



Comparative Assessment of Machine Learning Algorithms for Effective Diabetes Prediction and Care

Praveena Nuthakki^{1*}, Pavankumar T.²

¹Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vadlamudi 522302, Vijayawada, India

* Corresponding Author Email: praveena.4u@gmail.com - ORCID: 0000-0001-9197-9469

²Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vadlamudi 522302, Vijayawada, India

Email: pavankumar-ist@kluniversity.in - ORCID: 0000-0001-9149-8630

Article Info:

DOI: 10.22399/ijcesen.606
Received : 08 June 2024
Accepted : 11 December 2024

Keywords :

Artificial Intelligence,
Machine Learning,
Diabetic,
PIMA.

Abstract:

The prevalence and impact of diabetes have increased significantly over time, posing a major concern for the healthcare sector globally, especially in India. This study aims to enhance diabetes prediction and management through the use of artificial intelligence (AI) and machine learning (ML) methodologies. We present a range of AI-driven approaches that leverage ML algorithms to classify and predict diabetes more effectively. While most studies utilize the PIMA dataset, a few notable cases have also incorporated custom datasets curated from select healthcare organizations. This research provides a comparative assessment of state-of-the-art diabetes prediction methods alongside carefully selected care strategies. The study is organized into three categories, each exploring distinct approaches, and analyzes methodologies, ML algorithms, accuracy results, and validation metrics. By examining key parameters and techniques, this paper considers diabetes prediction and care tailored to the Indian population, accounting for various influencing factors

1. Introduction

Artificial intelligence (AI) and machine learning (ML) are increasingly applied in healthcare to enhance patient care, disease diagnosis, treatment, and overall system efficiency. These technologies play a crucial role in medical imaging analysis, disease diagnosis, personalized treatment plans, drug discovery, remote patient monitoring, administrative tasks, and workflow optimization. By analyzing various data sources, including medical images, patient records, biomedical data, and scientific literature, AI algorithms assist healthcare professionals in decision-making and augment their capabilities [1,2]. Diabetes is a global health concern, with an estimated 80% likelihood of occurrence. India, in particular, faces a high diabetes prevalence, as reported by the International Diabetes Federation (IDF). The country initiated diabetes studies in the 1950s and 1960s, establishing the National Diabetes Registry in 1965 and the Diabetic Association of India (DAI) in the 1970s. Rapid urbanization, lifestyle changes, and increased

obesity have contributed to the rise of type 2 diabetes in India. The National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases, and Stroke (NPCDCS) was launched in 2010 to address non-communicable diseases, including diabetes. The DAI emphasizes preventive measures, lifestyle modifications, early diagnosis, and access to quality healthcare. This paper conducts a comprehensive literature review to present the current state and recent trends in diabetes prediction, focusing on the use of technology, methodology, datasets, and algorithms in machine learning and artificial intelligence [2-5].

This paper provides a comprehensive overview of diabetes through an in-depth literature review. It covers various aspects, including diet-related recommendations for type-2 diabetic patients, detection of diabetes related complications, and the disease's history in India. The emphasis is on preventive measures, lifestyle changes, early diagnosis, and access to quality healthcare to manage and control diabetes in the Indian population. The primary focus is on the importance

of early diabetes prediction, especially in India. The study is organized into sections covering Background, Literature Review, Analysis, Discussion, and Conclusions to encompass its full scope.

2. Background

2.1. Artificial Intelligence (AI) and Machine Learning (ML)

In opinion of scientists and researchers all around the world artificial intelligence and machine learning holds great promise in healthcare, it is not meant to replace healthcare professionals but to enhance their capabilities and support decision-making. Ethical considerations, data privacy, and regulatory frameworks are crucial in the development and deployment of AI in healthcare to ensure patient safety and maintain trust in the system. It's worth emphasizing that successful AI and machine learning application in healthcare necessitates high-quality data, robust model creation, validation, and adherence to privacy and security requirements. Furthermore, healthcare personnel play an important role in evaluating and incorporating machine learning outputs into their decision-making processes, assuring patient safety and the ethical usage of these technologies. AI and ML are being utilized more and more in the healthcare sector to enhance patient care, diagnose diseases, optimize healthcare data utilization, and improve overall efficiency of healthcare systems. These technologies are applied in various areas such as medical imaging analysis, disease diagnosis and prognosis, electronic health record (EHR) analysis, drug discovery and development, predictive analytics and risk stratification, remote patient monitoring, wearable devices, health catboats, and virtual assistants to a certain extent [3-6].

