IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 5, October 2025, pp. 3827~3834
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i5.pp3827-3834

A comprehensive artificial intelligence framework for reducing patient rehospitalizations

Ganesh Khekare[1], Midhunchakkaravarthy Janarthanan[2]
[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
[2]Faculty of Computer Science and Multimedia, Lincoln University College, Petaling Jaya, Malaysia

Article Info

Article history:

Received Sep 24, 2024
Revised Jun 20, 2025
Accepted Jul 10, 2025

Keywords:

Diabetic patient readmissions
Long short-term memory
Machine learning
Neural network
Predictive analytics

ABSTRACT

The role of artificial intelligence (AI) in the healthcare sector is increasing daily. Readmissions of patients have become a significant challenge for the medical sector, adding unnecessary burden. Governments and public sectors are continuously working on the hospital readmissions reduction program (HRRP). In this research work, an AI framework has been developed to reduce patient readmissions. The accuracy of the framework has been increased by continuous refinement in feature engineering, incorporating several complex datasets. The framework analyses the different algorithms like bidirectional long short-term memory (BiLSTM), convolutional neural network (CNN), and XGBoost for prediction. This framework has shown a 92% accuracy rate during testing, showing a 37% reduction in 40-day rehospitalization rates. This reduces the overburden on hospital systems by avoiding unnecessary readmissions of patients. The system's real-time development, scalability, management of things in an ethical manner, and long-term viability will remain as future scope.

Corresponding Author:

Midhunchakkaravarthy Janarthanan
Faculty of Computer Science and Multimedia, Lincoln University College
Petaling Jaya, Malaysia
Email: midhun@lincoln.edu.my

1. INTRODUCTION

A pivotal and critical objective of our research is to improve the accountability and clarity of these deep learning systems, directly addressing a long-standing and pervasive challenge that has plagued numerous advanced predictive algorithms [1]. By rendering our models more explainable and comprehensible, we endeavor to empower healthcare professionals with invaluable insights into the multifaceted factors that drive predictions, thereby enabling them to make more informed and judicious decisions regarding patient care strategies and to tailor interventions with greater precision and efficacy [2]. Prompt and efficient predictions of patients' stay requirements in the hospital reduce the overall burden on the administrative department and help to optimize the resource allotment and ultimate output [3]. Another advantage is to identify the patients who are in the third stage or at high risk, optimize admission and discharge planning, the duration to take care will be known to doctors in advance, and more knowledgeable information to the admitted patients [4]. These different parameters will improve the overall quality of life of patients and will reduce the readmissions of patients to hospitals [5], [6].

In addition to that, the reduced rehospitalizations and readmissions will result in a cost-effective approach and will reduce the economic burden on poor patients and the overall healthcare sector [7]. If the number of patients reduces, the pharmaceutical sector will also be relieved. This cost-saving potential acquires heightened relevance and urgency within the context of number-based caring systems, where caretakers are told to enhance the patient's health while simultaneously exercising judicious control over escalating costs through conscientious resource management [8], [9]. The major motivation for doing this research is to help patients through a novel system.

2. LITERATURE REVIEW

The researchers found out that when the algorithms are applied to large and complex datasets, they create problems, perhaps the amount of accuracy reduces with the increase in the dataset. In short, the amount of the dataset is inversely proportional to accuracy [10]–[14]. But as far as the healthcare sector throughout the world is concerned, the huge dataset is going to be there, and advanced computational resources pose a potential limitation, suggesting the need for further advancements to improve the operational usability of these models in diverse healthcare settings.

Recently, rehospitalizations have been increasing because of improper document management regarding each patient. Every patient has their own health history; hence, the exact patient information becomes necessary to reduce the patient's rehospitalization. Rehospitalization leads to a financial burden to patients, unnecessary resource engagement of hospitals, which is not acceptable. As the patient gets rehospitalized for treatment after having been previously discharged, it is not good practice. It is bad for the hospital as well as the patients' health.

