OPTIMIZED CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

[1]Yoheswari S
[1] Department of Computer Science & Engineering, K.L.N College of Engineering, Pottapalayam – 630612, Tamilnadu, India
[1]yoheswari1988@gmail.com

Abstract: Cardiovascular diseases (CVD) represent a significant cause of morbidity and mortality worldwide, necessitating early detection for effective intervention. This research explores the application of machine learning (ML) algorithms in predicting cardiovascular diseases with enhanced accuracy by integrating optimization techniques. By leveraging data-driven approaches, ML models can analyze vast datasets, identifying patterns and risk factors that traditional methods might overlook. This study focuses on implementing various ML algorithms, such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, optimized through techniques like hyperparameter tuning, cross-validation, and feature selection to improve prediction accuracy.

The research methodology involves data preprocessing, feature engineering, model training, and performance evaluation. We employ optimization methods such as Genetic Algorithms and Grid Search to fine-tune model parameters, ensuring robust and generalizable models. The dataset used includes patient medical records, with features like age, blood pressure, cholesterol levels, and lifestyle habits serving as inputs for the ML models. Evaluation metrics, including accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC), assess the model's predictive power.

Our results demonstrate that optimized ML models outperform standard approaches, providing reliable predictions that could assist healthcare professionals in early diagnosis and personalized treatment planning. The study concludes by discussing the implications of these findings for clinical practice and the potential for future enhancements, such as incorporating real-time data from wearable devices or exploring deep learning techniques for even greater predictive accuracy.

Key words: Attribute-Based Keyword Search (ABKS), Secure Cloud Storage, Data Encryption, Access Control, Search Optimization

Corresponding Author: Yoheswari S
K.L.N. College of Engineering, Pottapalayam, Tamil Nadu, India
Mail: yoheswari1988@gmail.com

Introduction:

Cardiovascular diseases (CVD) have long been a leading cause of death globally, with the World Health Organization reporting millions of deaths annually due to heart-related conditions. The increasing prevalence of CVD underscores the need for early detection and timely intervention, which can significantly reduce the risk of severe outcomes. Traditional diagnostic methods, while effective, often rely on extensive clinical testing and the expertise of healthcare professionals, which can be time-consuming and resource-intensive. Consequently, there is a growing interest in leveraging technology, particularly machine learning (ML), to predict CVD risk more efficiently and accurately.

Machine learning, a subset of artificial intelligence, has shown great promise in various fields, including healthcare, by enabling the analysis of large and complex datasets. ML algorithms can automatically learn patterns and relationships within data, making them well-suited for predictive tasks. In the context of CVD prediction, ML can analyze patient data—such as medical history, lifestyle factors, and physiological measurements—to identify individuals at high risk of developing cardiovascular conditions.

This research explores the application of ML algorithms in predicting CVD, with a particular focus on enhancing model performance through optimization techniques. Optimization plays a crucial role in machine learning by improving the accuracy, robustness, and generalizability of models. Techniques such as hyperparameter tuning, feature selection, and cross-validation are employed to refine the models and ensure that they provide reliable predictions.

The study utilizes a comprehensive dataset of patient records, which includes variables such as age, gender, blood pressure, cholesterol levels, smoking status, and physical activity. These variables are used as inputs for various ML models, including Decision Trees, Random Forest, Support Vector Machines, and Neural Networks. Each model is subjected to rigorous optimization to ensure the best possible predictive performance.

The research is structured as follows: first, the data is preprocessed to handle missing values, outliers, and irrelevant features. Next, feature engineering is performed to enhance the input data's relevance to the predictive task. The ML models are then trained on the processed data, with optimization techniques applied to improve their performance. Finally, the models are evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to determine their effectiveness in predicting CVD.

This study aims to demonstrate the potential of optimized ML algorithms in predicting cardiovascular diseases, thereby providing a valuable tool for healthcare professionals. By improving the accuracy and efficiency of CVD prediction, these models could support early diagnosis and intervention, ultimately reducing the burden of cardiovascular diseases on individuals and healthcare systems.

Data Collection and Preprocessing:

The initial step involves gathering a comprehensive dataset, including patient medical records with features such as age, gender, blood pressure, cholesterol levels, and lifestyle habits. Data preprocessing is crucial, as it involves cleaning the data by handling missing values, removing outliers, and normalizing the data to ensure uniformity. This step also includes splitting the data into training and testing sets to validate the model's performance.

