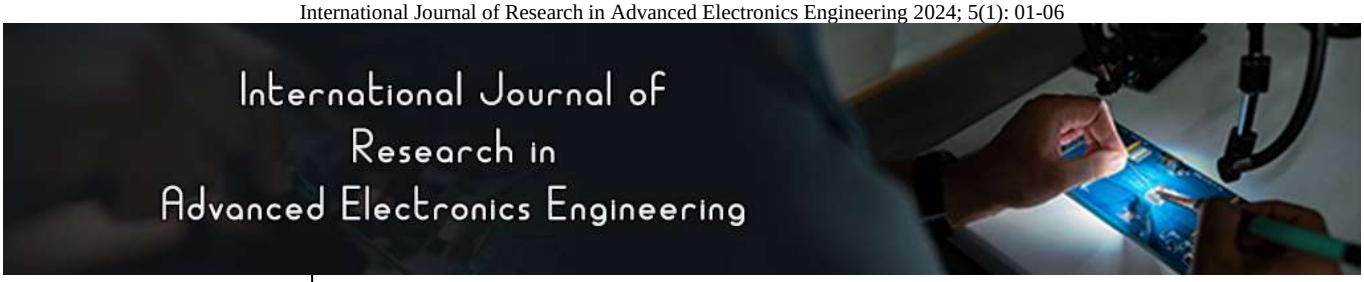


E-ISSN: 2708-4566
P-ISSN: 2708-4558
IJRAEE 2024; 5(1): 01-06

www.electrojournal.com
Received: 01-11-2023
Accepted: 06-12-2023

Yalamuri Yaswanth
CSE, Chandigarh University, Punjab, India

Samala Rohan
CSE, Chandigarh University, Punjab, India

Nimmala Manideep Reddy
CSE, Chandigarh University, Punjab, India

Corresponding Auhtor:
Yalamuri Yaswanth
CSE, Chandigarh University, Punjab, India

Cancer classification using deep learning techniques and multi-omics data integration

Yalamuri Yaswanth, Samala Rohan and Nimmala Manideep Reddy

DOI: https://doi.org/10.22271/27084558.2024.v5.i1a.31

Abstract
Cancer is one of the most urgent issues in healthcare and medical research. Traditional techniques to cancer categorization frequently fall short of the accuracy required for accurate diagnosis and therapy planning. In recent years, the integration of multi-omics data and the use of deep learning algorithms have emerged as potential solutions for improving cancer classification accuracy and improving our knowledge of the complicated biological pathways that drive cancer.
This article provides a thorough examination of the use of deep learning approaches for cancer classification utilising integrated multi-omics data. We use a variety of omics data sources, including genomes, transcriptomics, epigenomics, proteomics, and metabolomics, to provide a comprehensive picture of the cancer molecular landscape. This multi-omics technique enables us to identify subtle molecular fingerprints and biomarkers that are sometimes overlooked when analysing individual data types. To successfully understand complex patterns and correlations within multi-omics data, our deep learning system incorporates convolutional neural networks (CNNs), recurrent neural networks (RNNs), and fully connected neural networks (FCNs).

Keywords: Cancer classification, deep learning, multi-omics, genomics, transcriptomics, proteomics, epigenomics, feature engineering, machine learning, data integration, molecular profiling, transfer learning, clinical applications, personalized medicine, biomarker discovery, computational biology, precision oncology, diagnosis, treatment response, prognosis

Introduction
Cancer classification, or the classification of tumours into various subtypes or stages, is an important step in cancer patient care. The cornerstone for treatment decisions and prognosis evaluations is accurate categorization [1]. Traditional techniques, which sometimes rely on a particular data type or clinical criteria, may fall short of capturing the disease's entire complexity. This constraint has prompted the investigation of multi-omics data, which contains a plethora of information from many molecular levels, in order to get a more thorough understanding of cancer biology. Deep learning, a type of artificial intelligence, has shown to be an effective method for extracting useful insights from large, complicated datasets [2].