2.2. Diabetes Mellitus

Diabetes, a chronic condition, results from insufficient pancreatic insulin production or ineffective insulin use by the body, impacting blood sugar regulation. Diabetes type 1 involves inadequate insulin synthesis, requiring daily insulin therapy. Type 2 diabetes affects how the body uses glucose for energy and, if untreated, leads to elevated blood sugar levels, potentially causing harm to nerves and blood vessels. Preventable risk factors for type 2 diabetes include being overweight, lack of exercise, and genetic factors. Early diagnosis is crucial to avoid severe consequences, and regular check-ups can aid in early detection, although symptoms may be subtle. Type 2 diabetes, affecting over 95% of cases, was once considered adult-onset but is now increasingly prevalent in children.

Intermediate conditions, Impaired Glucose Tolerance (IGT) and Impaired Fasting Glycaemia (IFG), pose a high risk of developing type 2 diabetes. Additionally, gestational diabetes during pregnancy (type 3) and pre-diabetes (type 4) caused by increased insulin production due to genetic disorders or hormonal changes are noteworthy variations [5-8].

3. Literature review

This Diabetes is a complicated disease that is influenced by a number of factors, including genetic predisposition, lifestyle choices, and environmental factors. According to the international diabetes federation, India has the world's second-highest number of diabetes patients, with an estimated 77 million adults living with the disease in 2021, trailing only china, which has an estimated 116 million adults. The united states rank third, with an estimated 34 million adults living with diabetes in 2021. Only in India is this figure expected to rise to 101 million by 2030. Diabetes is a major public health concern in India, with efforts underway to raise awareness, improve access to care, and prevent the disease. Diabetes is a growing global health concern, and efforts are being made around the world to prevent and manage the disease. According to the International Diabetes Federation (IDF), India has one of the highest populations of diabetics in the world. When compared to many other populations around the world, Indians have a higher prevalence of diabetes. While the causes of diabetes in Indians differ from person to person, a number of key factors contribute to the high prevalence of diabetes in the Indian population. According to doctors and researchers, the main reasons are genetic predisposition, unhealthy diet, obesity and unhealthy lifestyle, insulin resistance, gestational diabetes, urbanization and socioeconomic factors, and cultural and social factors [4-7]. Diabetes in Indians differs from other populations in other countries in several ways. These distinctions can be found in terms of prevalence, risk factors, onset age, complications, and management. Diabetes in young Indians is caused by a variety of factors, including regional differences, genetic predisposition, lifestyle choices, and diabetes type. While providing a comprehensive analysis of all regions in India is difficult, here are some general factors that contribute to diabetes in young Indians, as well as their impact on health and quality of life [7].

Following extensive research presents the understanding of various papers was represented in three categories as below, which represent the state of the art in diabetes management and prediction. The first category represents diabetes and diet management for known type-2 diabetic patients and their basic background information, the second category describes diabetes prediction as a study of a specific group of people to analyze the known type-2 patient's population, which leads to the formulation of rules and suggestions primarily in the form of mobile applications, and the third category represents diabetes's multifaceted impact on different organs due to prolonged high blood sugar levels [7-9].

3.1 Category-I: Diabetes and Diet Management

This category describes the diet and food recommendations for diabetes management using recommendations taking into account various parameters, which include some health parameters and eating habits discussed by different researchers and their observations with regard to the prediction or classification of a person to either the non-diabetes or diabetes category. This table focuses on different methods used while accepting some user information in an app suggesting recommendations in diet and sometime exercise as well which works on rule based and category based rules applied for users. Rather this would have been something different to make it personalized and more precise. The below image represents the different categories researchers have find suitable for Diabetes Prediction and Care, we considered approximately 75 to 80 research papers for study of them few selected are used and cited here.

