Research Article

Semantic Based Data Fusion Technique For Internet Of Things

[1]D. Balakrishnan , [2]T. Dhiliphan Rajkumar, [3]S.Dhanasekaran, [4] B.S.Murugan

[1]Department of Computer Science and Engineering,
Kalasalingam Academy of Research and Education,
AnandNagar, Krishnankoil ,India.
d.balakrishnancse@gmail.com

[2]Department of Computer Science and Engineering,
Kalasalingam Academy of Research and Education,
AnandNagar, Krishnankoil ,India.
t.dhiliphan@klu.ac.in

[3]Department of Computer Science and Engineering,
Kalasalingam Academy of Research and Education,
AnandNagar, Krishnankoil ,India.
srividhans@gmail.com

[4]Department of Computer Science and Engineering,
Kalasalingam Academy of Research and Education,
AnandNagar, Krishnankoil ,India.
b.s.murugan@klu.ac.in

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

ABSTRACT: Handling large volume of data seems to have attracted by number of IoT researchers due to its increased number and volume of data generation. Growing size of data is more complex to handle which might consists of both relevant and irrelevant data even though they are collected from the same kind of IoT devices. This large volume of data handling can be improvised by integrating them together instead of keeping and maintaining multiple copies of same data. This can be performed by introducing the data fusion technique which allows IoT researchers to combine the data's from multiple sources together. There are various research techniques has been introduced earlier for performing the data fusion each follows different procedures. In the existing work, Distributed Hierarchical Data Fusion Architecture (DHDFA) is introduced for the data fusion process where data's collected from the multiple devices will be fused together in multiple levels. However accuracy of this existing research technique is lesser where meaning of data is not considered. This issue is rectified in the proposed research work by introducing the method called Semantic based Hierarchical Data Fusion Technique (SHDFT). In the proposed research work, data fusion is performed in hierarchical manner where data fusion is performed in three levels namely low level, middle level and high level. Here accuracy of the data fusion is improvised by performing the data fusion in the higher level of data fusion process by considering the semantic meaning of the data. Finally performance of the data fusion outcome is tested and analysed by introducing the Convolutional neural network based prediction framework which will learn and analyse the data fusion outcome in terms of error rate. Based on this outcome, data fusion performance can be analysed accurately. The overall evaluation of the research work is done in the matlab in terms of accuracy, error rate and memory consumption against the existing research technique to prove the proposed method effectiveness.

Keywords: Hierarchical Data fusion, Semantic data, IoT devices, distributed data, multiple level of data fusion, unlabelled data.

I. INTRODUCTION

Internet of Things (IoT) has attracted numerous researchers from every industries and organization due to its vast characteristics [1]. IoT enables researchers to create an world with fully digital connectivity with the help of internet [2]. The main advantage of introducing the IoT into the world is, it reduces the need of human access for controlling activities. Thus IoT leads to the anytime, anyplace and any user connectivity at the required time. This is achieved by implementing the IoT devices on every place where it is required to perform any small things using internet connectivity [3]. This characteristic of the IoT is defined by the European university as IoT enables users to access any kind of services any place and any time with the help of any network connectivity [4]. This characteristic of IoT allows various researchers to utilize IoT devices for the purpose.

This deployment of IoT can be performed easily by adapting the sensor technology [5]. A sensor is defined as the smaller device which enables users to monitor and measure the environmental factors such as temperature, humidity and so on. A sensor device can be placed in the environment where it is required to monitor physical measurements. It is required to place more than one sensor devices on an environment to monitor it, which makes it referred as sensor node [6]. Each sensor node has having the ability of monitoring and predicting the environment information which can be communicated to the server side with the help of internet connectivity either through wired or wireless medium [7]. One major advantage of utilizing the sensor network is, it is not required to sensor nodes to be homogeneous. Any kind sensor devices can be placed on the environment and that can be connected through internet connectivity.

The above mentioned IoT vision can be achieved by utilizing the sensor technology as IoT devices. This sensor device enables to IoT researchers to achieve this vision through any kind of sensor device [8]. IoT enables researchers to utilize the mobile sensor devices to achieve this goal. Thus the researcher's goal of constructing smart city can be achieved by integrating the sensor devices in the field which is more cost sensitive and can cover large volume of environment space [9].