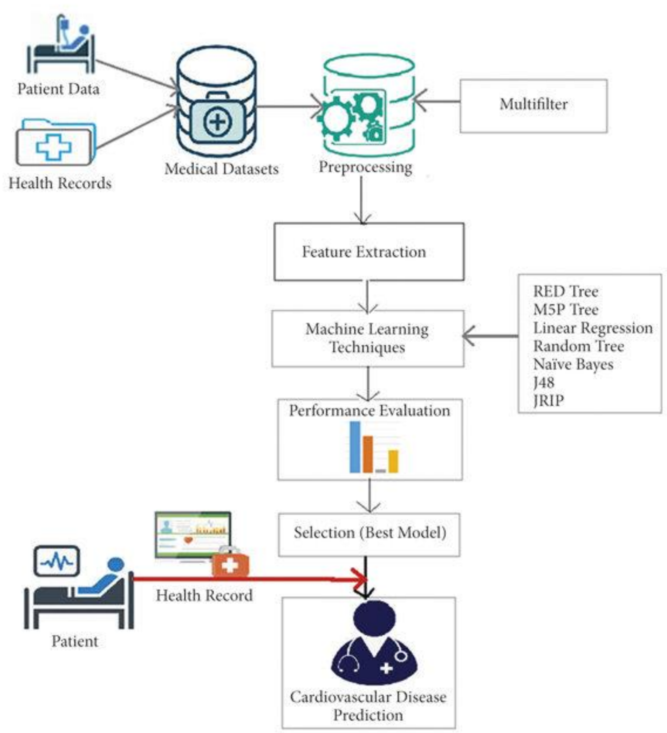

Fig.1. Framework of the proposed cardiovascular disease prediction system**:**

Feature Engineering:

Feature engineering enhances the relevance of the data by selecting and transforming variables that significantly impact the predictive task. Techniques such as Principal Component Analysis (PCA) or feature scaling are applied to reduce dimensionality and highlight the most informative features, thereby improving model accuracy and reducing computational complexity.

Model Selection:

Multiple machine learning algorithms are selected for comparison, including Decision Trees, Random Forest, Support Vector Machines, and Neural Networks. Each model is chosen based on its ability to handle non-linear relationships and high-dimensional data, which are common in cardiovascular datasets.

Model Optimization:

This step involves fine-tuning the chosen models using optimization techniques such as Grid Search, Genetic Algorithms, or Bayesian Optimization. Hyperparameters such as learning rate, depth of trees, and regularization parameters are adjusted to enhance model performance. Cross-validation is employed to prevent overfitting and ensure that the model generalizes well to new data.

Model Evaluation:

The optimized models are evaluated on the test set using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. These metrics provide insights into the model's predictive power and its ability to identify individuals at high risk of cardiovascular disease. The study concludes by summarizing the findings and discussing potential future enhancements, such as incorporating real-time data from wearable devices or exploring deep learning techniques for even greater predictive accuracy.

Conclusions:

The findings of this study highlight the potential of machine learning algorithms, optimized through advanced techniques, in accurately predicting cardiovascular disease risk. These optimized models demonstrate significant improvements in predictive accuracy, offering a valuable tool for healthcare professionals in early diagnosis and personalized treatment planning. Future research could explore the integration of real-time data from wearable devices, providing continuous monitoring and prediction. Additionally, deep learning techniques could be investigated to further enhance model performance, potentially offering even more precise and reliable predictions.

Reference:

1. Ramesh, G., Gorantla, V. A. K., & Gude, V. (2023). A hybrid methodology with learning based approach for protecting systems from DDoS attacks. *Journal of Discrete Mathematical Sciences and Cryptography*, *26*(5), 1317-1325.
2. Logeshwaran, J., Gorantla, V. A. K., Gude, V., & Gorantla, B. (2023, September). The Smart Performance Analysis of Cyber Security Issues in Crypto Currency Using Blockchain. In *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)* (Vol. 6, pp. 2235-2241). IEEE.