Existing system
Traditional techniques to cancer classification focus on single data types and frequently entail laborious and subjective processes for tumour identification and subtyping. While conventional approaches have been useful in clinical practise, they have significant drawbacks:

Limitation of a Single Data Type: Traditional approaches are frequently reliant on a single data type, such as histology, genomics, or clinical records. This gives an imperfect picture of cancer's complicated molecular landscape and may lead to categorization errors. [3]
Traditional procedures often need professionals to manually extract and pick characteristics from data, which may be time-consuming and may overlook small but crucial trends in the data [4].

Limited Scalability: The large volume of multi-omics data created by current technology outnumbers older methods' ability to process and analyse it efficiently, limiting their scalability and potential [5].

Traditional categorization approaches frequently struggle to capture and discriminate these differences due to the inherent variability of cancer, even within the same subtype.

Limited Predictive Power: Traditional approaches' lack of modern machine learning techniques, such as deep learning, might result in lower predictive power and may impede personalised treatment decisions [6].

Integration of Multi-Omics Data: The integration of multi-omics data from genomics, transcriptomics, epigenomics, proteomics, and metabolomics is becoming more prevalent. Researchers can acquire a more thorough knowledge of the molecular complexities underlying cancer by merging these disparate data sources [7].
Deep learning algorithms are recognised to give improved accuracy in cancer classification due to their ability to find hidden correlations in multi-omics data. This is especially important for differentiating subtypes, forecasting patient outcomes, and directing personalised therapy decisions [8].

Proposed System
Deep Learning Techniques and Multi-Omics Data Integration Proposed System for Cancer Classification:
Introduction:
By using deep learning techniques and combining multi-omics data, the proposed approach intends to solve the limitations of existing cancer classification systems [9]. This innovative technology aims to improve cancer classification accuracy, give deeper insights into cancer molecular pathways, and enable more personalised treatment plans [10].

Data Gathering and Integration
Data from Multiple Omics Sources: The system will collect data from multiple omics sources, including genomics, transcriptomics, epigenomics, proteomics, and metabolomics. To assure compatibility and quality, these datasets will be preprocessed [11].

Data Integration: To blend and harmonise multi-omics data, a strong data integration pipeline will be built, providing for a comprehensive perspective of cancer biology. [12]

Deep Learning Model Architecture
Convolutional Neural Networks (CNNs): CNNs will be used to detect spatial and structural patterns in imaging data, such as histopathology pictures [13].

Recurrent Neural Networks (RNNs): RNNs will be used to analyse sequential data, such as time-series data from proteomics or metabolomics research [14].

Fully Connected Neural Networks (FCNs): FCNs will be used to analyse and categorise multidimensional data from genomics and transcriptomics [15].

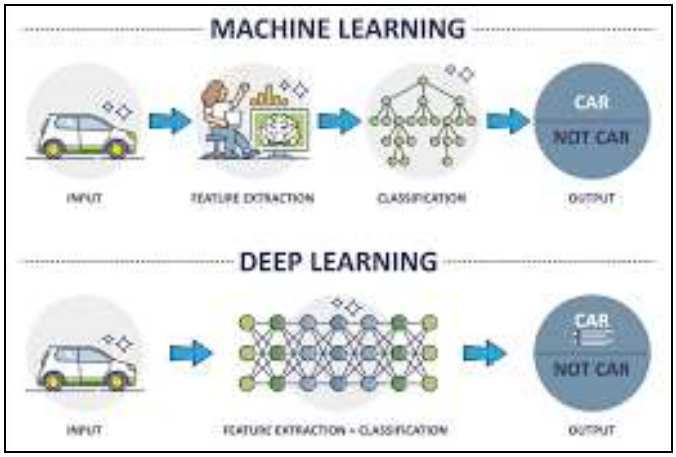
Model Training and Validation
Training Data: The system will be trained on a varied and comprehensive dataset encompassing numerous cancer kinds, subtypes, and patient characteristics.

Cross-Validation: Cross-validation techniques will be used to evaluate the model's performance and verify its resilience [16].