While integrating the heterogeneous devices on the environment, it will lead to generate more volume of data which will be more complex to handle. To manage this data, it is required to combine the data's collected from the multiple sensor devices based on their similarity [10]. Data fusion is the approach which allows IoT researchers to combine these data's that are gathered from the multiple sensor devices together, thus the combined data can be generated [11]. Data fusion is the process of combining the data's that are gathered from the multiple devices together based on their similarity, in order to avoid the repeated data handling [12].

In the above explanation, redundant is defined as the factor where the same data is gathered from the multiple sensor devices separately and keeping the same. In one way, this redundant data attempts to increase the accuracy of the sensor device interpretation and in another way it will consume the more memory space. This problem needs to be rectified in the IoT field to enable researchers to make efficient decision making outcome.

The main goal of this research work is to integrate the data's together that are collected from the multiple IoT devices in order to avoid the repeated data's, thus the memory overhead can be reduced considerably. This can be achieved by introducing the data fusion technique which can enable researchers to combine the similar data's together with the aim of reducing the memory storage and processing overhead.

II. RELATED WORKS

King et al [13] expected to give a short outline of data fusion methods and calculations that can be utilized to decipher wearable sensor information with regards to wellbeing observing applications. The utilization of these methods is then portrayed with regards to medical services including action and walking observing, step investigation, fall recognition, and biometric checking. A depiction of momentum industrially accessible sensors is likewise given, zeroing in on their detecting capacity, and an analysis on the holes that should be spanned to put up exploration for sale to the public.

Qi et al [14] proceeds to distinguish some new examination patterns and difficulties of information combination methods in the IoT empowered PARM contemplates, and talks about some key empowering strategies for handling them. Actual work Recognition and Measure (PARM) has been broadly perceived as a critical worldview for an assortment of shrewd medical care applications. Conventional strategies for PARM depends on planning and using Data combination or AI procedures in handling surrounding and wearable detecting information for characterizing kinds of active work and eliminating their vulnerabilities.

Berry et al [15] introduced point oriented decay coefficients (POD) for the data fusion task. This method fixes matrix number and side step number as 1.6 and 1.2 correspondingly to attain the accurate data fusion outcome. The accurate and efficient outcome of data fusion is guaranteed by adapting the symmetrical velocitrical information.

Pfeffer et al [16] introduced singular value decomposition based data fusion approach to enhance the data management task. This is achieved by adapting the various calculation factors such as closeness measure and mixed quality constraints.

Liu et al [17] assessed the general structure to build up a computerized twin combined with the mechanical Internet of Things innovation to propel aviation stages self-sufficiency. Information combination methods especially assume a huge job in the advanced twin structure. The progression of data from crude information to elevated level dynamic is impelled by sensor-to-sensor, sensor-to-display, and show to-demonstrate combination.

Traini et al [18] introduced the methodology that attempts to perform the data fusion in the double step way. The main goal of this research work is to attain the accurate data fusion outcome by adapting the data coming from the long transmission and low transmission ranges. This method ensures the accurate prediction of data fusion outcome by considering the context features.

Martín-Morató et al [19] introduced a fundamental examination of a few data fusion procedures pointed toward improving the acknowledgment exactness of an AED framework by exploiting the variety gave by numerous

mouthpieces in antagonistic acoustic conditions. The outcomes affirm that, under suitable handling plans, the acknowledgment rate can be expanded just as the relating autonomy on occasion area.
Singh et al [20] proposed an instrument named "Artificial Intelligent Energy Aware Routing Protocol (AIEARP)" for improving the energy utilization in WSN through the combination of artificial neural network (ANN) and Kohonen self-organizing management (KSOM) strategies. The clusters are framed and re-situated after cycle for viable circulation of energy and decrease of energy exhaustion at singular hubs.
Thupakula et al [21] discussed an data fusion model of multisensor data for object recognizing confirmation in an air terminal atmosphere to allow the traffic noticing system to choose the shape, type and position of a thing. As a future degree of this paper, the thing shape, type and position data from the article conspicuous confirmation stage is given as commitment to the accompanying stage noticeable all around terminal traffic checking system to follow the article advancements for crash assumption.