Table 1. Study of diet recommendation

Dietary Approach	Recommended Components	Specific Benefits
Low-Carbohydrate Diet	Reduces carb intake to 20-50g/day; focuses on proteins and fats	Lowers blood glucose levels, improves insulin sensitivity
Mediterranean Diet	High in fruits, vegetables, whole grains, olive oil, fish	Reduces risk of cardiovascular disease, supports weight loss
DASH Diet (Dietary Approaches to Stop Hypertension)	High in fruits, veggies, low-fat dairy, low in saturated fat	Lowers blood pressure, helps with weight control
Plant-Based Diet	Focuses on vegetables, fruits, whole grains, legumes, nuts	Improves insulin sensitivity, promotes weight loss
Ketogenic Diet	High-fat, very low-carb (5-10% carbs); moderate protein	Lowers blood sugar, can promote rapid weight loss
Paleo Diet	Emphasizes whole foods, lean proteins, fruits, and nuts	Reduces processed foods, promotes weight loss
High-Fiber Diet	25-30g fiber/day from whole grains, vegetables, fruits	Slows digestion, stabilizes blood glucose
Intermittent Fasting	Time-restricted eating (e.g., 16:8 or 5:2 fasting method)	May improve insulin sensitivity and assist in weight loss
Anti-Inflammatory Diet	Emphasis on foods like berries, leafy greens, fatty fish	Reduces inflammation, which is linked to insulin resistance
Glycemic Index (GI) Diet	Low-GI foods like legumes, nuts, whole grains	Minimizes blood sugar spikes

Table 2. Diabetes Prediction and Care Comparison based on Methodologies

Methodology	Key Component	Techniques/Models Used
Machine Learning (ML)	Predictive Modeling	SVM, Decision Trees, Random Forest, KNN
Deep Learning (DL)	Feature Extraction and Prediction	CNN, LSTM, RNN, MLP
Ensemble Learning	Model Combination	Bagging, Boosting, Stacking
Time Series Analysis	Temporal Data Analysis	ARIMA, LSTM, GRU
Natural Language Processing (NLP)	Text Data Analysis (e.g., medical records)	BERT, DistilBERT, LSTM for text classification
Computer Vision in Medical Imaging	Image-Based Diagnosis	CNN, ResNet, Transfer Learning
Fuzzy Logic Systems	Handling Uncertainty in Data	Fuzzy Inference Systems
IoT and Wearable Devices	Continuous Monitoring	Sensor networks, real-time data analysis
Self-Supervised Learning	Robust Model Training	Pretext tasks, contrastive learning

Table 3. Comparative Study Based on Dataset, Technology and Outcomes

Dataset and Methodology	Results obtained	Remarks
PIDD , Logistic Regression, Classification	highest Accuracy of 88.57% with ANN	The type 2 diabetes shows prominent effects with increased age and other physical factors
PIDD , Logistic Regression, Classification	Naive Bayes 76.30%	
PIMA, J48, K-mean, Feed forward-neural network, RB-Bayes, Naive Bayes, ANN with LSTM	LSTM with accuracy 86.08 in classification & 87.26% in prediction.	computational requirements and resource constraints , with no recommendations
LSVM, RBF-SVM, Poly-SVM, DT) as base learners and NSGA-II-Stacking model compared with k-NN	stacking-based ensemble approach, NSGA-II-Stacking worked better	No study about the exact feature impacting decision
PIMA, RF, SVM to compare with Twice-growth deep neural network (2GDNN) model of classification	2GDNN model of classification	Understanding which features contribute more to the classification decisions can provide insights into the biological or medical relevance of the features.
Clinical dataset using the snow sampling technique, 403 instances with 11 attributes.	Random Forest with highest accuracy around 90%	Different geographical data might need other validations too

- Approach Category 1 is in light blue colour represent the number of papers wherein researchers have created own dataset for that demographic conditions.
- Approach category 2 is in orange, which is most widely used by re-searchers where they have used PIMA standard dataset (PIDD that is PIMA Indian Diabetes Database also known as PIMA or PIDD)
- Approach Category 3 is in grey, which collects user’s data from Electronic health records for suggestions and data collected using sensors like smart watches, results but only for type-2 known patients. This is a bulleted list.

3.2 Category-II: Diabetes Prediction & Care

This category describes the use of various AI-ML techniques to identify and classify groups of people who may have known or unknown diabetes for the purposes of a research project or the development of a country's health care support system. The graph below represents the impact of Diabetes on different body organs which leads to damage to other organs like eye retina, different neuron/nerves in brains and at different parts, pulmonary and heart diseases and failure of multiple organs due to high blood glucose for very long time, represented as numbers of researches and research paper commented about this issues on Y-axis and impacts identified on X-axis.