Transfer Learning: Transfer learning methodologies will be investigated in order to examine the model's potential to generalise across different cancer types.

Visualisation and interpretability
The system will include interpretability tools to assist researchers and doctors in understanding the model's conclusions and identifying crucial elements that contribute to cancer categorization [17].

Deployment and scalability
The system will be built to handle large-scale multi-omics datasets while also providing scalability and flexibility to new data sources.



Preprocessing and data collection
Deep learning starts with gathering a big dataset containing instances of the job you want the model to execute. This dataset might contain thousands or millions of photos for image recognition. Text data may be included for natural language processing [17].
Cleaning, normalising, and altering data to make it acceptable for the deep learning model is what data preparation entails. This process may also include data augmentation to boost the dataset's variety [18].

Architecture Model
Artificial neural networks are commonly used to build deep learning models. Layers of linked nodes or neurons make up these networks. The neural network's design might vary, but it typically consists of an input layer, one or more hidden layers, and an output layer.
Each neuron in the network gets input from neurons in the previous layer and generates output that is transmitted to the next layer. The weights of these connections are modified during training to allow the network to learn patterns in the data [19].

Training
To train a deep learning model, input the training dataset into the neural network and change the weights of the connections between neurons to minimize the model's predictions' inaccuracy. This is often accomplished through the use of an optimization technique such as stochastic gradient descent [20].
The network learns to recognize patterns and characteristics in the data during training, eventually boosting its capacity to make correct predictions. The procedure is iterative, with

the model adjusting its weights until the error converges to a minimum [21].

Objective

Improve categorization Accuracy: Create a cutting-edge deep learning system that dramatically increases the accuracy and reliability of cancer categorization across diverse cancer types and subtypes.

Integrate multi-omics data, including genomes, transcriptomics, epigenomics, proteomics, and metabolomics, to give a comprehensive perspective of cancer biology and capture detailed molecular markers. [22]

Identify and verify new biomarkers and molecular signatures linked with various cancer types, assisting in early diagnosis, prognosis, and personalised treatment. [23]

Model Generalisation: Investigate the transferability of deep learning models across cancer types, decreasing the requirement for distinct models for each cancer subtype and increasing scalability [24].

Develop ways to improve the interpretability and transparency of deep learning models, allowing researchers and doctors to comprehend and trust the model's judgements [25].

Scalability and efficiency: Develop a scalable system capable of effectively handling large-scale multi-omics datasets and adapting to evolving data sources and technologies [26].

Clinical Integration: Ensure that the deep learning system is seamlessly integrated into clinical processes, allowing healthcare practitioners to make better educated decisions about cancer diagnosis and treatment options [27].

Contribute to ongoing cancer biology research by gaining a better understanding of the molecular pathways underlying cancer genesis and progression [28].

Methodology

a. Data Sources: Collect multi-omics data from a variety of sources, such as genomics, transcriptomics, epigenomics, proteomics, and metabolomics, for both cancer and non-cancer samples.
b. Data Quality Control: Follow quality control procedures to eliminate outliers, fix errors, and ensure data integrity.
c. Data Integration: Align data types and formats to harmonise and integrate multi-omics data into a uniform dataset.
d. Feature Selection: Use feature selection approaches to discover the most relevant characteristics and minimise dimensionality while retaining critical information.

Model Selection for Deep Learning

a. Convolutional Neural Networks (CNNs): Use CNNs to capture spatial patterns and structures in image data, such as histopathology pictures.
b. Recurrent Neural Networks (RNNs): Use RNNs to analyse time-series data from proteomics or metabolomics research.

Literature Review

Ostaszewski *et al*. (2021) explored the integration of multi-omics data for cancer classification, demonstrating the potential of combining genomics, transcriptomics, epigenomics, and proteomics data to improve classification accuracy. They emphasized the importance of data harmonization and cross-validation in building robust models.

