International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Machine-Learning Based Nano-Biosensing Study on Cancer Diagnosis and IoT Applications

[1]Kashish Bansal, [2]Nabeena Ameen, [3]Sumeet Mathur, [4]Arjun Vinay Avadhani, [5]Satwinder Singh, [6]Divya Chikati

Submitted:28/05/2023 **Revised**:15/07/2023 **Accepted**: 25/07/2023

Abstract: Cancer is a very big cause of death & an extremely costly illness to treat. The likelihood of a cure & survival rates increase with early cancer discovery, but sadly, most tumours are only discovered after they have spread to other parts of the body. Biosensors are tools created to identify a particular biological analyte by translating the intricate biological interactions into an electrical signal whose strength is related to the analyte's concentration. Nanotechnology with nanoparticles enhances & modifies the bio-recognition element portion to improve the bio-sensing phenomenon & makes it one of the hottest issues attracting the scientific fraternity. The primary goal of this study is to examine several nanostructures that have been applied to bio-sensing, & certain implementations in the areas of cancer diagnosis & IoT, & also a brief introduction to machine-learning-based bio-sensing. To categorise microarray data in this study, IOT & machine learning (ML) methods were applied. Two sets of data were used to generate them: one having 1,919 protein types & the other with 24,481 protein types for 97 individuals, 46 of whom had a reoccurring illness & 51 of whom did not. According to the study's findings, before feature reduction, logistic regression (LR) yielded the highest outcomes (90.23%) & also Random Forest yielded good outcomes (67.22%). Support Vector Machine had the best accuracy rates - 99.23% in both techniques in the first data & 87.87% in Random Logistic Regression (RLR) & 88.82% in LTE in the second data. To conclude, nanotechnology development surely helped biosensors to advance to new heights.

Keywords*: Nanostructures; Cancer Detection; Biosensor; Machine Learning*

1. Introduction

Cancer is an abnormal & uncontrolled cell growth that is caused by an accumulation of certain genetic & epigenetic abnormalities. This definition describes cancer as "cancer." It's possible that genetics & the environment both play a role in its development. Uncontrolled cell division results in the establishment of a tumour mass, which, with the passage of time, gradually loses its ability to maintain normal homeostatic checks & balances. It is increasingly impossible to treat cancer as it spreads to further body organs & systems as it advances in its progression. In general, cancer is a multifactorial disease, & it contains a wide range of genetic & epigenetic modifications. These variations disrupt the cellular microenvironment, which is connected with the beginning & progression of the disease (Khan et al., 2020).

Per the National Cancer Institute (NCI), 1,735,350 people were diagnosed having cancer in the US in 2018, with 609,640 people succumbing to the disease. Breast, lung, prostate, colon, along with rectum, & skin cancer, as well as melanoma, along with bladder cancer, & also non-Hodgkin lymphoma, along with kidney, as well as renal pelvis, & endometrial cancer, leukaemia, pancreatic along with thyroid, & liver cancer are extremely prevalent types of cancer, in that order. Pancreatic, thyroid, & liver malignancies are also prevalent. According to 2013–2015 figures, 38.4% of men & women will be diagnosed with cancer. Despite technological advances & novel treatment options, late diagnosis & a poor prognosis still contribute to a low 5-year cancer survival rate. The idea of bio-sensing is firmly ingrained into the majority of organic living forms, & from an evolutionary perspective, this has helped them to live in hostile situations & in the presence of predators. Typical illustrations of this phenomenon include the heightened olfactory capabilities of canines, sharks' electro-sensitive nature, & the toxin-sensing capacities of some algae (Aldewachi et al., 2018). Also, the fundamental operating principles of a biosensor are depicted in a schematic form in Figure 1, which can be found here.