3. Komatireddy, S. R., Meghana, K., Gude, V., & Ramesh, G. (2023, December). Facial Shape Analysis and Accessory Recommendation: A Human-Centric AI Approach. In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 182-191). IEEE.
4. Sriramulugari, S. K., Gorantla, V. A. K., Gude, V., Gupta, K., & Yuvaraj, N. (2024, March). Exploring mobility and scalability of cloud computing servers using logical regression framework. In *2024 2nd International Conference on Disruptive Technologies (ICDT)* (pp. 488-493). IEEE.
5. Gorantla, V. A. K., Gude, V., Sriramulugari, S. K., Yuvaraj, N., & Yadav, P. (2024, March). Utilizing hybrid cloud strategies to enhance data storage and security in e-commerce applications. In *2024 2nd International Conference on Disruptive Technologies (ICDT)* (pp. 494-499). IEEE.
6. Bharathi, G. P., Chandra, I., Sanagana, D. P. R., Tummalachervu, C. K., Rao, V. S., & Neelima, S. (2024). AI-driven adaptive learning for enhancing business intelligence simulation games. *Entertainment Computing*, *50*, 100699.
7. Rao, S. D. P. (2024). HARNESSING AI FOR EVOLVING THREATS: FROM DETECTION TO AUTOMATED RESPONSE.
8. Rao, S. D. P. (2022). PREVENTING INSIDER THREATS IN CLOUD ENVIRONMENTS: ANOMALY DETECTION AND BEHAVIORAL ANALYSIS APPROACHES.
9. Sanagana, D. P. R., & Tummalachervu, C. K. (2024, May). Securing Cloud Computing Environment via Optimal Deep Learning-based Intrusion Detection Systems. In *2024 Second International Conference on Data Science and Information System (ICDSIS)* (pp. 1-6). IEEE.
10. Kanth, T. C. (2023). EXPLORING SERVER-LESS COMPUTING FOR EFFICIENT RESOURCE MANAGEMENT IN CLOUD ARCHITECTURES.
11. Kanth, T. C. (2023). EFFICIENT STRATEGIES FOR SEAMLESS CLOUD MIGRATIONS USING ADVANCED DEPLOYMENT AUTOMATIONS.
12. Kanth, T. C. (2023). CONTEMPORARY DEVOPS STRATEGIES FOR AUGMENTING SCALABLE AND RESILIENT APPLICATION DEPLOYMENT ACROSS MULTI-CLOUD ENVIRONMENTS.
13. Thangapalani, L., Dharini, R., & Keerthana, R. (2023, May). Securing Medical Image Transmission using Memetic Algorithm. In *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-8). IEEE.
14. Vennila, D., Vinotha, C., Shanthakumari, A., & Thangapalani, L. Convex Optimization Algorithm for Product Recommendation Using Microblogging Information. *Journal of Data Mining and Management*, *2*(1).
15. Lawan, L. A., & Roy, S. K. Assessing the Predictive Capability of the Theory of Planned Behavior in the Nigerian Context: A Study of Intention to Founding New Business. In *Constructive Discontent in Execution* (pp. 231-248). Apple Academic Press.