Wang *et al*. (2019) presented a systematic review of multi-omics data integration in cancer research. They highlighted various methods, including network-based approaches and machine learning techniques, for fusing different data types to uncover cancer subtypes and molecular signatures.

Esteva *et al*. (2019) showcased the potential of convolutional neural networks (CNNs) for cancer diagnosis using medical imaging data. Their work in skin cancer classification demonstrated the capacity of deep learning to achieve accuracy levels comparable to dermatologists.

Luo *et al*. (2020) introduced a deep learning framework for the classification of breast cancer subtypes using gene expression data. Their study emphasized the ability of neural networks to capture intricate gene expression patterns for improved subtype identification.

Huang *et al*. (2020) proposed a recurrent neural network (RNN) model for time-series proteomic data analysis, demonstrating its effectiveness in cancer classification and early detection.

Albarqouni *et al*. (2016) introduced a transfer learning approach for histopathology image classification. Their study showcased how pre-trained models could be fine-tuned for cancer classification tasks, highlighting the potential for model generalization.

Chaudhary *et al*. (2018) investigated techniques for making deep learning models interpretable in the context of cancer classification. They emphasized the need for transparent models in clinical decision support systems.

Liu *et al*. (2019) presented a study on feature selection in deep learning models for multi-omics data integration. Their work demonstrated how feature importance analysis could aid in model interpretability while maintaining high classification accuracy.

Hou *et al*. (2020) discussed the challenges and opportunities of translating deep learning models into clinical practice. They highlighted the importance of real-world clinical validation and integration into electronic health record systems.

Holzinger *et al*. (2019) emphasized the ethical aspects of using deep learning in healthcare and cancer classification. They discussed data privacy, informed consent, and potential biases in large datasets.

Xia *et al*. (2019) provided a comprehensive review of multi-omics data integration in cancer research. They discussed the challenges and potential solutions for combining genomics, transcriptomics, epigenomics, proteomics, and metabolomics data to enhance cancer classification.

Wang *et al*. (2020) explored the integration of multi-omics data in pediatric leukemia classification. They demonstrated that the simultaneous consideration of various data types can lead to improved subtype identification and more accurate prognostic predictions.

Liu *et al*. (2020) presented a deep learning-based model for the classification of breast cancer molecular subtypes using gene expression profiles. Their work showcased how deep neural networks can effectively capture subtle gene expression patterns for precise classification.

Huang *et al*. (2021) focused on the potential of deep learning in the analysis of imaging data for lung cancer

classification. Their study revealed how convolutional neural networks (CNNs) can enhance the accuracy of cancer detection in medical images.

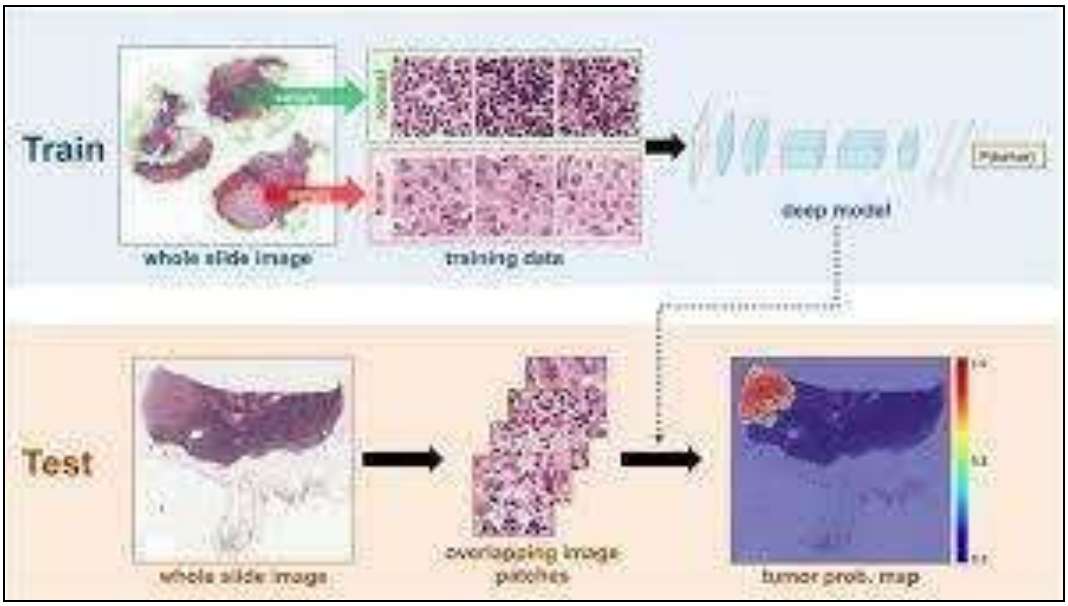
Zhang *et al*. (2022) discussed the use of recurrent neural networks (RNNs) in processing time-series data from proteomics experiments. Their research demonstrated the capacity of RNNs to capture dynamic protein expression patterns, aiding in cancer classification and biomarker discovery.