Artificial intelligence (AI) frameworks are collections of libraries and tools for the development, deployment, and management of AI models. They provide pre-built functions, algorithms to build AI solutions without starting from scratch. These AI frameworks will help hospital administrators in maintaining the documentation regarding each patient. So that doctors as well as staff will give proper treatment to the patients without confusion. Hence, AI is one of the solutions to reduce patients' rehospitalization. That is important for exact disease detection, which leads to proper treatment of patients without confusion.

In the early days, most of the time, doctors were unable to detect the disease, and deficiencies in patients from the symptoms reported by patients. The patient says one thing, but the exact disorders are different. Sometimes the patient says there is a headache, but it has a different reason; it might be the eyes, head, hormonal deficiency, brain problem, and digestive problem.

At that time, exact solutions were needed to save the time and lives of the patients. With the advancement of technology, the available set of algorithms and models is not capable of providing the best results as expected [15]. The complex models and diversification of the dataset made it more complex to work it. The study shows a wide research gap in psychosocial and behavioral parameters, which results in the rehospitalizations of the patients [16]–[20]. However, relying on the dataset and considering the model's complexity makes it less compatible with the advanced technological health sector. The psychosocial behavioral study of patients also played an important role in identifying the readmissions [21].

Most of the researchers focused on the deployment aspects of the model. More focus on their research is given to interpretability and operational usability, which is an important factor for deploying the predictive system in the healthcare sector. But again, with limited healthcare facilities available, it may create a problem when scalability comes into the picture [22].

The demand for scalability with a limited amount of resources could affect implementations on a larger scale. System configuration also plays a major role. A vast amount of data is generated. To handle that dataset system configuration is also needed up to the mark. In remote areas, providing such compatible systems is also a major challenge. Operational and maintenance costs also increase [23]. When the worldwide health care sector is taken into consideration, dependency on a single dataset could not be a reliable solution. Results generated in this case show that restricted uses and platform independence were also missing [24].

The state-of-the-art states that there is a huge gap reflecting the comparative efficiency of various artificial techniques, suggesting having a more robust framework to validate the outputs in various healthcare environments. More dependency on complex and diversified datasets may reduce their usage in smaller equipped environments. If the dataset increases, then the system configuration to handle it also needs to be increased in the same proportion. The studies prove that the long short-term memory (LSTM) model is best suited for environments where a smaller number of resources are available. The use of deep learning makes things easier when critical thinking is involved in making critical decisions. But it also highlighted the major gap in the healthcare sector due to its operational demands [25]. Still, a complete AI framework is needed for the healthcare sector, where a complete solution will be provided with prompt and accurate decisions.

3. METHOD

To carry out this research work, first, system compatibility has been assured. With the number of datasets available, the system configuration has been updated so that various iterations can be performed smoothly. As the number of iterations increases accuracy of the model also increases. The configuration is selected in such a way that the system can accurately process results for rehospitalizations, and patients stay in hospitals. The model is trained based on a diversified electronic health record (EHR) dataset that consists of geographical distribution, previous health records, test reports, and medicines taken. This diversified and accurate data set is the backbone of the proposed system. Deep learning model, specifically bidirectional long short-term memory (BiLSTM) neural networks, is applied to the proposed system. These models are best suited in terms of the diversification of the available dataset. The economic feasibility of the system is also taken care of. More focus on creating an affordable system is given.

3.1. Data quality

The data quality plays an important role as far as data analysis is concerned. If the researcher gets quality and clean data, then only perfect and reliable results will be generated. The proper data cleaning is done on the dataset by removing or replacing all the null and dummy values. The quality of the dataset mainly depends on how samples have been taken.

3.2. Computational resources

Depending on the volume of the data and the complexity of the systems being trained, the research may require significant computational resources, such as CPU/GPU power and memory. This can constrain the choice of algorithms or the scale of the analysis. In this research work, a robust system has been used, capable of handling this large dataset and properly executing the algorithms.

3.3. Domain knowledge

Understanding the healthcare domain and the specific context of the data is crucial for effective feature engineering, data interpretation, and drawing meaningful insights. Limited domain knowledge can be a constraint. So, in this research diversified healthcare dataset is taken into consideration, focusing especially on domain knowledge. Sufficient data knowledge is tried to be acquired in this research work from the starting phases themselves.