16. Ibrahim, M., & Roy, S. K. (2023). Advancement of Nonlife Insurance in Both Public and Private Sectors in Bangladesh. In *Constructive Discontent in Execution* (pp. 209-230). Apple Academic Press.
17. Jain, M. B., & Roy, S. K. (2022). Student Motivation in Online Learning. *International Journal of Early Childhood*, (01), 4339-4346.
18. Jain, B., & Roy, S. K. (2022). Exploring the Pros and Cons of Promoting Interaction in Online Learning. *NeuroQuantology*, *20*(5), 5401.
19. Ibrahim, M., & Roy, S. K. (2022). Assessment of Profitability Achievement of Stateowned Non-life Insurance in Bangladesh. *NeuroQuantology*, *20*(6), 2883.
20. Roy, S. K. (2014). Factors Affecting (CRM) Practices in Commercial Banks a Case of Select Banks in India. *International journal of current research*, *6*(11), 10344-10351.
21. Gupta, R. C., & Roy, S. K. (1970). Studies on the pollen grains of Urena lobata Linn. *Cur Sci*.
22. Alikhan, J. S., Alageswaran, R., & Amali, S. M. J. (2023). Self-attention convolutional neural network optimized with season optimization algorithm Espoused Chronic Kidney Diseases Diagnosis in Big Data System. *Biomedical Signal Processing and Control*, *85*, 105011.
23. Alikhan, J. S., Alageswaran, R., & Amali, S. M. J. (2023). Dingo optimization based network bandwidth selection to reduce processing time during data upload and access from cloud by user. *Telecommunication Systems*, *83*(2), 189-208.
24. Rokade, U. S., Doye, D., & Kokare, M. (2009, March). Hand gesture recognition using object based key frame selection. In *2009 International Conference on Digital Image Processing* (pp. 288-291). IEEE.
25. Kshirsagar, K. P. (2015). Key Frame Selection for One-Two Hand Gesture Recognition with HMM. *International Journal of Advanced Computer Research*, *5*(19), 192.
26. Kumbhar, K., & Kshirasagar, K. P. (2015). Comparative study of CCD & CMOS sensors for image processing. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, *3*, 194-196.
27. Kshirsagar, K. P., & Doye, D. (2010, October). Object Based Key Frame Selection for Hand Gesture Recognition. In *2010 International Conference on Advances in Recent Technologies in Communication and Computing* (pp. 181-185). IEEE.
28. Ingle, A., Gour, A., & Kshirsagar, K. (2017). DDoS attack detection algorithms based on pattern classification and machine learning. *J Univ Shanghai Sci Technol. ISSN*, 1007-6735.
29. Kshirsagar, K. P., & Shinde, R. A. (2017, October). Comparing techniques of segmenting hand region. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 721-724). IEEE.
30. Rokade, R., Kshirsagar, K., Sonawane, J., & Munde, S. (2019). Analysis of Human-Machine Interaction Through Facial Expression and Hand-Gesture

Recognition. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(9S3).

31. Kshirsagar, K. P., Rokade, R., & Kamble, D. (2019). IoT based gesture recognition for smart controlling. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(4).
32. Kshirsagar, K. P., & Doye, D. D. (2015). Comparing key frame selection for one-two hand gesture recognition using different methods. *International Journal of Signal and Imaging Systems Engineering*, *8*(5), 273-285.
33. Kshirsagar, K. P., Sahu, R. M., Bankar, S. M., Moje, R. K., & Doye, D. D. (2013). K one hand gesture recognition. *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng*, *1*(7), 330-334.
34. Gujarkar, P., Lonkar, S., Jain, T., Nigal, S., Patil, P., Deshpande, P., ... & Ratnaparkhi, A. (2023, March). IoT based Smart Attendance System. In *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 1-6). IEEE.
35. Chishty, S., Langare, A., Sawant, S., & Kshirsagar, K. Industrial Data Acquisition. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, *11*(12), 14590-14594.
36. Raikwar, A. R., Sadawarte, R. R., More, R. G., Gunjal, R. S., Mahalle, P. N., & Railkar, P. N. (2018). Long-Term and Short-Term Traffic Forecasting Using Holt-Winters Method.
37. Kshirsagar, K. P. (2015). Segmentation Of One And Two Hand Gesture Recognition Using Key Frame Selection. *Segmentation of One and Two Hand Gesture Recognition using Key Frame Selection*.
38. Mukati, N., Namdev, N., Dilip, R., Hemalatha, N., Dhiman, V., & Sahu, B. (2023). Healthcare assistance to COVID-19 patient using internet of things (IoT) enabled technologies. *Materials today: proceedings*, *80*, 3777-3781.
39. Bansal, B., Jenipher, V. N., Jain, R., Dilip, R., Kumbhkar, M., Pramanik, S., ... & Gupta, A. (2022). Big data architecture for network security. *Cyber Security and Network Security*, 233-267.
40. Shrivastava, A., Nayak, C. K., Dilip, R., Samal, S. R., Rout, S., & Ashfaque, S. M. (2023). Automatic robotic system design and development for vertical hydroponic farming using IoT and big data analysis. *Materials Today: Proceedings*, *80*, 3546-3553.
41. Pandey, J. K., Jain, R., Dilip, R., Kumbhkar, M., Jaiswal, S., Pandey, B. K., ... & Pandey, D. (2022). Investigating role of iot in the development of smart application for security enhancement. In *IoT Based Smart Applications* (pp. 219-243). Cham: Springer International Publishing.
42. Gupta, N., Janani, S., Dilip, R., Hosur, R., Chaturvedi, A., & Gupta, A. (2022). Wearable sensors for evaluation over smart home using sequential minimization optimization-based random forest. *International Journal of Communication Networks and Information Security*, *14*(2), 179-188.

