate in the algorithm. This is calculated by total number of true negatives and a total number of false positives. Precision is the value required information rate among relevant objects among the retrieved objects. F1 score is the average weight of Precision and sensitivity in the technique which is used.

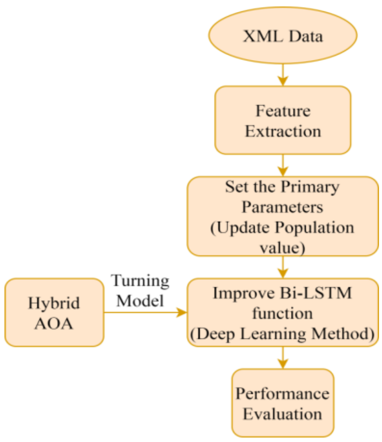

Figure 1. Overall system design of the proposed model

3.1. Pseudo code of proposed AINODL framework

The Algorithm 1 has the input as an XML document. The primary parameters are initialized L (lower bound), U (upper bound), C3, C4 control variable to improve the performance of the classification.

The iteration is executing until to reach the robust performance by updating acceleration (a), density (d) and velacity (v) of the object with optimal value. The manipulation of a, d and v depends on the previous object position in the current execution.

Figure 2 shows the flow diagram of the proposed AINODL framework for XML data classification. It is consists of one root element and more than one child element in the given document. These elements are extracted using a vector space model. It uses the ke Gaussianssian function to remove the details from the hierarchical structure and identify the path of each element from the root element. The number of features is generated based on the derived classification. The required parameters are initialized based on the design of the algorithm such as density (d), velocity (v), acceleration (v), and E-LSTM parameters (number of layers, learning rate). Identify whether the generated parameters are sufficient to improve the performance of the classification. Compute fitness objective function TO. If TO is greater than 1, compute the values of d, a, v using (7), (8) and (9) respectively. Otherwise, compute this using LSTM numerical computation. The primary parameters are updated based on the number of iterations.

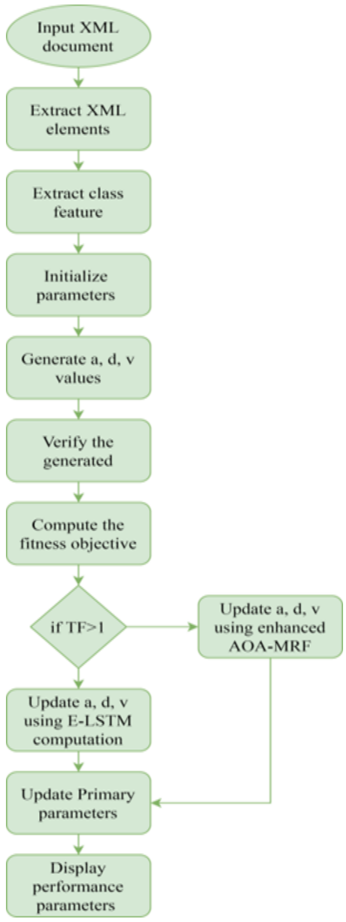

Figure 2. Flow diagram of the proposed AINODL framework

Algorithm 1. AINODL

Input: XML document
Output: Metrics of Performance parameters
Initiation of parameters: $iter, U, L, xr_t, \beta_i, G_i, m, d, D, SA, f_{obj}, dmin$

$L > -1.5$: Lower
$U > 1.5$: Upper
C3->0.1,C4->0.1
Set the class label C
function call AOAMRF($iter, U, L, xr_t, m, D, SA, f_{obj}, C3, C4, X_{test}, Y_{test}$)
C1=0.5,C2=0.5
U=0.3,l=0.5
Compute $a, d\&v$ from (7), (8), and (9), respectively
for t = 1: $iter$

```
TO=exp(((t-iter)/(iter)));
if TO>1
TO=1;
end if
Update a,d&v values
end for
for k = 1: dim
T=TO*C3
if T>1
T=1;
end if
Again, update a,d&v values
```