III. CONTEXT AWARE DATA FUSION

In the proposed research work, data fusion is performed in hierarchical manner where data fusion is performed in three levels namely low level, middle level and high level. Here accuracy of the data fusion is improvised by performing the data fusion in the higher level of data fusion process by considering the semantic meaning of the data. Finally performance of the data fusion outcome is tested and analysed by introducing the Convolutional neural network based prediction framework which will learn and analyse the data fusion outcome in terms of error rate. Based on this outcome, data fusion performance can be analysed accurately. The overall processing flow of the research work is given in the following figure 1.

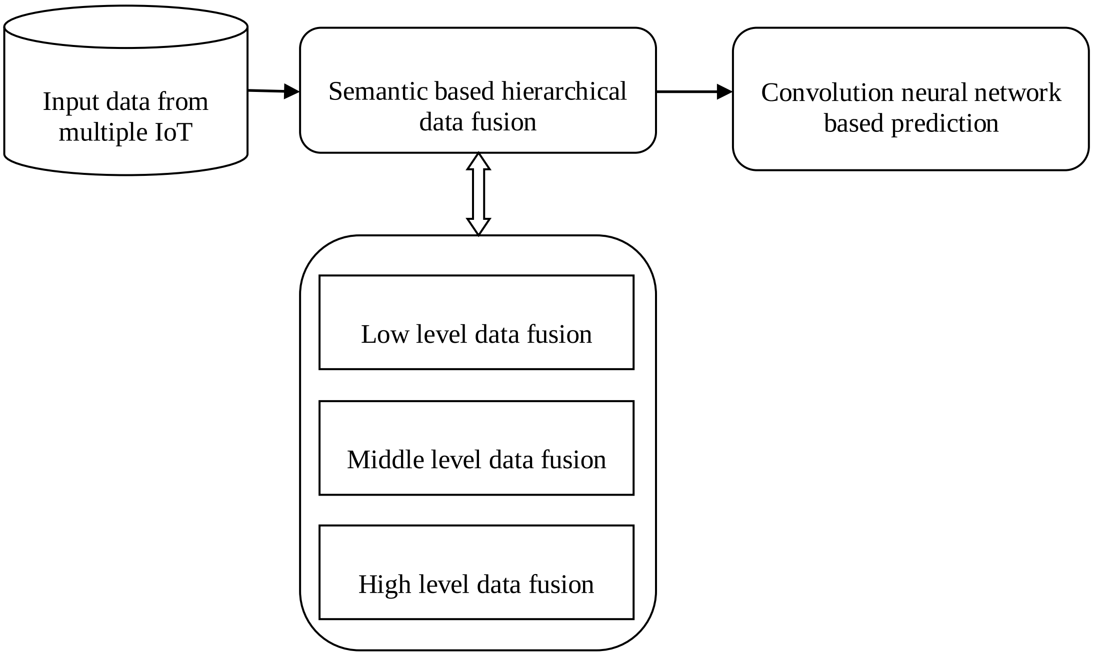

Figure 1. Processing flow of Semantic based Data Fusion

The detailed analysis of the proposed research method is given in the following sub sections.

3.1. Semantic Based Hierarchical Data Fusion

The main contribution of this research work is to perform the data fusion on the data's that are gathered from the multiple IoT devices in order to reduce the multiple copies into single copy. In this research work, semantic based hierarchical data fusion is performed to combine the data's that are collected from the multiple IoT devices. In order to improvise the fusion accuracy, in this work semantic meaning of the data's are considered for the data fusion. And also, efficient and accurate data fusion is guaranteed by performing the data fusion in the three levels namely low level data fusion, middle level data fusion and high level data fusion. In this work semantic meaning of data's are considered at the high level data fusion processing step. The detailed explanation of the semantic based hierarchical data fusion is given in following sub sections:

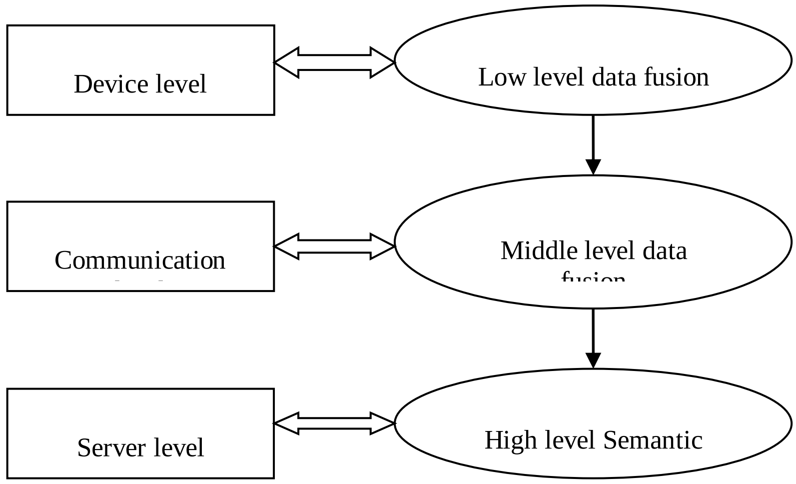

Figure 2. Data fusion levels

3.1.1. Low Level Data Fusion

In general, data's will be collected from the multiple IoT devices that are fixed on the different locations. Those data's will be forwarded to the server for decision through any communication medium. In this research work, data fusion is performed in the three level where in the first level which is called as low level data fusion, data fusion is performed at the IoT devices before transmitting it to the server. By doing so, emergency communication and most important parameters can be processed at first without any delay. For example consider the hospital scenario where patient health status are monitored using the IoT devices which is fused in the hospital itself before transmitting it to the server for handling the most critical patient information first. In this phase, most personal information of patients will be handled first before transmitting it to the server, thus the privacy consideration can be handled. Here simply, similar data information will be gathered and average will be taken for the data fusion process.

3.1.2. Middle Level Data Fusion

In the middle level data fusion process, fusion is performed at the communication level. Once the critical and important information are fused in the device range, it will be transmitted to the server through some communication medium. Middle level data fusion task performs fusion in the communication level to find out more demanding data's and the common factors. In case of hospital scenario, this middle level fusion process enables researchers to find out information about the more commonly spreading disease information and more common factors find out among the patients who are located in same physical location. In this research work middle level data fusion is performed by adapting the clustering technique. In this work existing cluster namely k means algorithm is utilized for the middle level data fusion. The main goal of this algorithm is to find out the most similar data's and grouping them together. Once the similar cluster of data's are formed, here data fusion on each cluster will be carried out by applying the average function.

3.1.3. Semantic based High level Data Fusion

Once the middle level data fusion is performed, the data's will be forwarded to the server, where these data's will be processed and the decision making about certain task can be carried out. In this work, semantic meaning of the data's are considered for the data fusion process, in order to improve the accuracy of the data fusion outcome. In this work, Hownet is considered for the semantic similarity computation between different words, thus the more relevant and similar data's can be combined together with lesser error rate. Hownet is the platform which is utilized to derive the terms and relationship between different concepts and the attributes to reveal the basic content meaning. The concept similarity between the two data's can be calculated by using the following formulae:

$$\mathrm{Sim}\,(w_1, w_2) = \max_{i=1,..,m,\, j=1,..,n} \mathrm{Sim}\,(S_{1i}, S_{2j})$$

Where w_1, w_2 → English words
S → concepts1

By using the concept similarity measured above, semantic similarity can be computed as like below:

$$\mathrm{Sim}\,(p_1, p_2) = \frac{\alpha}{d + \alpha}$$

Where p_1 and p_2 → mean values
α→ adjustable parameter
d → positive integer

By using the above formulae, semantic similarity between the data items can be calculated based on which final prediction can be made.

3.2. Convolution Neural Network Based Prediction

In the proposed work, convolutional neural network is used for the information mining and execution analysis of data fusion. In this work there are three layers are applied to guarantee computation overhead decreased expectation. Those are input layer, convolution layer, lastly soft max layer. CNN is generally made out of two sections. To a limited extent 1, convolution activity is utilized to produce profound features of the raw information. What's more, to a limited extent 2, the features are associated with a MLP for characterization. Here are a few subtleties for each layer:

(i) Input layer: In the proposed Convolutional neural network, numbers of input layers are assigned based on variate number and length of each data. The number of layers assigned in this network is defined as Nxk.

(ii) Convolutional layer: The main goal of this layer is to perform convolution on each received input layer to obtain some masked input. Here convolution operation is performed with the help of filtering parameters which will convolute the input data based on previous iteration outcome. This is obtained with the help of adapting the non linear transformation function f. The format of convolution operation is defined as follows: $\left\lfloor \frac{N-l}{s} + 1 \right\rfloor$ where $\lfloor . \rfloor$ is function which will round off the obtained outcome value.

