How to Cite:
Sujatha, M. P., & Kumar, S. N. (2022). Anomaly detection of industrial control systems based on transfer learning. *International Journal of Health Sciences*, *6*(S4), 5782–5792. https://doi.org/10.53730/ijhs.v6nS4.9413

Anomaly detection of industrial control systems based on transfer learning

Mrs. Sujatha. M. P
M.P, Assistant Professor, Department of Information Technology, Dhanalakshmi Srinivasan College of Engineering and Technology

MR. S. Niresh Kumar
Assistant Professor, Dhanalakshmi Srinivasan College of Engineering and Technology

***Abstract*---**Industrial Control Systems (ICSs) are the lifeline of a country. Therefore, the anomaly detection of ICS traffic is an important endeavor. This paper proposes a model based on a deep residual Convolution Neural Network (CNN) to prevent gradient explosion or gradient disappearance and guarantee accuracy. The developed methodology addresses two limitations: most traditional machine learning methods can only detect known network attacks and deep learning algorithms require a long time to train. The utilization of transfer learning under the modification of the existing residual CNN structure guarantees the detection of unknown attacks. One-dimensional ICS flow data are converted into two-dimensional grayscale images to take full advantage of the features of CNN. Results show that the proposed method achieves a high score and solves the time problem associated with deep learning model training. The model can give reliable predictions for unknown or differently distributed abnormal data through short-term training. Thus, the proposed model ensures the safety of ICSs and verifies the feasibility of transfer learning for ICS anomaly detection.

***Keywords*---**anomaly detection, transfer learning, deep learning, industrial control system (ICS).

Introduction

Modern Industrial Control Systems (ICSs) have higher production efficiency than traditional industrial systems and can well process big data. However, increases in the type and frequency of network attacks and hacking incidents threaten the security of ICSs based on data transmission. The National Institute of Standards and Technology has proposed the main sources of security issues for modern

ICSs, which include non secure communication protocols, poor network isolation and access controls, and the lack of an ICS anomaly detection system.

Intrusion detection technology is an important research direction in the field of network security. The original flows of network equipment and servers have been comprehensively analyzed. When industrial control networks are invaded or traffic data are abnormal, intrusion detection technology can effectively predict and take active defensive measures in a timely manner. Deep learning has shown great research significance in intrusion detection technology. Feature values are extracted through a great amount of data training, parameters are constantly changed, and a system that can identify abnormal traffic data is constructed. Deep learning and traditional machine learning show certain similarities. The core aim of traditional machine learning is to map features to the target space. In traditional machine learning algorithms, the recognition rate increases with increasing data size; however, because a bottleneck period is often encountered during processing, these models cannot handle massive amounts of data. Machine learning performs well in intrusion detection in closed environments. However, machine learning will be exposed when entering an open-world scenario with various random traffic or noise, which could adversely affect its availability.

Therefore, traditional machine learning algorithms are unsuitable for detecting abnormal traffic in ICSs, and finding abnormal data quickly and implementing active measures with high accuracy are quite challenging. Compared with traditional machine learning, deep learning has a strong generalizability for extracting high dimensional data. Deep learning uses back-propagation algorithms to change and adjust parameters continuously to achieve optimal results. This learning method can handle large amounts of data; indeed, the larger the data size, the better the resulting effect. Unfortunately, although deep learning has good generalizability in processing images, it relies on labeled data and cannot handle unknown abnormal data types. In this article, we solve some of the problems of traditional machine learning by using a residual Convolution Neural Network (CNN) structure to model the source dataset and modify the relevant parameters by transfer learning. We then apply the transfer learning algorithm using the relevant information of the source domain and predicting the target domain. Transfer learning is finally employed to train the model quickly and detect differently distributed or unknown datasets. ICS flow data can usually be processed with one dimensional data sequences through preprocessing; in this work, however, we use mapping to convert ICS flow data to an image format suitable for CNNs to take full advantage of the features of the latter. Fine-tuning is utilized during transfer learning to ensure timeliness.