1Student, Electrical Engineering, Indian Institute of Technology Indore
Email ID- kashibansal15@gmail.com
2Assistant professor(Sr grade) Department of Information Technology BS Abdur Rahman Crescent Institute of Science and Technology Chennai, India,
Email ID- nabeena@crescent.education
3Assistant Professor, Department of Computer Science & Engineering, University of Waikato Joint Institute at Zhejiang University Hangzhou, China
Email ID- sumeet.mathur@hotmail.com
4Research Scholar, Department of Computer Science Engineering, PES University Bengaluru, Karnataka
Email ID- avadhaniarjun@gmail.com
5Student, Indian Institute of Management Amritsar, Punjab
Email ID- satwinders.mba08@iimamritsar.ac.in
6Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
Email ID- chdivya333@kluniversity.in

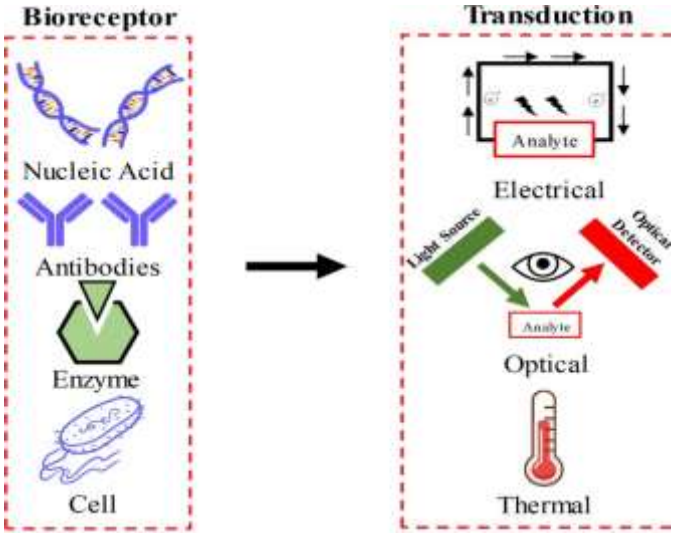

Fig 1: Given schematic displaying the working principle of biosensors based on various transduction methods (Banerjee et al., 2021).

The previous literature on this subject is extensively discussed in the section that follows.

2. Literature Review:

TABLE 1: LITERATURE REVIEW

AUTHORS AND YEAR	METHODOLOGY	FINDINGS
(Benjamin & Júnior, 2023)	traditional methods & novel biosensors for cancer diagnosis	investigated biosensors & nanomaterials for cancer detection as reliable diagnostic techniques.
(Song et al., 2023)	introduced wearable biosensor research. First, wearable biosensor-detected bodily fluids are suggested. Next, micro-nano fabrication methods & wearable biosensor properties are reviewed. The paper also discusses application & information processing. Wearable physiological pressure along with sweat sensors, & also self-powered biosensors are instances of groundbreaking research.	Readers were shown how these sensors detect with examples. Finally, present problems & future prospects are offered to further this study area & extend practical applications.
(Gasmi, 2021)	IoT wearable networks & wireless microcontrollers are utilized for contact tracking, bio-sensory point-of-care testing, & the Internet of Medical Things to limit viral disease spread.	Internet of Medical Things to fight viral diseases, IOT DSN & D2D to detect COVID-19, & simulations to validate IoT wearables.
(Saeed et al., 2021)	The manufactured sensors can automatically conduct quantitative analysis using digital & also microfluidic technologies, advancing wearable sensors & creating new opportunities & problems. Material innovation along with AI development is pushing the development of MIPs-based tumour detection sensors, which have not yet been employed in hospitals.	Discussed the sensors' difficulties & prospects as intelligent & implantable devices.
(Perdomo et al., 2021)	Low-detection sensors identify diseases early, contain infections & viral propagation, monitor biodata, control food quality, & elucidate high-resolution genome-phenome correlations. Introduced signal transduction components & also target-selective receptors (organic molecules like DNA along with enzyme,	current advances in nanomaterials for biomarker detection & quantification, advantages & cons of transduction techniques, & field difficulties & prospects.

	as well as antibody, & more or low molecular weight synthetic moieties) of a Bio-Nano sensor. Nanostructured materials, especially 2D materials along with nanoparticles, have improved biocompatibility, sensitivity, selectivity, wearability, & detection limits in electronic, electromechanical, optical, & electrochemical transducers.	

This study provides a concise summary of current developments in the creation of Nano-biosensors by merging functional biomolecules with various types of nanomaterials. These nanomaterials include carbon nanotubes, graphene, carbon dots, & others. In addition to this, this study also considers the use of Nanosensors to bring about substantial advancements in the cancer diagnostic process using point-of-care devices.