43. Gite, P., Shrivastava, A., Krishna, K. M., Kusumadevi, G. H., Dilip, R., & Potdar, R. M. (2023). Under water motion tracking and monitoring using wireless sensor network and Machine learning. *Materials Today: Proceedings*, *80*, 3511-3516.
44. Dilip, R., & Bhagirathi, V. (2013). Image processing techniques for coin classification using LabVIEW. *OJAI 2013*, *1*(1), 13-17.
45. Krishna, K. M., Borole, Y. D., Rout, S., Negi, P., Deivakani, M., & Dilip, R. (2021, September). Inclusion of cloud, blockchain and iot based technologies in agriculture sector. In *2021 9th international conference on cyber and IT service management (CITSM)* (pp. 1-8). IEEE.
46. Dilip, R. (2019). DESIGN AND DEVELOPMENT OF INTELLIGENT SYSTEM FOR HUMAN BODY DESIGN AND DEVELOPMENT OF INTELLIGENT SYSTEM FOR HUMAN BODY. *no. July*, 0-3.
47. Veeraiah, V., Thejaswini, K. O., Dilip, R., Jain, S. K., Sahu, A., Pramanik, S., & Gupta, A. (2024). The Suggested Use of Big Data in Medical Analytics by Fortis Healthcare Hospital. In *Adoption and Use of Technology Tools and Services by Economically Disadvantaged Communities: Implications for Growth and Sustainability* (pp. 275-289). IGI Global.
48. Dilip, R., Milan, R. K., Vajrangi, A., Chavadi, K. S., & Puneeth, A. S. (2021, November). Jumping robot: a pneumatic jumping locomotion across rough terrain. In *Journal of Physics: Conference Series* (Vol. 2115, No. 1, p. 012008). IOP Publishing.
49. Dilip, R., Borole, Y. D., Sumalatha, S., & Nethravathi, H. M. (2021, September). Speech based biomedical devices monitoring using LabVIEW. In *2021 9th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-7). IEEE.
50. Rekha, C. M., Shivakumar, K. S., & Dilip, R. (2020, October). Comparison of spacefactor, capacitance value and impregnated temperature in mpp oil impregnated polypropylene film AC capacitors. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 544-547). IEEE.
51. Dilip, R., & Bhagirathi, V. (2013). LAN Based Industrial Automation with GSM Connectivity. ICSEM-2013 Conference Proceedings.
52. Janani, S., Dilip, R., Talukdar, S. B., Talukdar, V. B., Mishra, K. N., & Dhabliya, D. (2023). IoT and Machine Learning in Smart City Healthcare Systems. In *Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities* (pp. 262-279). IGI Global.
53. Dilip, R., Solabagoudar, M. P., Chapi, N., & Vaidya, P. B. (2023). A Review of Surveillance and Fire Fighter Drone. *International Journal of Unmanned Systems Engineering*, *5*(2), 123-145.
54. Janani, S., Dilip, R., Talukdar, S. B., Talukdar, V. B., Mishra, K. N., & Dhabliya, D. (2023). IoT and Machine Learning in Smart City Healthcare Systems. In *Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities* (pp. 262-279). IGI Global.
55. Dilip, R., & Ramesh, K. B. (2020). Development of Graphical System for Patient Monitoring using Cloud Computing.