After building an eight-layer residual neural network, only the three deepest layers of the neural network are fine-tuned. The residual structure, which effectively prevents gradient explosion or gradient disappearance while ensuring the depth of the model effect, is introduced. The detection ability of the model in unknown domain datasets is excellent. After training the source domain, that is the KDDCUP99 dataset, the model is used on a gas pipeline dataset through transfer learning. The model shows good anomaly detection effects on the gas pipeline dataset, and its precision, recall, and F1-score are fairly high.

This article is organized as follows. Section 2 introduces the background of this study and the related work. Section 3 describes the method. Section 4 introduces the evaluation index. Section 5 describes the experimental process and results. Section 6 provides the conclusions and directions for future work.

Background and Related Work

Given rapid developments in informatization and industrialization, ICS has been widely used in national infrastructures. However, platform hardware and software vulnerabilities and the openness of the network environment render ICSs vulnerable to security attacks. Therefore, ICS anomaly detection is very important. Anomaly detection has a significant effect on the active defense process of ICSs. The anomaly detection approaches of ICSs mainly include three types, namely, knowledge-based, statistics-based, and machine learning-based. Reference proposed an anomaly detection method based on state recognition to detect attacks in ICSs by using a data-driven clustering method to identify the normal and critical states of a system. A statistical model for traffic detection in the time domain has also been introduced to detect network anomalies and evaluate the performance of the method in different scenarios. Results show that the model can detect network anomalies in all scenarios faster than other methods. Considering their consequences, network attacks aimed at ICSs are very serious. More importantly, they are difficult to detect. In the context of industrial control environments, anomaly detection based on machine learning does well in improving the accuracy of finding abnormal behavior and is of great importance in the establishment of efficient and intelligent intrusion detection models.

Decision tree, random forest, Support Vector Machine (SVM), and logistic regression are traditional machine learning methods. An earlier study used an intelligent Markov model based on statistical learning to establish a multimodel intrusion detection system for industrial process automation that could effectively detect actual attack operations. Other researchers used SVM to establish a data detection model utilized in an industrial control communication protocol . Decision tree is a machine learning method with a tree structure and high efficiency. It is easy to understand and highly effective for processing discrete data. SVM is based on the principle of structural risk minimization. In this method, the optimal classification is found by learning the classification model of data samples in the feature space.

Traditional machine learning methods have a number of disadvantages, such as low efficiency in processing large-scale data and inability to solve samples with uneven distributions. Compared with traditional machine learning methods, deep learning models have more complex architectures and multiple layers. The most important advantages of deep learning over traditional machine learning are that it can learn features directly and automatically from the original data and has good performance . Deep learning models are quite effective in the field of detecting industrial process anomalies. Almalawi et al. proposed two novel techniques that are an automatic identification of inconsistent states of SCADA data and an automatic extraction of proximity detection rules from identified states. Gao developed an anomaly-based intrusion detection system for the SCADA network and found a combined Intrusion Detection System (IDS) which

includes signature-based IDS and anomaly-based IDS. These studies show that deep learning methods have good performance in anomaly detection and attack classification. CNNs are suitable for image classification . Compared with other image recognition algorithms, CNN uses not only deep learning methods but also some special structures for feedforward neural networks and has relatively little data to preprocess. Transfer learning is an effective approach to exploit deep neural networks on small datasets. The essence of transfer learning is to transfer and reuse knowledge in other fields. Mathematically, transfer learning includes two concepts, namely, domain and learning task. Modelbased transfer learning methods are usually combined with deep learning models to transfer the structure and parameters of models that have been trained on largescale datasets (e.g., AlexNet, VGGNet, and ResNet) to new tasks and use the weights trained on the large dataset as the initial weights for the new task[18]. In contrast to deep learning, transfer learning can detect unknown information. Fine-tuning pretrained CNNs on images is an effective strategy to achieve transfer learning. This technique is widely used in the field of image recognition and provides new insights into the recognition of small scale datasets. In transfer learning, the deep network structure is trained on a large natural image dataset, after which the model is transferred to a small dataset by fine-tuning its parameters. The features extracted from the pretrained deep neural network are universal and applicable to other datasets. Figure 1 shows a schematic of the model.