3. Methodology:

The data analysed in this work come from the Python language's numerical NumPy package, which is utilised to interpret multidimensional data like matrix arrays & allows one to perform mathematical operations on them. This study used the Pandas library, which enables data structuring, the Scikit-learn library; includes ML algorithms for classification, along with regression, & also clustering, & the Keras library; offers deep learning (DL) applications. The study's data (Abdullah Hamad et al., 2022) incorporated 1,927 features directly from 133 individuals; 11 - healthy & 122 - patients. This was the first amount of data used. The other set from the study (Song et al., 2023) has 24481 traits belonging to 97 individuals, 51 of whom are not, 46 of whom had a breast cancer recurrence. The matrix containing these data has 133 rows & 1919 columns. The other data consists of 24481 columns & 97 rows. Patients & non-patients were coded in the first data. To facilitate training & testing, these data are split into two categories. The machine is taught using the training group, & also the effectiveness of the classification is evaluated using the test group.

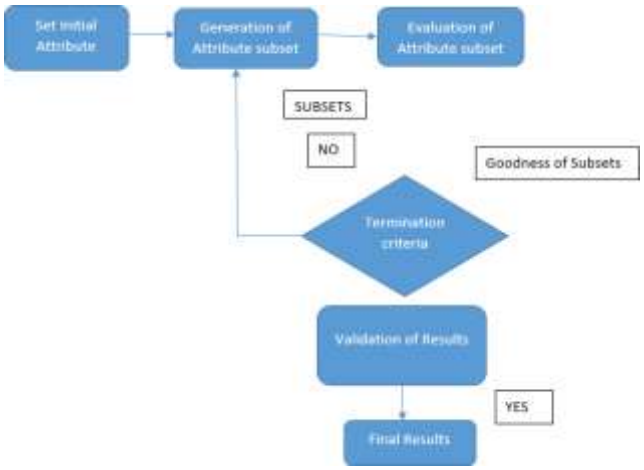

Fig 1: Diagnosing Cancer Utilising IOT along with Machine Learning Methods

4. Results and Discussions:

Python was used to apply seven traditional ML methods to two different sets of cancer data: the first set contained 133 samples & 1919 characteristics, & the second set contained 97 samples & 24481 features. Both sets of data were about cancer. In the course of carrying out these procedures, tests were carried out without making use of any feature elimination strategy. According to this, the best results were found using LR, which had a percentage of 99.23% in initial data, & also the random forest approach, which had a percentage of 67.42% in the second data. In the graphs that may be seen on the following pages, you'll also find the results that were found using alternative methodologies. After that, the same dataset was reduced in size using the LTE approach in order to select the top 50 features to maintain. As a consequence of this, the SVM approach obtained the best classification outcome in the first data set with 99.23%, & in the second data set, it found the best classification outcome with 88.82%.

The graph below also displays the outcomes of different methodologies' use of the data. To reduce the size of the dataset even more, the RLR feature selection approach was utilised once more, & the result was the retention of only the top 50 features. As a consequence of this, the SVM approach turned out to be the method with the highest accuracy, achieving a score of 99.23% in the first data & 87.87% in the second data.

It is clear that the K-FCV algorithm has not altered & that the other five algorithms provide better results compared to the first example, while the LR approach following RLR has a lesser rate. It is also clear that the results from the first instance are worse when using the LR algorithm, that the K-FCV algorithm is equivalent, & that the results from the other five ways are better. Finally, it is seen that the decision tree method has the lowest results, & also the LR method has the highest results, as shown in the figure below.

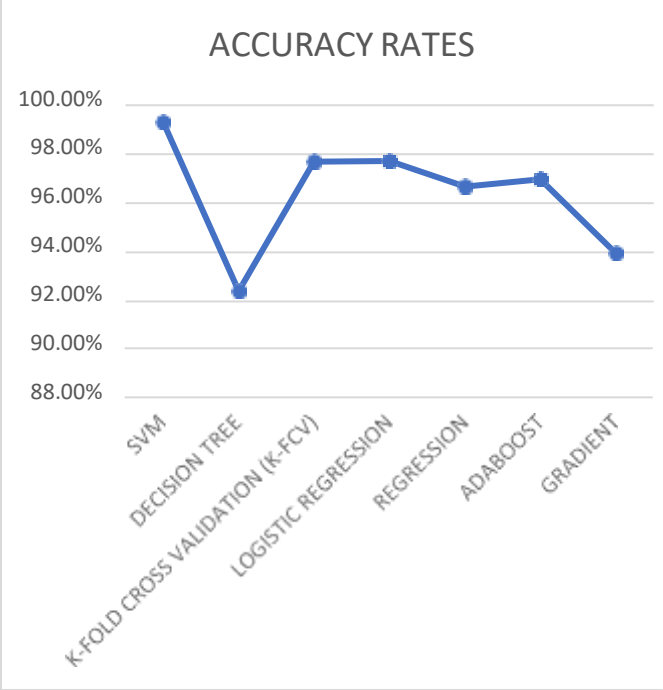

Fig 2: Results pre feature selection, post RLR feature selection method, & post LTE feature elimination method.