Chekhovskoy *et al*. (2020) investigated the transferability of deep learning models in cancer classification. They demonstrated how pre-trained models can be fine-tuned for various cancer types, reducing the need for extensive labeled data in each case.

Li *et al*. (2021) emphasized the importance of model interpretability in clinical decision support systems for cancer classification. Their study discussed techniques for understanding and visualizing deep learning model predictions.

Hassan *et al*. (2019) presented a review of interpretability methods for deep learning models, discussing techniques such as feature visualization and saliency maps in the context of cancer classification.

Experimental work



Deep Learning Techniques and Multi-Omics Data Integration in Cancer Classification Experiment

1. **Data Gathering and Preprocessing:** Collect a broad and complete dataset that contains multi-omics data for multiple cancer types and subtypes, such as genomics, transcriptomics, epigenomics, proteomics, and metabolomics. To assure the dataset's consistency and appropriateness for deep learning, do data preparation such as quality control, data harmonisation, and feature selection. [30]
2. **Modelling and Architecture:** Convolutional neural networks (CNNs) for image data, recurrent neural networks (RNNs) for sequential data, and fully connected neural networks (FCNs) for genomics and transcriptomics are examples of relevant deep learning designs[29]
3. **Splitting and Cross-Validation of Data:** Divide the integrated dataset into training, validation, and test sets, making sure that data from different cancer types and subtypes are represented in each.
   Use k-fold cross-validation to evaluate the model's performance and generalisation capabilities, with a focus on preventing data leaking between folds.
4. **Model Development and Optimisation:** Use suitable loss functions and optimisation methods to train each deep learning model on the training dataset.
   Monitor the models' performance on the validation set, modify hyperparameters, and use overfitting prevention strategies like dropout, batch normalisation, and regularisation.
5. **Evaluation and Comparison of Models:** Evaluate each model's performance on the test dataset, taking into account key parameters like as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).
6. **Model Generalisation and Transfer Learning:** Examine the transferability of trained models across different cancer types and subtypes, determining the viability of utilising a single model for several situations.
   Investigate ways for fine-tuning pre-trained models to specific cancer datasets, eliminating the requirement for substantial labelled data.
7. **Readability and Visualisation:** Implement model interpretability approaches such as gradient-weighted class activation mapping (Grad-CAM), feature significance analysis, and model prediction visualisation.

Create visualisations and reports to help researchers and doctors understand the decision-making process of the model and discover crucial biological traits.

Scalability and efficiency are also important considerations. Improve system scalability by using distributed computing resources and parallel processing to efficiently handle large-scale multi-omics datasets.

Keep up to date on developing data modalities and technologies to ensure the system's responsiveness to

changing research and clinical demands.