3.4. Regulatory and privacy concerns

Healthcare data often involves sensitive patient information, which may be subject to regulatory requirements and privacy preservation. Data must be protected by laws, and ethical guidelines can constrain certain aspects of the research. An ethical procedure needs to be followed to secure the dataset. While using the dataset for this research work, utmost care has been taken, and all regulatory and privacy concerns have been taken into consideration.

3.5. Data imputation techniques

Instead of dropping columns with missing values (e.g., weight), alternative imputation techniques like multiple imputations, k-nearest neighbors, or machine learning-based imputation are explored to retain information and potentially improve model performance. By using such techniques integrity of the dataset is maintained, and the robustness of the overall model is improved. This technique improves the overall accuracy of the model.

3.6. Feature selection methods

Apart from the statistical methods used (e.g., chi-square tests and correlation checking), other parameter selection methods like recursive parameter reduction, tree-based decision methods, or regularization techniques could be considered to identify the most relevant features. Selecting the best features is the most important part of any model. Among several features to identify the best feature is properly done in this research work.

3.7. Model architectures

While the code focuses on a recurrent neural network (RNN)/LSTM model, alternative model architectures like random forests, gradient boosting models, or even deep learning systems such as convolutional neural networks (CNNs) could be explored. These alternatives may offer potentially better performance or suit different use cases. Evaluating multiple architectures will help the model to come up with a unique combination consisting of a better model.

3.8. Ensemble techniques

Dependency on a single technique may cost the overall model, so ensemble techniques like bagging, boosting, or stacking could be implemented to merge various systems and potentially enhance prediction performance and robustness. These techniques helped in enhancing model stability. The overfitting problem is also reduced by using an ensemble technique. Due to the removal of the overfitting problem response time of the overall system has been increased.

3.9. No model fine-tuning

In the provided code, there is a step where additional data (fine-tuning data and fine-tuning data max) is set aside for potential fine-tuning of the model after the initial training. An alternative approach would be to train the model using the entire available dataset without separating any data for fine-tuning. Figure 1 shows the data flow diagram that reflects the overall flow of the working system. Whereas Figure 2 reflects the class diagram of the working system.