All data is summarised in Figure 3, with the SVM having the highest classification accuracy at 98.98% & the decision tree having the lowest at 90.28%. Now, SVM, a DL algorithm, was used to analyse the same dataset. However, rather than employing the same number of hidden layers, a variety of neuron types were utilised in each layer.

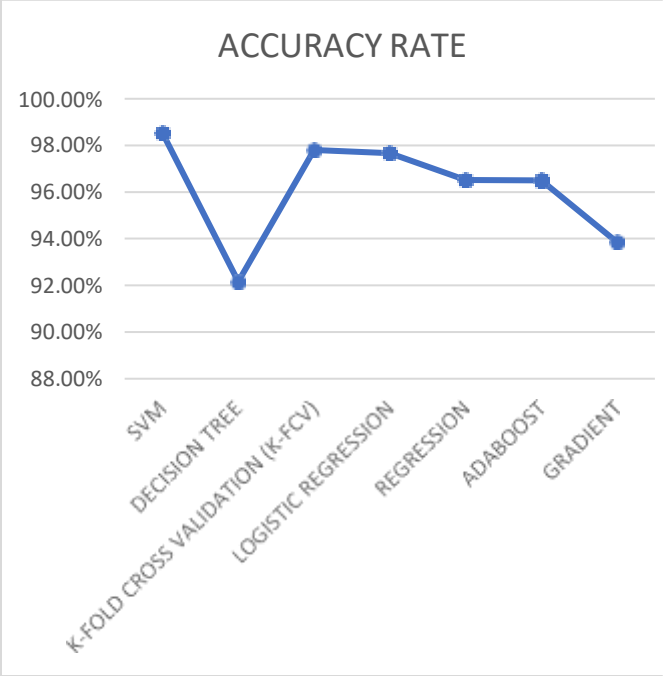

Fig 3: All results pre & post feature elimination.

Under each column in Figure 4, the total number of hidden layers that were utilised, as well as the number of neurons that were present in each layer, are displayed.

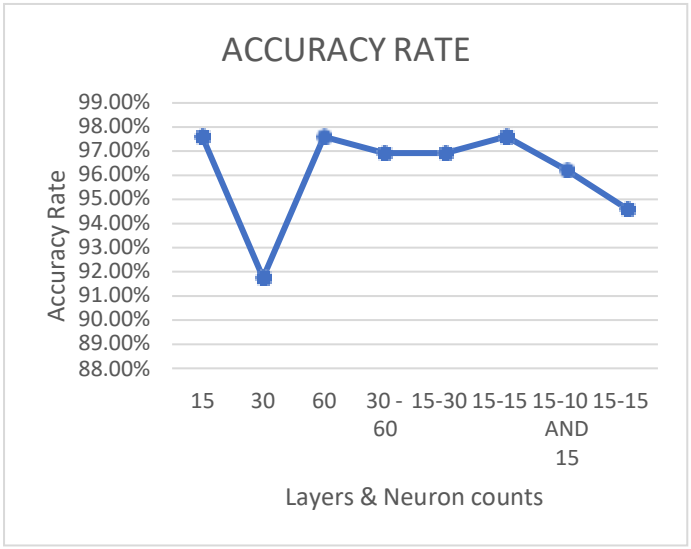

Fig 4: Results of DL by different layers.

The second dataset consists of 24,481 different attributes & 97 different samples. In this dataset, the feature elimination approach was applied after the fact, & the outcomes were compared. Initially, the method wasn't used because it wasn't used in the first case. In Figure 5, it is clear that each of the seven ML algorithms only achieves a poor level of accuracy when classifying this data before any feature removal strategy is applied. When compared to the state they were in before feature deletion, the accuracy rates of all algorithms have seen significant improvements. The SVM method produced the greatest results, with an accuracy rate of 87.87%, & it can be seen that all of the algorithms' accuracy rates have greatly improved in comparison to the "first case." Once again, it can be shown that the SVM algorithm achieves the highest classification rate.