This article uses the dataset reed.xml to analyze the performance evaluation of the classification technique. The sample XML document element is shown in Figure 3. The root element is root. This root element approach has the children element of course. This course element has more than one sub-element such as reg_num, subj, crse, sect, title, units, instruction etc. This article uses the tttttttttt as features to perform the classification operation. The path of the unit element in Figure 4 is root\course\units.

```
<root>
<course>
    <reg_num>20573</<reg_num>
    <subj>ANTH</subj>
    <crse>344</crse>
    <sect>S01</sect>
    <title>Sex and Gender</title>
    <units>1.0</units>
    <instructor>Makley</instructor>
    <days>T-Th</days>
    <time>
        <start_time>10.30AM</start_time>
        <end_time>11.50AM</end_time>
    </time>
    <place>
        <building>VOLUME</building>
        <room>120</room>
    </place>
</course>
```

Figure 3. Sample XML elements

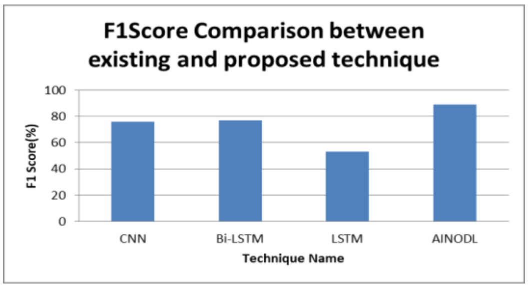

Figure 4. Performance F1 Score comparison between the existing and proposed model

4. EXPERIMENTAL SET-UP AND RESULTS

The experiments are demonstrated in the platform WINDOWS 7/10 with a minimum configuration of 4GB RAM. The proposed AINODL framework is implemented in the MATLAB R2019a version. The dataset reed.xml is used in the performance evaluation. Table 1 gives the overall result generated by the proposed AINODL framework and the existing technique convolutional neural network (CNN), long short-term memory (LSTM), and Bidirectional (BI)-LSTM. The performance parameters computation is referred to from [6]. This article uses the element units to complete the classification of the dataset reed.xml dataset.

Figure 4 shows the performance comparison of the F1 score parameter of the proposed AINODL framework compared with existing techniques CNN, BI-LSTM, LSTM respectively. Usually, a high F1 score is mandatory for any algorithm. In this paper, the proposed AINODL framework generates robust F1score value of 89% is significantly increased as compared with CNN (76%), BI-LSTM (77%), and LSTM (77%). Subsequently, the precision value reached out high by the proposed AINODL framework. Figure 5 shows the performance comparison of the precision value of the proposed AINODL framework compared with existing techniques CNN, BI-LSTM, LSTM respectively. It is observed that the proposed AINODL framework produces a better result than the current techniques. That means nearly 45% improvement has been achieved as compared to LSTM. Similarly, 20% and 10% increment from other two methods BI-LSTM and CNN respectively. On the other hand, the specificity parameter also should maintain as high as possible. The proposed AINODL framework carries it out. That is, it has retained the maximum specificity of (94%) which is reasonably good as compared with CNN (93%), BI-LSTM (89%), and LSTM (77%). It is shown in Figure 6. Consequently, Figure 7 seems that the sensitivity of the proposed AINODL framework has produced a significantly better result than the existing technique. That implies nearly 30% improvement has been achieved as compared to LSTM. Similarly, 15% and 12% increments from the other two methods BI-LSTM and CNN, respectively. In addition to this, for practical processing, FPR value is maintained as low as possible. Based on the result obtained for all four techniques, the proposed AINODL framework produces less FPR rate than the existing technique. Thus, the proposed AINODL framework reduces the FPR value to (5%), which is significantly as low as compared with CNN (6%), BI-LSTM (10%), and LSTM (22%). It is shown in Figure 8. The performance comparison of accuracy between the existing and the proposed technique is shown in Figure 9. The accuracy of the proposed AINODL framework, nearly, 50% improvement has been achieved as compared to LSTM. Similarly, 2% and 12% increments from the other two techniques BI-LSTM and CNN, respectively. It is proven that the proposed AINODL framework significantly yields better results in all the parameters such as F1score, precision, specificity, sensitivity, FPR, and accuracy than the existing techniques CNN, BI-LSTM, LSTM, respectively.