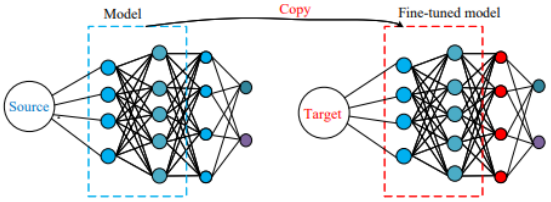


Fig.2. 1 Transfer learning by model fine-tuning.

He, a researcher at Microsoft Research Asia, designed residual neural networks with a deeper network structure and a simpler network structure[19]. The residual network consists of multiple residual blocks, and each residual block comprises a convolutional layer and a pooling layer. The blocks of the convolutional layer are skipped by using shortcut connections. The use of identity shortcuts requires the same input and output sizes[20]. In this case, the problem of model attenuation caused by the disappearance of gradients is avoided by the superposition of gradients. This deep residual neural network won five championships in two major technical competitions, namely, ImageNet and MS COCO. A unique feature of this network is that it includes a network depth greater than 152 layers, which had never been achieved before. In previous experiments, the gradient disappeared as the number of network layers increased and the error rate of such a network was higher than that of a neural network with a lower number of layers. The emergence of the residual neural network solves these issues well. Experiments showed that higher accuracy could be obtained as the number of network layers trained by the residual network increased, which means the residual network can allow deeper network layer training, and the performance of the model was greatly enhanced. Because traditional machine learning algorithms have poor generalizability, deep learning algorithms present greater time costs and may be prone to gradient

disappearance or explosion. Although finetuning technology based on deep neural networks can solve the problem of high time cost and detect unknown attacks, model pretraining remains subject to gradient disappearance or explosion during deep learning. Finetuning based on a deep residual neural network can solve these problems simultaneously.

Material and Method

In this section, we will describe the processing flow in detail. Because of the correlations among the features of industrial control flow data, we convert the onedimensional data stream into a two-dimensional matrix and then convert this matrix into a Mahalanobis Distance (MD) matrix. The obtained matrix is converted into one-dimensional data, saved, normalized, and then mapped into a black-and-white image. After building an eight-layer CNN with a residual structure, we use the KDDCUP99 dataset for pretraining and then train the obtained model on the gas pipeline dataset by fine-tuning through the transfer learning method.

This experiment uses a deep migration learning model and the KDDCUP99 dataset as the source domain to perform migration learning on the gas pipeline dataset. Because the data may be disturbed by noise, missing values, and inconsistent data, the presence of low quality data is inevitable. We improve the credibility of the data by preprocessing them and then improve the performance of model recognition. A preliminary exploration of the data reveals the presence of attributes with exactly the same feature values in the training data; these attributes are not beneficial to the establishment of the model and affect its construction. Therefore, the redundant features are removed. Some normalization methods are used to process the training data, improve the convergence speed of deep learning, and complete the task of anomaly detection.

The steps are as follows: (1) Data cleaning: A large amount of abnormal data will greatly affect the normalized results by affecting the data distribution. Data cleaning is used to clean duplicates, erroneous data, and useless features, thus improving the reliability and integrity of the data as well as the accuracy of the analysis results. The gas pipeline dataset includes some negative and unreasonably large values of key measurement data, which means cleaning is necessary. Next, the attributes “commandlength”, “commwritefun”, “reset”, “gain”, “deadband”, “cycletime”, “rate”, and “crcrate” are redundant attributes in the dataset. We delete these attributes from the dataset because they interfere with data classification. (2) Feature mapping: This experiment uses MD to perform feature mapping on the data. MD was proposed by Mahalanobis as a distance measurement method and refers to the covariance distance of the data. In contrast to the Euclidean distance, MD ignores differences in measurement units and considers the relationship between features, thus aligning the relationship between features with the actual situation. Therefore, MD is not affected by the measurement scale and the interference of correlations between variables can be eliminated. Figure 2 shows the pseudo code for feature mapping of data using MD.

We standardize the source domain, that is the KDDCUP99 dataset. Standardized data are subtracted from the mean and then divided by the variance (or standard deviation). When this data standardization method is completed, the data are converted into a standard normal distribution.