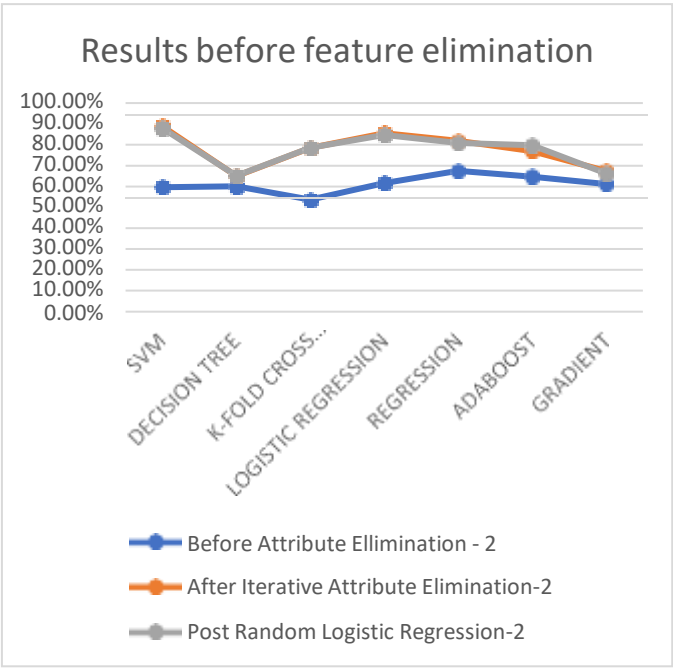

Fig 5: Results pre feature elimination in the 2nd data, by implementing RLR to 2nd data, & also by HT implementation to 2nd data.

The classification results showed that the SVM had the greatest accuracy overall with 78.81%, while the decision tree had the lowest accuracy overall with 63.39%. As may be seen in Figure 7, MLP did not attain particularly high levels of success. This study observed that the highest result, which is achieved with 68.72%, is achieved with 30 neurons being used in a single hidden layer, but the worst result is achieved with 60 neurons being utilised in a single hidden layer. It has been demonstrated that utilising 15-15 or 15-30 neurons within two hidden layers produces better outcomes than utilising 30-60 neurons in a single hidden layer. It can be observed that using three hidden layers with the ratio 15-10-5 yields the second-best result. If the no.of hidden layers is kept constant, it may say that increasing the number of neurons will have a negative impact on the final output. On the whole, this can have stated that the performance of ML algorithms on this very small number of data is superior.

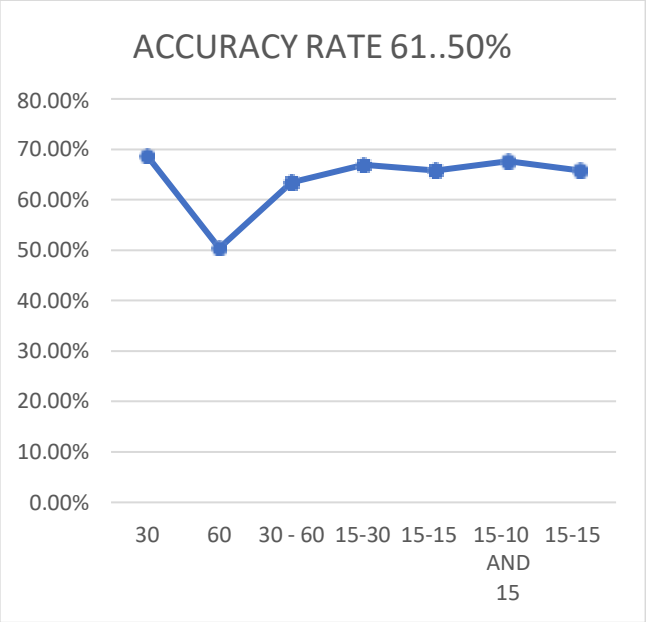

Fig 7: Results of DL in 2nd data by various layers.