Table 1. Performance parameter results between existing and proposed technique

Technique information		Performance parameters					
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	FPR (%)	F1 score (%)
Existing technique	CNN	89	74	93	81	6	76
	Bi-LSTM	78	77	89	77	10	77
	LSTM	52	62	77	54	22	53
Proposed technique	AINODL	90	93	94	91	5	89

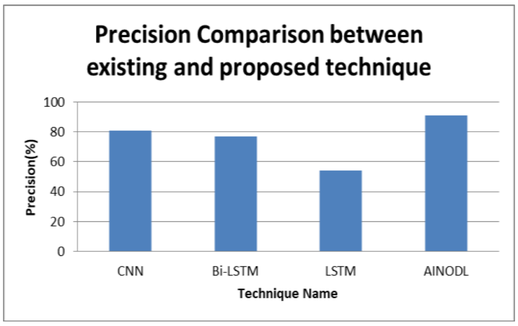

Figure 5. Performance of precision comparison between the existing and proposed model

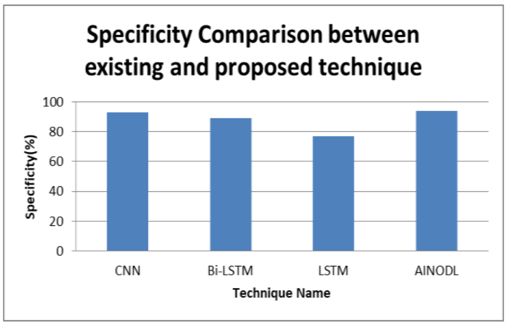

Figure 6. Performance of specificity comparison between the existing and proposed model

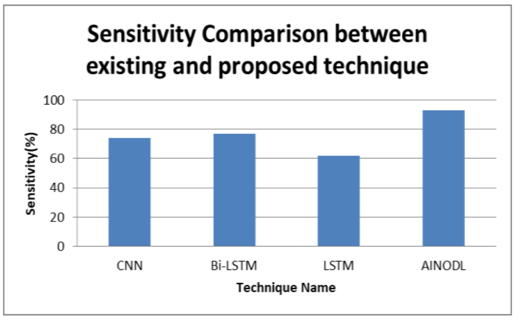

Figure 7. Performance of sensitivity comparison between the existing and proposed model

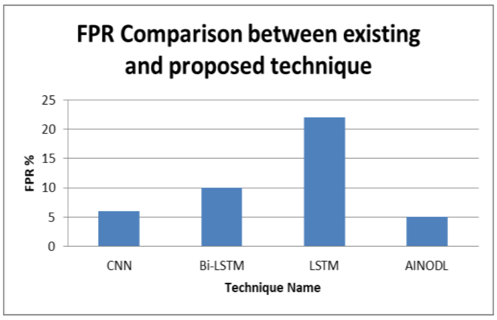

Figure 8. Performance of FPR comparison between the existing and proposed model

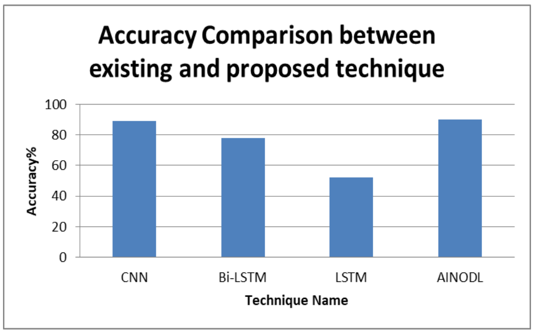

Figure 9. Performance of accuracy comparison between the existing and proposed model

5. CONCLUSION

To obtain a robust performance of the fuzzy XML document, the technique AINODL has been proposed in this study. First, the tree structure of the XML document has been generated using parsing method. Then the feature has been extracted using the Kernel PCA model. The technique AINODL has been able to classify the XML data based on the extracted features of the XML document. The various performance parameters have been evaluated by using the AINODL classification technique. The results of the performance parameters accuracy (90%), sensitivity (93%), and specificity (94%) of the proposed AINODL framework are significantly enhanced compared with the existing approaches CNN, LSTM, and BI-LSTM.

REFERENCES

[1] N. T. Thasleena and S. C. Varghese, “Enhanced associative classification of XML documents supported by semantic concepts,” *Procedia Computer Science*, vol. 46, pp. 194–201, 2015, doi: 10.1016/j.procs.2015.02.011.