3.3 Category-III: Diabetes’s multifaceted impact

Compared various computer vision algorithms for detecting and identifying diabetes-induced

Retinopathy, focusing on proliferative and non-proliferative conditions. The paper utilized standard

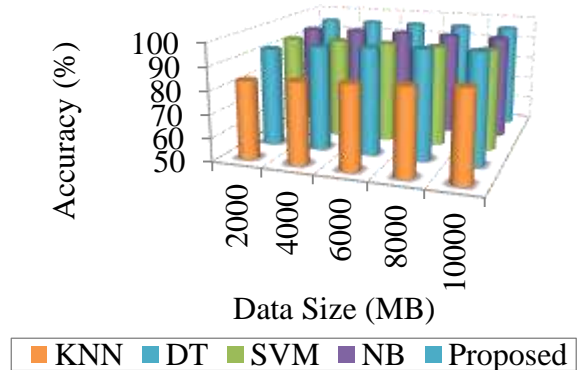


Figure 1. Categories in Diabetes Predictions Approaches wise

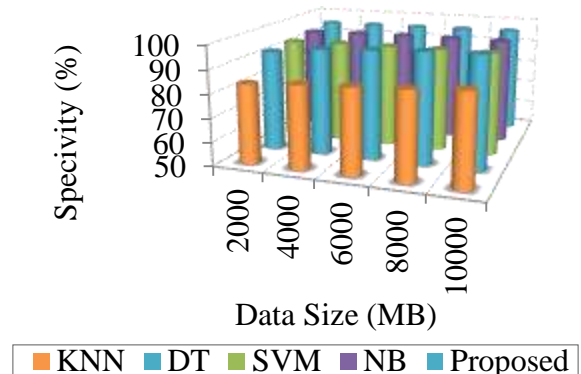


Figure 2. Diabetes’s Multifaceted impact on different body parts

CNN, VGG16, and AlexNet modules, achieving acceptable results with the evaluation of SSIM and FSIM. The study addresses the early detection of proliferative retinopathy, emphasizing the impact of diabetes on the retina. Two datasets were employed, with Retinopathy linked to increased risks of systemic vascular complications. Kumar et al. [10] introduced a study employing GAN-based visualization (Patho-GAN) for encoding pathological descriptors in predicting diabetic retinopathy grades. The research utilized retinal images and unseen binary vessel segmentation, employing machine learning and deep learning techniques. The paper presents explainable deep learning models such as LIME, DeepRED extension of CRED algorithms, ReLu, lekyReLu, and ADAM optimizer to encode pathological descriptors from activated neurons for precise predictions. The research papers summarized overall study shown in a pie chart to illustrate the essence of various studies, emphasizing the focus on known type-2 diabetes patients. The majority employ the PIMA or PIDD dataset, experimenting with AI-ML algorithms to classify and predict patient outcomes. While some studies explore the impact of prolonged diabetes on vital organs, predictive modelling in healthcare emerges as a valuable tool for informed decision-making, improving patient outcomes, and resource optimization. Its applications span disease diagnosis, treatment planning, patient monitoring, public health interventions, and re-source management. Ethical considerations and privacy concerns are crucial when dealing with healthcare data, necessitating transparent, interpretable, and validated models across diverse patient populations before widespread adoption. Predictive modelling in healthcare involves using statistical and machine learning techniques to analyse historical data, developing mathematical models that learn from past patterns to forecast future health scenarios.

4. Analysis

The various studies conducted by researchers with various machine learning algorithms using different datasets have yielded positive results, with some even suggesting recommendations to patients. These studies, summarized in a systematic manner, include details on machine learning algorithms, approach, techniques, and results providing a comprehensive overview of all parameters and outcomes at a glance.

5. Discussion

The type 2 diabetes has effects shown with increased age and other physical factors that are important for

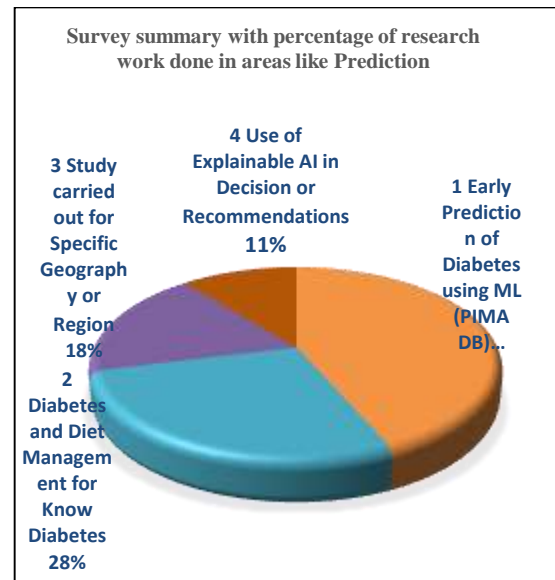


Figure 3. Diabetes's Multifaceted impact on different Survey Summary as per the categories.