(iii) Softmax layer: The role of softmax layer is to find the probability of distribution over the input data over the varying distributed events. The main goal of this function is to evaluate the probability of occurring in the next event to attain the target goal. Here the probability range is calculated within the range of 0 to 1. The data with higher probability value will be forwarded to the next even by taking exponential sum of all positive outcomes. Various multi class problems tends to adapt this softmax layer to attain the accurate predictive outcome.

(iv) Output layer: After performing the above mentioned operations, final outcome will be obtained through n number of neurons where n denotes the number of class of features. Based on this outcome, class label will be assigned where the highest number of neurons will take over the most predicted class label.

The overall goal of this research work is to test the performance of data fusion outcome by predicting the fused data class labels accurately. The training samples given as input to the input layer of convolution network is defined as ((a_1, b_1), (a_2, b_2),...,(a_n, b_n)) where $a_t \in R^{N \times k}$, $b_t \in R^n$ for $1 \le t \le N$. Here a_t defines the higher order input and b_t denotes the target output. The processing steps of Convolutional neural network is given in the following steps.

Step 1: Initialize the Convolutional network with initial parameter fixing. The number of input layers and the output layers will be defined at the time network initialization to improvise the classification outcome. The representation of input layer initialization is defined as the sigmoid function as like follows:

$$f(x) = sigmoid(x) = \frac{1}{1+e^{-x}}$$

Step 2: Once the network is initialized, pick the training data in the random way

Step 3: Find the output layer outcome as like follows:

$$C_r(t) = f\left(\sum_{i=1}^{l}\sum_{j=1}^{k} a(i + s(t-1), j)\omega_r(i,j) + b(r)\right)$$

Where s→ convolution stride

$C_r(t)$ → t^{th} component of r^{th} feature

ω_r→ weight value

b(r) → bias value

The derived output layer from the input layer after performing several intermediate actions are given as follows:

$$O(j) = f\left(\sum_{i=1}^{M} z(i)\omega_f(i,j) + b_f(j)\right), j = 1,2,\ldots,n$$

Where z→ final feature map

b_f → Output layer bias value

w_f → weight value of output layer

Based on the outcome from the CNN, performance of data fusion outcome is tested by adapting the mean square error which is calculated as like as follows:

$$E = \frac{1}{2}\sum_{k=1}^{n} e(k)^2 = \frac{1}{2}\sum_{k=1}^{n}(O(k) - y(k))^2$$

Step 4: In every iteration, weight and bias will be updated based on outcome layer and the mean square error by using gradient descent method as like follows:

$$p = p - \eta \frac{\partial E}{\partial p}$$

Step 5: Repeat the iteration with another set of input training samples from step 3
Step 6: repeat until convergence obtained

IV. RESULTS AND DISCUSSION

In this section, numerical evaluation of the proposed research methodology is done in terms of various performance measures to analyze the performance improvement of the proposed and existing research methodologies. The matlab simulation environment is used to implement the proposed research methodology. The performance measures considered in this work are listed as follows: “Accuracy, Precision, Recall and F-Measure”. The performance metrics values are given in the following table 1.

Table 1. Performance metric values

Metrics	**Methods**	
	DHDFA	**SHDFT**
Accuracy	65	87
Precision	74	98
Recall	79	98.6
F-Measure	87	99