5. Conclusions:

To conclude, massive datasets can now be utilised to identify cancer, a potentially fatal disorder, utilising microarray technology. IOT & ML methods were employed in this work to categorise microarray data. These were produced using two information sets: one having 1,919 protein types & also one having 24,481 protein types for 97 individuals, of whom 46 had a reoccurring illness & 51 did not. Separate applications of each classification technique were made to the data without any size or feature reduction. Second, the first scenario was contrasted with two additional feature reduction strategies. ML methods like Adaboost along with Gradient Boosting Machine are applied in this situation. The best results were obtained by the LR method (90.23%), followed by the Random Forest method (67.22%) before any feature reduction techniques were used. SVM had the highest accuracy rate in the first set of data (99.23%) for both techniques, whereas in the second set of data (87.87% for RLR & 88.82% for LTE), SVM had the highest accuracy rate. With MLP, DL was also accomplished. Using it at various depths, the relationship between depth & classification accuracy was investigated. The accuracy rate gradually decreased as the no. of layers grew over time. In the first set of data, the maximum accuracy rate was 97.69%, whereas in the second, it was 68.72%. So, DL does not become more accurate when additional layers are added.

References

[1] Khan, M. S., Iftikhar, A., Shubair, R. M., Capobianco, A. D., Braaten, B. D., & Anagnostou, D. E. (2020). Ultra-Wideband Antenna with MIMO Diversity for 5G Wireless Communication. *arXiv preprint arXiv:2007.02294*.

[2] Ding, R., Ye, M., Zhu, Y., Zhao, Y., Liu, Q., Cao, Y., & Xu, J. (2023). Toward Dynamic Detection of Circulating Tumor Cells Exploiting Specific Molecular Recognition Elements. *Chemosensors*, *11*(2), 99.

[3] Aldewachi, H., Chalati, T., Woodroofe, M. N., Bricklebank, N., Sharrack, B., & Gardiner, P. (2018). Gold nanoparticle-based colorimetric biosensors. *Nanoscale*, *10*(1), 18-33.

[4] Banerjee, A., Maity, S., & Mastrangelo, C. H. (2021). Nanostructures for biosensing, with a brief overview on cancer detection, IoT, & the role of machine learning in smart biosensors. *Sensors*, *21*(4), 1253.

[5] Benjamin, S. R., & Júnior, E. J. M. R. (2023). Detection of Cancer Biomarker by Advanced Biosensor. In *Targeted Cancer Therapy in Biomedical Engineering* (pp. 437-464). Singapore: Springer Nature Singapore.

[6] Gasmi, A. (2021). Enabled IoT Applications for Covid-19. In *Computational Intelligence Techniques for Combating COVID-19* (pp. 305-331). Cham: Springer International Publishing.

[7] Saeed, N., Loukil, M. H., Sarieddeen, H., Al-Naffouri, T. Y., & Alouini, M. S. (2021). Body-centric terahertz networks: Prospects & challenges. *IEEE Transactions on Molecular, Biological & Multi-Scale Communications*, *8*(3), 138-157.

[8] Song, Z., Zhou, S., Qin, Y., Xia, X., Sun, Y., Han, G., ... & Zhang, Q. (2023). Flexible & Wearable Biosensors for Monitoring Health Conditions. *Biosensors*, *13*(6), 630.

[9] Perdomo, S. A., Marmolejo-Tejada, J. M., & Jaramillo-Botero, A. (2021). Bio-nanosensors: Fundamentals & recent applications. *Journal of The Electrochemical Society*, *168*(10), 107506.

[10] Abdullah Hamad, A., Thivagar, M. L., Bader Alazzam, M., Alassery, F., Hajjej, F., & Shihab, A. A. (2022). Applying dynamic systems to social media by using controlling stability. Computational Intelligence & Neuroscience, 2022.

[11] Karunakaran, V. ., & Geetha, A. . (2023). A Secured Software Defined Network Architecture for Mini Net using POX Controller. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 226–231. https://doi.org/10.17762/ijritcc.v11i4.6443

[12] Botha, D., Dimitrov, D., Popović, N., Pereira, P., & López, M. Deep Reinforcement Learning for Autonomous Robot Navigation. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/140

[13] Aoudni, Y., Donald, C., Farouk, A., Sahay, K. B., Babu, D. V., Tripathi, V., & Dhabliya, D. (2022). Cloud security based attack detection using transductive learning integrated with hidden markov model. Pattern Recognition Letters, 157, 16-26. doi:10.1016/j.patrec.2022.02.012