prediction and recommendations in diabetes care. However, relying solely on a limited-parameter dataset and applying AI-ML algorithms may not sufficiently predict diabetes risk for unseen inputs. Comprehensive classification justification requires considering additional factors, both internal and external. Various methods, including some screening tests, risk assessment tools (considering age, family history, BMI, physical activity), and machine learning algorithms analyzing extensive data, can identify individuals at high risk of developing diabetes. It's crucial to note that no single test can accurately predict diabetes; a combination of methods allows for early intervention and prevention efforts. Artificial Intelligence is an interesting method and used in different fields in the literature [11-23].

6. Conclusion

The Detailed examination of Machine Learning algorithms and its comparative evaluations of research on diabetes prediction reveals a wealth of AI-ML algorithms and optimization techniques that can effectively predict and classify known type-2 diabetes among numerous widely used datasets. On the other hand, these algorithms can also be tailored to a particular group of people in a restricted demographic who exhibit similar behavioural patterns. All of the study results can be used to forecast and categorize a person as having diabetes or not, but recommendations for food, exercise, and other factors can be made depending on the specific needs of each individual. Most of the work uses the same conventional datasets, which are thus useful for first classifications; some other models do better

by stacking extra information. Diabetes prediction and prevention would take into account the various other elements that contribute to the disease, correlate it with other data, and offer suggestions based on clinical study references and domain expert consultation to derive some guidelines or implications to outcomes.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- [1]Ali, A., & Iqbal, M.M. (2022). A Cost and Energy Efficient Task Scheduling Technique to Offload Microservices based Applications in Mobile Cloud Computing. *IEEE Access*, 10;46633-46651, DOI: 10.1109/ACCESS.2022.3170918
- [2]Castillo, O., Valdez, F., Soria, J., Yoon, J.H., Geem, Z.W., Peraza, C., Ochoa, P., & Amador-Angulo, L. (2020). Optimal Design of Fuzzy Systems Using Differential Evolution and Harmony Search Algorithms with Dynamic Parameter Adaptation. *Applied Sciences*. 10(18), 6146; <https://doi.org/10.3390/app10186146>
- [3]Nguyen, T.A., Min, D., Choi, E., & Lee, J. (2021). Dependability and Security Quantification of an Internet of Medical Things Infrastructure Based on Cloud-Fog-Edge Continuum for Healthcare Monitoring Using Hierarchical Models. *IEEE Internet of Things Journal*, 8(21);15704-15748, doi: 10.1109/JIOT.2021.3081420.
- [4]Kumar, N.G., & J, P.D. (2022). Shielded Identity Based Data and Contour Sharing for Mobile Healthcare Through Cloud Computing. *International Journal for Research in Applied Science and Engineering Technology*. 10(2);1250-1253 DOI: 10.22214/ijraset.2022.40497
- [5]Chinnasamy, P., Albakri, A., Khan, M., Raja, A.A., Kiran, A., & Babu, J.C. (2023). Smart Contract-Enabled Secure Sharing of Health Data for a Mobile Cloud-Based E-Health System. *Applied Sciences*. 13(6);3970; <https://doi.org/10.3390/app13063970>
- [6]Khan, M.A., & Algarni, F. (2020). A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS. *IEEE Access*, 8, 122259-122269. doi: 10.1109/ACCESS.2020.3006424
- [7]Umak, R. (2021). Sharing Healthcare Records in the Cloud Using Attribute-Based Encryption and De-Duplication. *International Journal for Research in Applied Science and Engineering Technology*. 9(VII);3345-3350. DOI: 10.22214/ijraset.2021.37109
- [8]M. N. Bajwa et al., "Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks," *Applied Sciences*, 10(7);2488, doi: 10.3390/app10072488.
- [9]X. Zhang, S. Wang, J. Liu and C. Tao, "Computer-aided diagnosis of four common cutaneous diseases using deep learning algorithm," *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, USA, 2017, pp. 1304-1306, doi: 10.1109/BIBM.2017.8217850.
- [10]Kumar, N.G., & Bhuvana J, P.D. (2022). Survey: Shielded Identity Based Data and Contour Sharing for Mobile Healthcare Through Cloud Computing. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2022.40936>
- [11]M. Husain Bathushaw, & S. Nagasundaram. (2024). The Role of Blockchain and AI in Fortifying Cybersecurity for Healthcare Systems. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1120-1129. <https://doi.org/10.22399/ijcesen.596>
- [12]Rama Lakshmi BOYAPATI, & Radhika YALAVARTHI. (2024). RESNET-53 for Extraction of Alzheimer's Features Using Enhanced Learning Models. *International Journal of Computational and Experimental Science and Engineering*, 10(4);879-889. <https://doi.org/10.22399/ijcesen.519>
- [13]AY, S. (2024). Vehicle Detection And Vehicle Tracking Applications On Traffic Video Surveillance Systems: A systematic literature review. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1059-1068. <https://doi.org/10.22399/ijcesen.629>
- [14]Sheela Margaret D, Elangovan N, Sriram M, & Vedha Balaji. (2024). The Effect of Customer Satisfaction on Use Continuance in Bank Chatbot Service. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1069-1077. <https://doi.org/10.22399/ijcesen.410>
- [15]jaber, khalid, Lafi, M., Alkhatib, A. A., AbedAlghafer, A. K., Abdul Jawad, M., & Ahmad, A. Q. (2024). Comparative Study for Virtual Personal Assistants (VPA) and State-of-the-Art Speech Recognition Technology. *International Journal of Computational and Experimental Science*