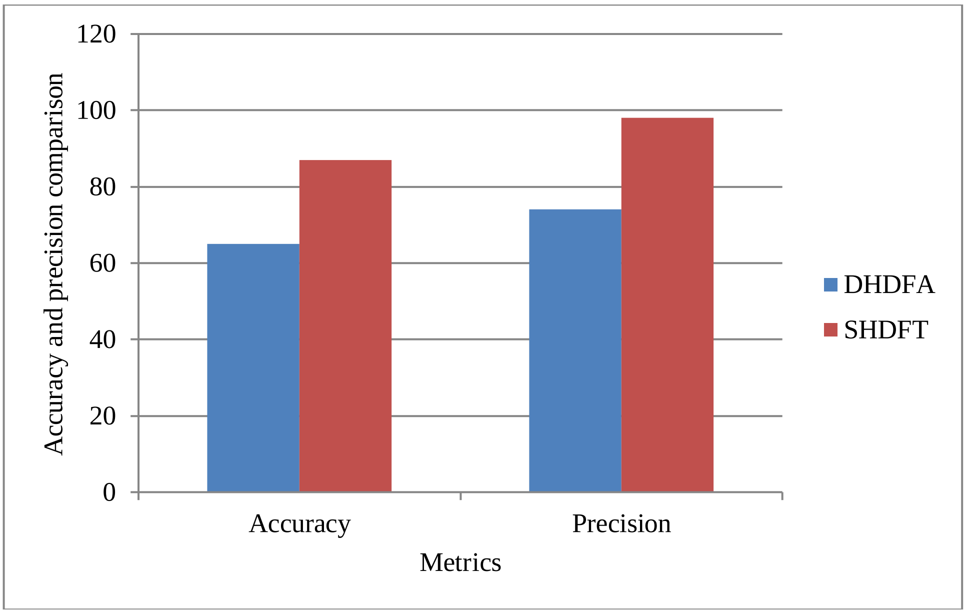

Figure 3. Accuracy and precision comparison

In figure 3, comparison analysis of the proposed method and the existing method namely DHDFA is given. From this analysis it is proved that the proposed shows better performance than the existing technique. Proposed SHDFT 27% increased accuracy than DHDFA. In terms of precision, 24% increased precision than the existing DHDFA.

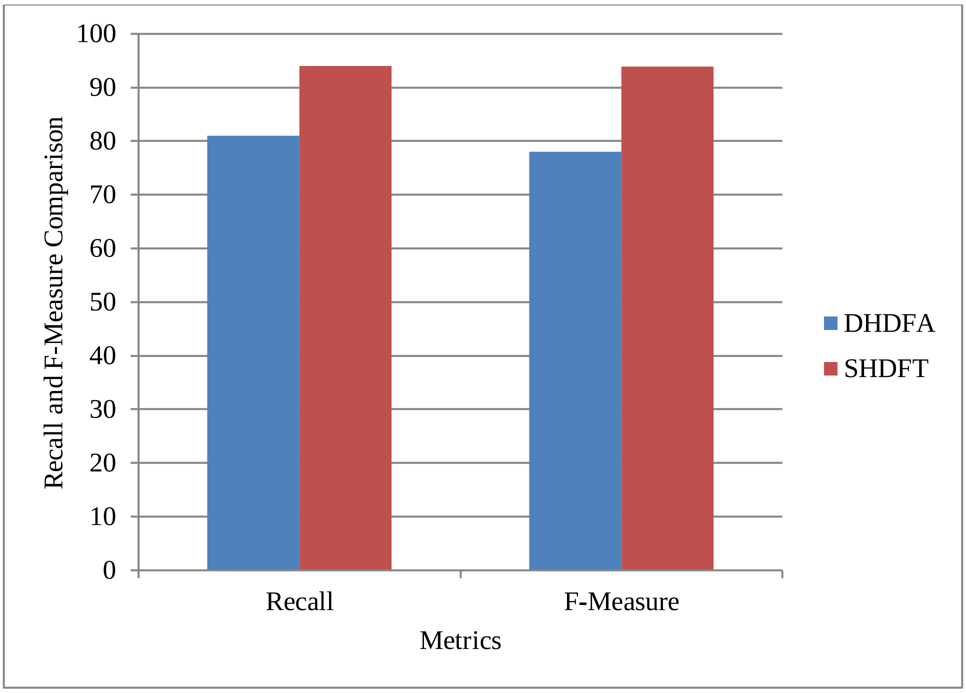

Figure 4. Recall and F-Measure comparison

In figure 4, comparison analysis of the proposed method and the existing method namely DHDFA is given. From this analysis it is proved that the proposed shows better performance than the existing technique. Proposed SHDFT shows 13% increased recall than DHDFA. In terms of F-Measure 19% increased F-Measure than DHDFA.

V. CONCLUSION

In the proposed research work, data fusion is performed in hierarchical manner where data fusion is performed in three levels namely low level, middle level and high level. Here accuracy of the data fusion is improvised by performing the data fusion in the higher level of data fusion process by considering the semantic meaning of the data. Finally performance of the data fusion outcome is tested and analysed by introducing the Convolutional neural network based prediction framework which will learn and analyse the data fusion outcome in terms of error rate. Based on this outcome, data fusion performance can be analysed accurately. The overall evaluation of the research work is done in the matlab in terms of accuracy, error rate and memory consumption against the existing research technique to prove the proposed method effectiveness.