56. Mathuravalli, S. M. D., Narayanansamy Rajendran, D. K. B., Dilip, R., Ranjan, A., Das, I., & Chauhan, A. (2023). Deep Learning Techniques For Exoticism Mining From Visual Content Based Image Retrieval. *Journal of Pharmaceutical Negative Results*, 925-933.
57. Dilip, R., Samanvita, N., Pramodhini, R., Vidhya, S. G., & Telkar, B. S. (2022, February). Performance Analysis of Machine Learning Algorithms in Intrusion Detection and Classification. In *International Conference on Emerging Technologies in Computer Engineering* (pp. 283-289). Cham: Springer International Publishing.
58. Kumar, V. S., Thansekhar, M. R., Saravanan, R., & Amali, S. M. J. (2014). Solving multi-objective vehicle routing problem with time windows by FAGA. *Procedia Engineering*, *97*, 2176-2185.
59. Selvan, M. A., & Amali, S. M. J. (2024). RAINFALL DETECTION USING DEEP LEARNING TECHNIQUE.
60. Sudha, S., Baskar, S., Amali, S. M. J., & Krishnaswamy, S. (2015). Protein structure prediction using diversity controlled self-adaptive differential evolution with local search. *Soft Computing*, *19*, 1635-1646.
61. Kadhar, K. M. A., Baskar, S., & Amali, S. M. J. (2015). Diversity Controlled Self Adaptive Differential Evolution based design of non-fragile multivariable PI controller. *Engineering Applications of Artificial Intelligence*, *46*, 209-222.
62. Sivaramkumar, V., Thansekhar, M. R., Saravanan, R., & Miruna Joe Amali, S. (2018). Demonstrating the importance of using total time balance instead of route balance on a multi-objective vehicle routing problem with time windows. *The International Journal of Advanced Manufacturing Technology*, *98*, 1287-1306.
63. Brindha, S. (2021). A robust and adaptive fuzzy logic based differential evolution algorithm using population diversity tuning for multi-objective optimization. *Engineering Applications of Artificial Intelligence*, *102*, 104240.
64. Sivaramkumar, V., Thansekhar, M. R., Saravanan, R., & Miruna Joe Amali, S. (2017). Multi-objective vehicle routing problem with time windows: Improving customer satisfaction by considering gap time. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *231*(7), 1248-1263.
65. Zhang, J., Chen, C., Xiang, Y., Zhou, W., & Xiang, Y. (2012). Internet traffic classification by aggregating correlated naive bayes predictions. *IEEE transactions on information forensics and security*, *8*(1), 5-15.
66. Kirubahari, R., & Amali, S. M. J. (2024). An improved restricted Boltzmann machine using Bayesian optimization for recommender systems. *Evolving Systems*, *15*(3), 1099-1111.
67. Jemimah, J. P. P., & Miruna Joe Amali, S. (2024). An efficient PCCM masking scheme for PAPR reduction and encryption in OFDM-VLC system. *Telecommunication Systems*, 1-31.
68. Rekha, K. S., Amali, M. J., Swathy, M., Raghini, M., & Darshini, B. P. (2023). A steganography embedding method based on CDF-DWT technique for data hiding

application using Elgamal algorithm. *Biomedical Signal Processing and Control*, *80*, 104212.

69. Sashi Rekha, K., & Miruna Joe Amali, S. A. (2022). Efficient feature subset selection and classification using levy flight-based cuckoo search optimization with parallel support vector machine for the breast cancer data. *International Journal of Imaging Systems and Technology*, *32*(3), 869-881.
70. Kiran, A., Kalpana, V., Madanan, M., Ramesh, J. V. N., Alfurhood, B. S., & Mubeen, S. (2023). Anticipating network failures and congestion in optical networks a data analytics approach using genetic algorithm optimization. *Optical and Quantum Electronics*, *55*(13), 1193.
71. Lalithambigai, M., Kalpana, V., Kumar, A. S., Uthayakumar, J., Santhosh, J., & Mahaveerakannan, R. (2023, February). Dimensionality reduction with DLMNN technique for handling secure medical data in healthcare-IoT model. In *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 111-117). IEEE.
72. Kalpana, V., Mishra, D. K., Chanthirasekaran, K., Haldorai, A., Nath, S. S., & Saraswat, B. K. (2022). On reducing energy cost consumption in heterogeneous cellular networks using optimal time constraint algorithm. *Optik*, *270*, 170008.
73. Kalpana, V., & Karthik, S. (2020). Route availability with QoE and QoS metrics for data analysis of video stream over a mobile ad hoc networks. *Wireless Personal Communications*, *114*(3), 2591-2612.
74. Kalpana, V., & Karthik, S. (2018, February). Bandwidth Constrained Priority Based Routing Algorithm for Improving the Quality of Service in Mobile Ad hoc Networks. In *2018 International Conference on Soft-computing and Network Security (ICSNS)* (pp. 1-8). IEEE.