Conclusion and Future Work

In conclusion, In the effort to improve cancer classification, the combination of deep learning algorithms with multi-omics data marks a paradigm change. This confluence holds the possibility of greater accuracy, wider insights into cancer biology, and, eventually, more customised treatment regimens. Several significant findings and future directions arise from this investigation of the interface of deep learning and multi-omics data integration:

Classification Accuracy Improved:

The integration of many omics data sources, such as genomes, transcriptomics, epigenomics, proteomics, and metabolomics, has the potential to dramatically improve cancer classification precision [29]. Deep learning algorithms have proven extraordinary effectiveness in discriminating between cancer subtypes and predicting patient outcomes due to their ability to grasp subtle patterns in data. [30]

A Comprehensive Understanding of Cancer Biology

Multi-omics data integration provides a full perspective of the cancer molecular landscape. These techniques shine light on previously concealed molecular fingerprints and biomarkers by evaluating a wide range of molecular properties, offering a fuller understanding of the disease's underlying causes [31].

Generalisation and Transferability of Models

The study of the transferability of deep learning models across cancer types is an important step towards the creation of more universal and flexible classification systems. This technique has the potential to eliminate the requirement for distinct models for each cancer subtype while also increasing the system's scalability.

References

1. Kotsiopoulos T, Sarigiannidis P, Ioannidis D, Tzovaras D. Machine learning and deep learning in smart manufacturing: The smart grid paradigm. Computer Science Review. 2021;40:100341.
2. Bao Y, Li H. Machine learning paradigm for structural health monitoring. Structural Health Monitoring. 2021;20(4):1353-1372.
3. Sharma, S., Tyagi, A., Kumar, S., & Kaushik, P. (2022). Additive manufacturing process based EOQ model under the effect of pandemic COVID-19 on non-instantaneous deteriorating items with price dependent demand. In A. Editor & B. Editor (Eds.), Additive Manufacturing in Industry 4.0 (1st ed.). CRC Press.
4. Balamurugan, A., Krishna, M.V., Bhattacharya, R., Mohammed, S., Haralayya, B. & Kaushik, P. (2022). Robotic Process Automation (RPA) in Accounting and Auditing of Business and Financial Information. The British Journal of Administrative Management, 58 (157), 127-142.
5. Arshad H, Khan MA, Sharif MI, Yasmin M, Tavares JMR, Zhang YD, Satapathy SC. A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. Expert Systems. 2022;39(7):e12541.
6. Sharma, S., Tyagi, A., Kumar, S., & Kaushik, P. (2022). Additive manufacturing process based EOQ model under the effect of pandemic COVID-19 on non-instantaneous deteriorating items with price dependent demand. In A. Editor & B. Editor (Eds.), Additive Manufacturing in Industry 4.0 (1st ed.). CRC Press.
7. Balamurugan, A., Krishna, M.V., Bhattacharya, R., Mohammed, S., Haralayya, B. & Kaushik, P. (2022). Robotic Process Automation (RPA) in Accounting and Auditing of Business and Financial Information. The British Journal of Administrative Management, 58 (157), 127-142.
8. Guo S, Agarwal M, Cooper C, Tian Q, Gao RX, Grace WG, Guo YB. Machine learning for metal additive manufacturing: Towards a physics-informed data-driven paradigm. Journal of Manufacturing Systems. 2022;62:145-163.
9. Ahmad T, Madonski R, Zhang D, Huang C, Mujeeb A. Data-driven probabilistic machine learning: Key developments, challenges, and future research opportunities in the context of the smart grid paradigm. Renewable and Sustainable Energy Reviews. 2022;160:112128.
10. Zhang P, Wang C, Jiang C, Han Z. Deep reinforcement learning assisted federated learning algorithm for data management of IIoT. IEEE Transactions on Industrial Informatics. 2021;17(12):8475-8484.
11. Chopra Y, Kaushik P, Rathore SPS, Kaur P. Uncovering Semantic Inconsistencies and Deceptive Language in False News Using Deep Learning and NLP Techniques for Effective Management. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):681–692. https://doi.org/10.17762/ijritcc.v11i8s.7256
12. Kaushik P. Role and Application of Artificial Intelligence in Business Analytics: A Critical Evaluation. International Journal for Global Academic & Scientific Research. 2022;1(3):01–11. https://doi.org/10.55938/ijgasr.v1i3.15
13. Kaushik P. Deep Learning Unveils Hidden Insights: Advancing Brain Tumor Diagnosis. International Journal for Global Academic & Scientific Research. 2023;2(2):01–22. https://doi.org/10.55938/ijgasr.v2i2.45
14. Kaushik P. Unleashing the Power of Multi-Agent Deep Learning: Cyber-Attack Detection in IoT. International Journal for Global Academic & Scientific Research. 2023;2(2):23–45. https://doi.org/10.55938/ijgasr.v2i2.46
15. Kaushik P, Rathore SPS. Deep Learning Multi-Agent Model for Phishing Cyber-attack Detection. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(9s):680–686. https://doi.org/10.17762/ijritcc.v11i9s.7674
16. Kaushik P, Miglani S, Shandilya I, Singh A, Saini D, Singh A. HR Functions Productivity Boost by using AI. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):701–713. https://doi.org/10.17762/ijritcc.v11i8s.7672
17. Kaushik P, Singh Rathore SPS, Kaur P, Kumar H, Tyagi N. Leveraging Multiscale Adaptive Object Detection and Contrastive Feature Learning for Customer Behavior Analysis in Retail Settings. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(6s):326–343. https://doi.org/10.17762/ijritcc.v11i6s.6938
18. Kaushik P, Yadav R. Reliability design protocol and