[2] X. G. Zhao, G. Wang, X. Bi, P. Gong, and Y. Zhao, “XML document classification based on ELM,” *Neurocomputing*, vol. 74, no. 16, pp. 2444–2451, Sep. 2011, doi: 10.1016/j.neucom.2010.12.038.

[3] C. Gupta, R. K. Singh, and A. K. Mohapatra, “A survey and classification of XML based attacks on web applications,” *Information Security Journal: A Global Perspective*, vol. 29, no. 4, pp. 183–198, Jul. 2020, doi: 10.1080/19393555.2020.1740839.

[4] M. Gopianand and P. Jaganathan, “An effective quality analysis of XML web data using hybrid clustering and classification approach,” *Soft Computing*, vol. 24, no. 3, pp. 2139–2150, Feb. 2020, doi: 10.1007/s00500-019-04045-9.

[5] Z. Ma, L. Bai, Y. Ishikawa, and L. Yan, “Consistencies of fuzzy spatiotemporal data in XML documents,” *Fuzzy Sets and Systems*, vol. 343, pp. 97–125, Jul. 2018, doi: 10.1016/j.fss.2017.03.009.

[6] C. Mata, A. Oliver, A. Lalande, P. Walker, and J. Martí, “On the use of XML in medical imaging web-based applications,” *IRBM*, vol. 38, no. 1, pp. 3–12, Feb. 2017, doi: 10.1016/j.irbm.2016.10.001.

[7] A. Algergawy, E. Schallehn, and G. Saake, “Improving XML schema matching performance using Prüfer sequences,” *Data and Knowledge Engineering*, vol. 68, no. 8, pp. 728–747, Aug. 2009, doi: 10.1016/j.datak.2009.01.001.

[8] A. Gaurav and R. Alhajj, “Incorporating fuzziness in XML and mapping fuzzy relational data into fuzzy XML,” in *Proceedings of the ACM Symposium on Applied Computing*, 2006, vol. 1, pp. 456–460, doi: 10.1145/1141277.1141386.

[9] S. Agreste, P. D. Meo, E. Ferrara, and D. Ursino, “XML matchers: Approaches and challenges,” *Knowledge-Based Systems*, vol. 66, pp. 190–209, Aug. 2014, doi: 10.1016/j.knosys.2014.04.044.

[10] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk, and W. Al-Atabany, “Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems,” *Applied Intelligence*, vol. 51, no. 3, pp. 1531–1551, Mar. 2021, doi: 10.1007/s10489-020-01893-z.

[11] W. Zhao, Z. Zhang, and L. Wang, “Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications,” *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103300, Jan. 2020, doi: 10.1016/j.engappai.2019.103300.

[12] Y. Luan and S. Lin, “Research on text classification based on CNN and LSTM,” in *Proceedings of 2019 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2019*, Mar. 2019, pp. 352–355, doi: 10.1109/ICAICA.2019.8873454.

[13] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, Feb. 2018, doi: 10.1109/TETCI.2017.2772792.

[14] K. Zhang, “The design of regional medical cloud computing information platform based on deep learning,” *International Journal of System Assurance Engineering and Management*, vol. 12, no. 4, pp. 757–764, Aug. 2021, doi: 10.1007/s13198-021-01075-1.

[15] S. Markkandan, A. Sharma, S. P. Singh, V. Solanki, S. Sethuramalingam, and S. P. Singh, “SVM-based compliance discrepancies detection using remote sensing for organic farms,” *Arabian Journal of Geosciences*, vol. 14, no. 14, p. 1334, Jul. 2021, doi: 10.1007/s12517-021-07700-4.

[16] S. Sahunthala, A. Geetha, and L. Parthiban, “Computational fuzzy inference logic for effectively analyzing customer survey,” in *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020*, Oct. 2020, pp. 491–497, doi: 10.1109/I-SMAC49090.2020.9243516.

[17] M. Vidhyalakshmi and S. Sudha, “Text detection in natural images with hybrid stroke feature transform and high performance deep Convnet computing,” *Concurrency and Computation: Practice and Experience*, vol. 33, no. 3, 2021, doi: 10.1002/cpe.5271.