Main code of feature mapping method based on MD

Input: industrial control network data stream
Output: transformed feature matrix
Do while X_i:
Diag = convert_to_diag(X_i)
# The function is responsible for converting the data stream into a diagonal matrix, and the mapped matrix is Map_Matrix=map_matrix(diag)
# The function is responsible for converting the diagonal matrix into a matrix Save_matrix(Map_Matrix)
# The function saves the transformed matrix
$X_i = X_{i+1}$
End while

Fig. 3.1 Pseudo code for feature mapping of data using the MD.

In general, the standard deviation is 1 and the mean is 0. One-dimensional industrial flow data are transformed into a two-dimensional matrix via the feature-mapping method. This section introduces the feature matrix visualization method employed in this article. In this paper, every element in the Map Matrix matrix is regarded as a pixel, and the element value corresponds to the gray value of the pixel.

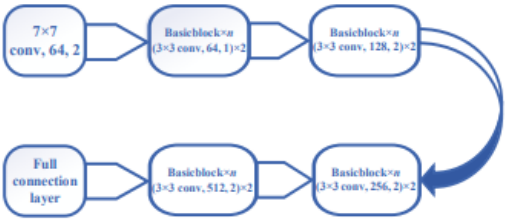


Fig. 3.2 Original residual convolutional neural network model structure

Three residual blocks are utilized in the model. Each residual module is composed of two weight layers and two Rely activation functions. The weight layer is composed of a convolutional layer and a batch normalization layer. The batch-normalization layer transforms the input value distribution of any neuron in each layer of neural network into a standard normal distribution via a certain normalization method. Therefore, the batch normalization layer prevents the model from gradient vanishing and greatly accelerates its training speed. The Relu function performs a nonlinear transformation on the input. The input is not a linear combination of the outputs of the previous layer but can be approximated to any function, thus ensuring the significance of the deep neural network.

Fine-tuning is performed according to the neural network. As the network deepens, the extracted features become more abstract. For two similar domains, the previous layer for extracting common features can be retained after source domain training, and the target domain only needs to train the deepest several layers of the network. In this study, because the KDDCUP99 and gas pipeline datasets are anomaly detection datasets with fixed-dimensional data features and

certain correlations between features, we can use transfer learning on the basis of the data features of the datasets described above. The KDDCUP99 dataset has a sufficient sample

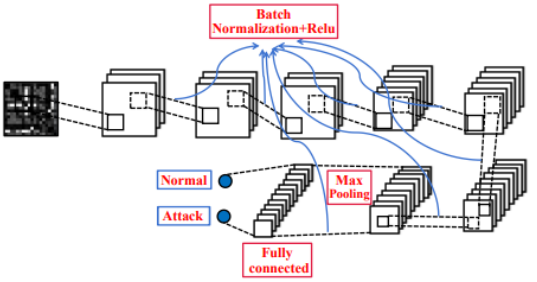


Fig. 3.3 Model structure used in this article.

number and contains over 490 000 datasets, while the gas pipeline dataset is relatively small and contains only over 90 000 datasets. Employing the premise initially introduced in this subsection, we use the KDDCUP99 dataset to pretrain the model completely, retain the parameters of the convolutional layer that could extract low-dimensional features, and then train and adjust the last three layers of the neural network through finetuning.

Result and Discussion

The datasets used in this experiment are the gas pipeline and KDDCUP99 datasets. The gas pipeline dataset is an industrial control network laboratory-scale ICS dataset based on Modbus application layer protocol published by Professor T. Morris of Mississippi State University. The KDDCUP99 dataset is a network connection dataset obtained from a simulated US Air Force LAN costing 9 weeks.

Because the KDDCUP99 dataset has a total of 5 million items, which is massive, we take only 10 % of these items for experimentation. The experimental data include approximately 100 000 items, which accounts for approximately 20% of this dataset. The gas pipeline dataset has a total of 97 019 items, of which 61 156 are normal samples. Every data stream in the target is processed into 324 pieces of data and source domains are processed into 41 pieces of data by preprocessing each set of traffic data. Figure 4.1 shows the results of data visualization. Each dataset in the source domain is processed into a grayscale image of 7 pixel 7 pixel by a 6-bit 0 supplement, and each dataset in the target domain is processed into a grayscale image of 18 pixel18 pixel pixels. Processing samples in this form to the residual CNN is clearly feasible.