REFERENCES

1. Ray, P. P. (2018). A survey on Internet of Things architectures. Journal of King Saud University-Computer and Information Sciences, 30(3), 291-319.
2. Li, S., Da Xu, L., & Zhao, S. (2018). 5G Internet of Things: A survey. Journal of Industrial Information Integration, 10, 1-9.
3. Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities.
4. Lampropoulos, G., Siakas, K., & Anastasiadis, T. (2019). Internet of things in the context of industry 4.0: An overview. International Journal of Entrepreneurial Knowledge, 7(1), 4-19.
5. Wang, J., Yeh, W. C., Xiong, N. N., Wang, J., He, X., & Huang, C. L. (2019). Building an improved Internet of things smart sensor network based on a three-phase methodology. IEEE Access, 7, 141728-141737.
6. Yahya, O., Alrikabi, H., & Aljazaery, I. (2020). Reducing the data rate in internet of things applications by using wireless sensor network.
7. Kiani, F., & Seyyedabbasi, A. (2018). Wireless sensor network and internet of things in precision agriculture. International Journal of Advanced Computer Science and Applications (ijacsa), 9(6).
8. Manrique, J. A., Rueda-Rueda, J. S., & Portocarrero, J. M. (2016, December). Contrasting internet of things and wireless sensor network from a conceptual overview. In 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (pp. 252-257). IEEE.

9. Pramanik, P. K. D., Upadhyaya, B. K., Pal, S., & Pal, T. (2019). Internet of things, smart sensors, and pervasive systems: Enabling connected and pervasive healthcare. In Healthcare data analytics and management (pp. 1-58). Academic Press.
10. Luis Bustamante, A., Patricio, M. A., & Molina, J. M. (2019). Thinger. io: An Open Source Platform for Deploying Data Fusion Applications in IoT Environments. Sensors, 19(5), 1044.
11. Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., & Jayne, J. T. (2017). Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. Atmospheric Measurement Techniques, 10(9), 3575.
12. Goffi, A., Bordogna, G., Stroppiana, D., Boschetti, M., & Brivio, P. A. (2020). A Scalable Synthesis of Multiple Models of Geo Big Data Interpretation. Journal of Software Engineering and Applications, 13(6), 104-128.
13. King, R. C., Villeneuve, E., White, R. J., Sherratt, R. S., Holderbaum, W., & Harwin, W. S. (2017). Application of data fusion techniques and technologies for wearable health monitoring. Medical engineering & physics, 42, 1-12.
14. Qi, J., Yang, P., Newcombe, L., Peng, X., Yang, Y., & Zhao, Z. (2020). An overview of data fusion techniques for Internet of Things enabled physical activity recognition and measure. Information Fusion, 55, 269-280.
15. Berry, M. G., Stack, C. M., Magstadt, A. S., Ali, M. Y., Gaitonde, D. V., & Glauser, M. N. (2017). Low-dimensional and data fusion techniques applied to a supersonic multistream single expansion ramp nozzle. Physical Review Fluids, 2(10), 100504.
16. Pfeffer, M., Uschmajew, A., Amaro, A., & Pfeffer, U. (2019). Data fusion techniques for the integration of multi-domain genomic data from uveal melanoma. Cancers, 11(10), 1434.
17. Liu, Z., Meyendorf, N., & Mrad, N. (2018, April). The role of data fusion in predictive maintenance using digital twin. In AIP Conference Proceedings (Vol. 1949, No. 1, p. 020023). AIP Publishing LLC.
18. Traini, S., Sciullo, L., Trotta, A., & Di Felice, M. (2019, January). Practical Indoor Localization via Smartphone Sensor Data Fusion Techniques: A Performance Study. In 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC) (pp. 1-7). IEEE.
19. Martín-Morató, I., Cobos, M., & Ferri, F. J. (2017, October). Analysis of data fusion techniques for multi-microphone audio event detection in adverse environments. In 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.
20. Singh, P., Pareek, V., & Ahlawat, A. K. (2017). Designing an energy efficient network using integration of KSOM, ANN and data fusion techniques. International Journal of Communication Networks and Information Security, 9(3), 466-473.
21. Thupakula, K. (2017). Data Fusion Techniques for Object Identification in Airport Environment (No. 2017-01-2109). SAE Technical Paper.