[18] M. Thangamani, S. K. Bharathi, and N. S. Kumar, "Exploring nano biotechnology for detecting specific disease in medical diagnosis and therapeutic," *IOP Conference Series: Materials Science and Engineering*, vol. 1091, no. 1, p. 012031, Feb. 2021, doi: 10.1088/1757-899X/1091/1/012031.
[19] T. T. Leonid and R. Jayaparvathy, "Statistical–model based voice activity identification for human-elephant conflict mitigation," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 5269–5275, May 2021, doi: 10.1007/s12652-020-02005-y.
[20] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014, doi: 10.1109/TCYB.2014.2307349.
[21] S. Sahunthala, A. Geetha, and L. Parthiban, "Analysing computational complexity for prediction function in health record dataset," in *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*, Nov. 2020, pp. 1643–1649, doi: 10.1109/ICECA49313.2020.9297598.
[22] Z. Zhao, Z. Ma, and L. Yan, "An Efficient Classification of fuzzy XML documents based on kernel ELM," *Information Systems Frontiers*, vol. 23, no. 3, pp. 515–530, Jun. 2021, doi: 10.1007/s10796-019-09973-3.
[23] L. Bai, A. He, M. Liu, L. Zhu, and Y. Xing, "Adaptive query relaxation and result categorization of fuzzy spatiotemporal data based on XML," *Expert Systems with Applications*, vol. 168, p. 114222, Apr. 2021, doi: 10.1016/j.eswa.2020.114222.
[24] Y. Liu and Z. Hong, "Mapping XML to RDF: An algorithm based on element classification and aggregation," *Journal of Physics: Conference Series*, vol. 1848, no. 1, p. 012012, Apr. 2021, doi: 10.1088/1742-6596/1848/1/012012.
[25] M. Ledmi, A. Ledmi, and M. El Habib Souidi, "Classification of XML Documents Using Semantic Resources," in *Proceedings - 2021 IEEE International Conference on Recent Advances in Mathematics and Informatics, ICRAMI 2021*, Sep. 2021, pp. 1–5, doi: 10.1109/ICRAMI52622.2021.9585995.
[26] S. Sahunthala, A. Geetha, and L. Parthiban, "An effective classification technique for XML documents using hyper parameter optimized classifiers," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 6, pp. 499–509, 2021.
[27] G. B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, Oct. 2007, doi: 10.1016/j.neucom.2007.02.009.

BIOGRAPHIES OF AUTHORS

Sahunthala Sanmugam she is Assistant Professor at Anand Institute of Higher Technology, Research Scholar, Hindustan Institute of Technology and Science, India, Having teaching experience in S. Veerasamy Chettiar College of Engineering and Technology, National College of Engineering and Anand Institute of Higher Technology. Good exposure on handling various domains of subjects for both UG and PG. She can be contacted at email: sahukalai2011@gmail.com.

Angelina Geetha she is a Professor, Hindustan Institute of Technology and Science. Total Experience (in years) is 25, Research area: data science, web analytics, natural language processing, and knowledge management. She can be contacted at email: angelinag@hindustanuniv.ac.in.

Latha Parthiban she is a Assistant Professor, Department of Computer Science Pondicherry University CC, Lawspet-605008, She has teaching for 26 years (UG/PG) in engineering colleges in Chennai including 8 year at Pondicherry University community college. She has published 109 peer reviewed Scopus indexed/SCI journals and guided 18 Scholars towards completion of PhD. She got best paper award in IEEE conference IAMA 2009 and NCASG 2013. Editor in chief of International Journal of Human Machine Interface of Artificial Intelligence Association of India. She is a member of 5 professional societies like IEEE, CSI, IACSIT, IAE, ISTE. Has reviewed sponsored project proposal for Czech science foundation, one of the leading scientific organizations in the world. Reviewed papers for many international journals including IEEE, Elsevier, Springer link. Completed 3 research projects sponsored by CSI, NVDIA and MHRD. She has completed a PDF and has filed for a patent and presented over 60 papers in conferences and has given over 50 invited talks. She can be contacted at email: athaparthiban@yahoo.com.