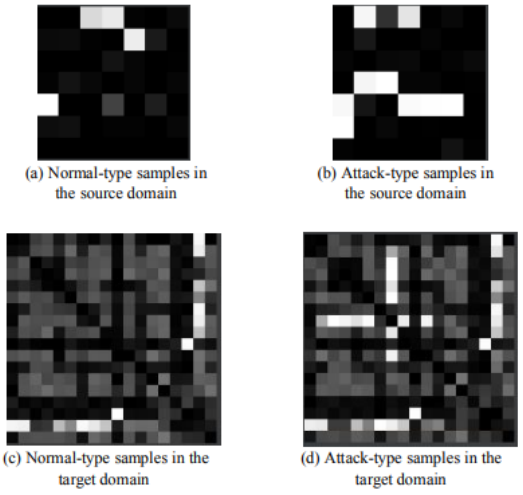


Fig. 4.1 Data visualization results.

In this experiment, all eight layers of the model are fine-tuned by utilizing transfer learning. The deep learning method used in this article initializes the model randomly and then optimizes all model parameters without transfer learning. The result in Fig. 4.2 shows that the pretrained model using fine-tuning converges faster and has a smaller loss and higher accuracy than the model using deep learning when the gas pipeline dataset is used for training. This finding indicates that transfer learning is significant in this environment. Both training methods

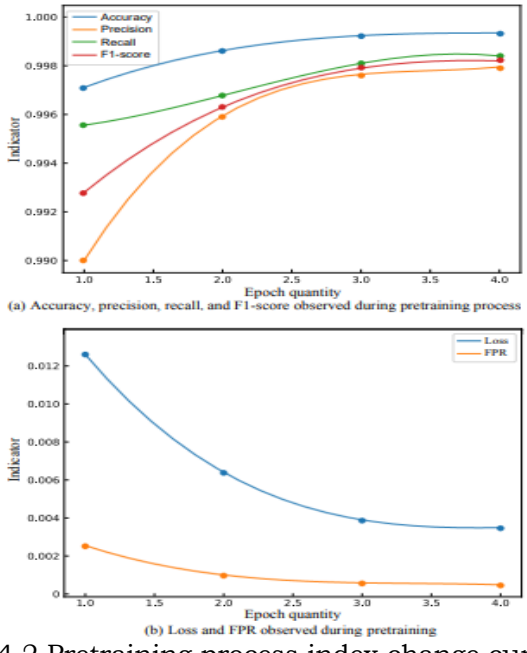
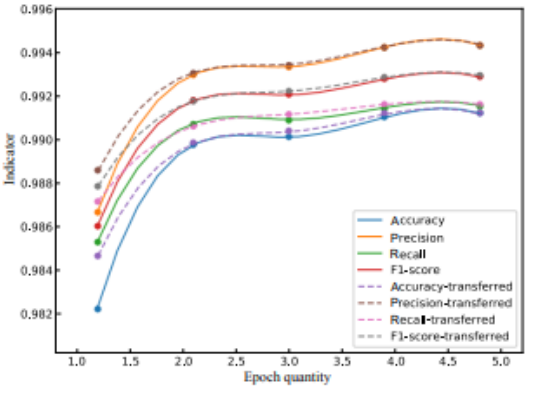


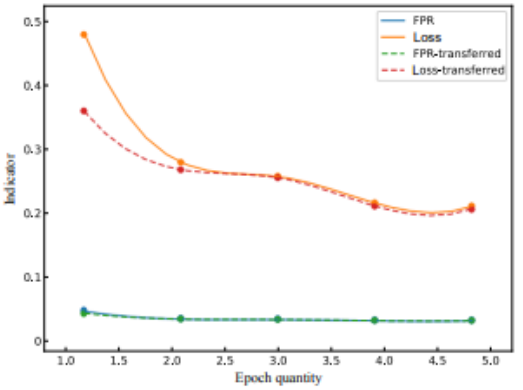
Fig. 4.2 Pretraining process index change curves.

are completed in approximately 81 minutes. The comparison shown in Table 1 reveals that the score of the method of fine-tuning three layers in the ICS flow anomaly detection index is close to the first two methods, which also greatly reduces the training time of the model.