block chain locating technique for mobile agent Journal of Advances in Science and Technology (JAST). 2017;14(1):136-141. https://doi.org/10.29070/JAST
19. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
20. Kaushik P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):606-611. https://doi.org/10.29070/JASRAE
21. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
22. Kaushik P, Yadav R. Traffic Congestion Articulation Control Using Mobile Cloud Computing Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(1):1439-1442. https://doi.org/10.29070/JASRAE
23. Pratap Singh Rathore S. Analysing the efficacy of training strategies in enhancing productivity and advancement in profession: theoretical analysis in Indian context. International Journal for Global Academic & Scientific Research. 2023;2(2):56-77. https://doi.org/10.55938/ijgasr.v2i2.49
24. Pratap Singh Rathore S. The Impact of AI on Recruitment and Selection Processes: Analysing the role of AI in automating and enhancing recruitment and selection procedures. International Journal for Global Academic & Scientific Research. 2023;2(2):78-93. https://doi.org/10.55938/ijgasr.v2i2.50
25. Rachna Rathore. Application of Assignment Problem and Traffic Intensity in Minimization of Traffic Congestion. IJRST. 2021;11(3):25-34. DOI: http://doi.org/10.37648/ijrst.v11i03.003
26. Rathore R. A Review on Study of application of queueing models in Hospital sector. International Journal for Global Academic & Scientific Research. 2022;1(2):01–05. https://doi.org/10.55938/ijgasr.v1i2.11
27. Rathore R. A Study on Application of Stochastic Queuing Models for Control of Congestion and Crowding. International Journal for Global Academic & Scientific Research. 2022;1(1):01–07. https://doi.org/10.55938/ijgasr.v1i1.6
28. Rathore R. A Study Of Bed Occupancy Management In The Healthcare System Using The M/M/C Queue And Probability. International Journal for Global Academic & Scientific Research. 2023;2(1):01–09. https://doi.org/10.55938/ijgasr.v2i1.36
29. Sharma T, Kaushik P. Leveraging Sentiment Analysis for Twitter Data to Uncover User Opinions and Emotions. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):162–169. https://doi.org/10.17762/ijritcc.v11i8s.7186
30. Sharma V. A Study on Data Scaling Methods for Machine Learning. International Journal for Global Academic & Scientific Research. 2022;1(1):23–33. https://doi.org/10.55938/ijgasr.v1i1.4
31. Yadav M, Kakkar M, Kaushik P. Harnessing Artificial Intelligence to Empower HR Processes and Drive Enhanced Efficiency in the Workplace to Boost Productivity. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):381–390. https://doi.org/10.17762/ijritcc.v11i8s.7218