- and *Engineering*, 10(3);427-433.
<https://doi.org/10.22399/ijcesen.383>
- [16] Guven, M. (2024). A Comprehensive Review of Large Language Models in Cyber Security. *International Journal of Computational and Experimental Science and Engineering*, 10(3);507-516. <https://doi.org/10.22399/ijcesen.469>
- [17] M, V., V, J., K, A., Kalakoti, G., & Nithila, E. (2024). Explainable AI for Transparent MRI Segmentation: Deep Learning and Visual Attribution in Clinical Decision Support. *International Journal of Computational and Experimental Science and Engineering*, 10(4);575-584. <https://doi.org/10.22399/ijcesen.479>
- [18] ÖZNAÇAR, T., & ERGENE, N. (2024). A Machine Learning Approach to Early Detection and Malignancy Prediction in Breast Cancer. *International Journal of Computational and Experimental Science and Engineering*, 10(4)911-917. <https://doi.org/10.22399/ijcesen.516>
- [19] Türkmen, G., Sezen, A., & Şengül, G. (2024). Comparative Analysis of Programming Languages Utilized in Artificial Intelligence Applications: Features, Performance, and Suitability. *International Journal of Computational and Experimental Science and Engineering*, 10(3);461-469. <https://doi.org/10.22399/ijcesen.342>
- [20] Venkatraman Umbalacheri Ramasamy. (2024). Overview of Anomaly Detection Techniques across Different Domains: A Systematic Review. *International Journal of Computational and Experimental Science and Engineering*, 10(4);898-910. <https://doi.org/10.22399/ijcesen.522>
- [21] Jafar Ismail, R., Samar Jaafar Ismael, Dr. Sara Raouf Muhamad Amin, Wassan Adnan Hashim, & Israa Tahseen Ali. (2024). Survey of Multiple Destination Route Discovery Protocols. *International Journal of Computational and Experimental Science and Engineering*, 10(3);420-426. <https://doi.org/10.22399/ijcesen.385>
- [22] guven, mesut. (2024). Dynamic Malware Analysis Using a Sandbox Environment, Network Traffic Logs, and Artificial Intelligence. *International Journal of Computational and Experimental Science and Engineering*, 10(3);480-490. <https://doi.org/10.22399/ijcesen.460>
- [23] Serap ÇATLI DİNÇ, AKMANSU, M., BORA, H., ÜÇGÜL, A., ÇETİN, B. E., ERPOLAT, P., ... ŞENTÜRK, E. (2024). Evaluation of a Clinical Acceptability of Deep Learning-Based Autocontouring: An Example of The Use of Artificial Intelligence in Prostate Radiotherapy. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1181-1186. <https://doi.org/10.22399/ijcesen.386>