The comparison results in Table 3 show that the precision indicators of reciprocal data and Auto Encoder (AE)COne Class Support Vector Machine (OCSVM)



(a) Accuracy, precision, recall, and F1-score of fine-tuning and deep learning observed during pretraining process



(b) Loss and FPR of fine-tuning and deep learning observed during pretraining process

Fig. 4.3 Fine-tuning and training of the model using deep learning convergence comparison curves.

are higher than those of other machine learning algorithms. However, the recall indicator of these algorithms is low, which means the detected positive samples are actually not all positive. The recall index of Generative Adversarial Networks (GAN) is high but its other indices are low, thus indicating that its comprehensive performance is poor. Although the F1- score of AE+OCSVM is high, which indicates that it has good overall performance, its recall value is quite low, which means some negative samples (abnormal types) may be misclassified as positive samples. These indicators are unsuitable for abnormal detection in ICSs. The algorithms proposed in this paper are clearly superior to those algorithms in terms of the indicators of interest.

Conclusion and Future Work

Network security is a popular and important topic. The network security of ICSs is of great importance for a country. This paper uses data visualization to convert flow data into images. Specifically, we build an eight-layer residual neural network and use fine-tuning technology for transfer learning to detect abnormal datasets of ICSs. Experimental results show that transfer learning for residual CNNs is effective in this field. The depth of the model also ensures that it has a certain generalizability. The residual structure effectively prevents gradient explosion or gradient disappearance. The model can provide reliable predictions for unknown or differently distributed abnormal data through short-term training by transfer learning. Compared with other anomaly detection algorithms, the algorithm proposed in this paper results in superior indicators. The method we proposed not only solves the problem associated with training time for deep learning models by transfer learning, but also meets the requirements of ICSs in terms of evaluation indicators. At present, the model we constructed solves the two classification problem, but a refined classification of abnormal traffic data is still desirable. In the future work, we will perform multi classification of abnormal traffic data, track the characteristics of different abnormal data types, and then reliably classify them to further ensure network security in ICSs.

References

1. A. R. Sadeghi, C. Wachsmann, and M. Waidner, Security and privacy challenges in industrial Internet of Things, in Proceedings of the 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 2015, pp. 1–6.
2. L. Obergon, InfoSec reading room secure architecture for industrial control systems, SANS Institute InfoSec, GIAC(GSEC) Gold Certification, vol. 1, pp. 1–27, 2014.
3. C. Markman, A. Wool, and A. A. Cardenas, A new burstDFA model for SCADA anomaly detection, in Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy, Dallas, TX, USA, 2017, pp. 1–12.
4. M. Mantere, I. Uusitalo, M. Sailio, and S. Noponen, Challenges of machine learning based monitoring for industrial control system networks, in Proceedings of the 2012 26th International Conference on Advanced Information Networking and Applications Workshops, Fukuoka, Japan, 2012, pp. 968–972.
5. R. Zhao, R. Q. Yan, Z. H. Chen, K. Z. Mao, P. Wang, and R. X. Gao, Deep learning and its applications to machine health monitoring: A survey, Mechanical System and Signal Processing, vol. 115, pp. 213–237, 2019
6. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Q. Zhou, W. Li, and P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020
7. S. N. Shirazi, A. Gouglidis, K. N. Syeda, S. Simpson, A. Mauthe, I. M. Stephanakis, and D. Hutchison, Evaluation of anomaly detection techniques for SCADA communication resilience, in Proceedings of the 2016 Resilience Week (RWSr), Chicago, IL, USA, 2016, pp. 140–145.

8. Rinartha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
9. Rinartha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
10. Malaka, I. G., Syarif, S., Arsyad, M. A., Baso, Y. S., & Usman, A. N. (2021). Development of women's reproductive health application as android-based learning media of adolescent knowledge. *International Journal of Health & Medical Sciences*, *4*(2), 182-188. https://doi.org/10.31295/ijhms.v4n2.1685