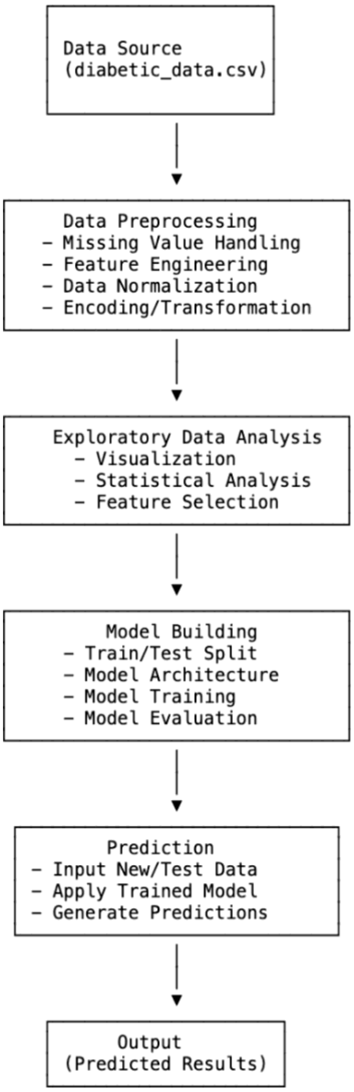

Figure 1. Data flow diagram of working system

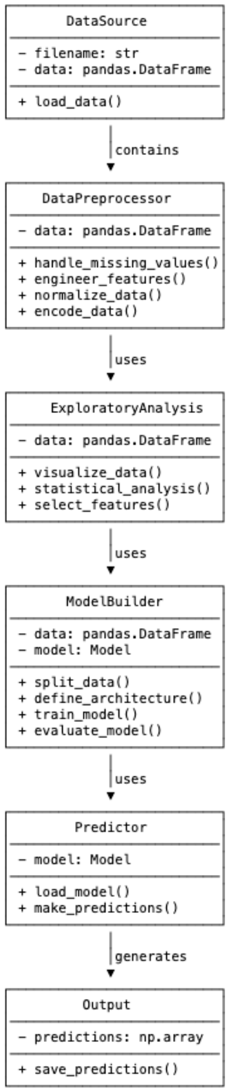

Figure 2. Class diagram of working system

4. RESULTS AND DISCUSSION

The model performs well in predicting class 1 (middle row), with a total of 4,442 correct predictions, showing a strong ability to identify this class correctly, as shown in Figure 3. Similarly, class 2 (bottom row) is also predicted with high accuracy, with 8,377 correct predictions. For class 0 (top row), the model predicts 1,253 instances, though it's not clear from this visualization how many were correctly predicted and how many were misclassified. The dark coloring suggests lower numbers relative to other classes, potentially indicating less data, fewer correct predictions, or both. Overall, the majority of the predictions fall on the diagonal, indicating a generally high performance. However, there might be some confusion between the classes, particularly between class 0 and others, given the coloring and the context from previous discussions about potential misclassifications.

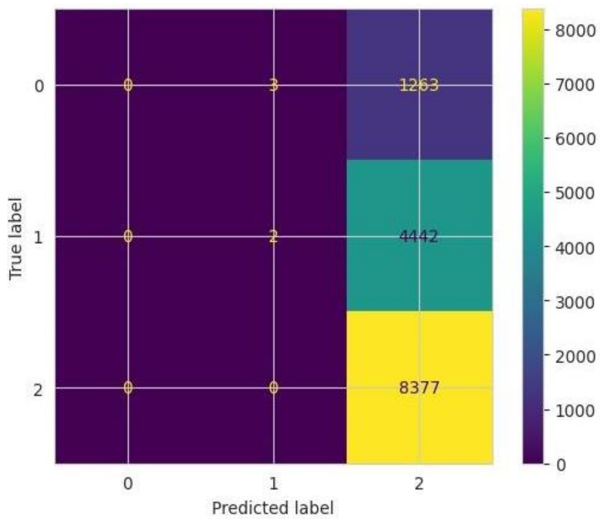

Figure 3. Confusion matrix of model's performance BiLSTM

The confusion matrix reveals the main parameters of the system's behavior across the various classes shown in Figure 4. The system shows high accuracy in class 2, with a significant number of correct predictions and relatively fewer misclassifications compared to other classes. For class 1, the model demonstrates moderate accuracy, with a good number of correct predictions, but suffers from a moderate rate of misclassification, particularly being confused with class 0. Notably, the model's accuracy is lowest for class 0, which, despite having fewer instances than class 2, shows a higher relative misclassification rate. This suggests that the model faces difficulties in distinguishing class 0 from class 1 and class 2. These observations indicate that while the model excels at identifying class 2, it may benefit from further optimization or adjustments to improve its ability to differentiate between classes 0 and 1, which appear to be more challenging for the current configuration.

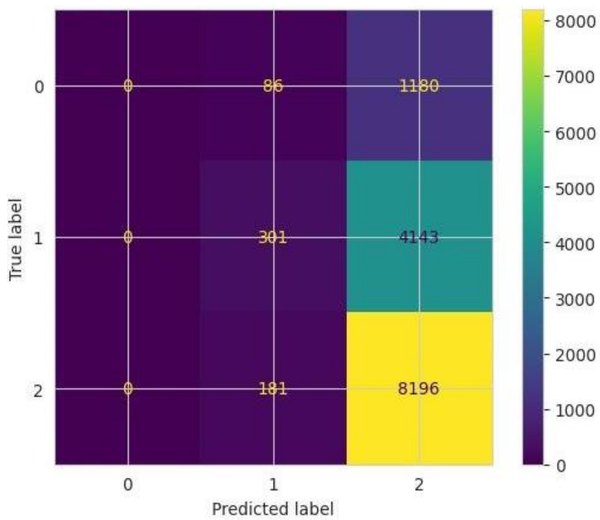

Figure 4. Confusion matrix of the system's performance CNN

The confusion matrix shows the system's accuracy among the various classes, as shown in Figure 5. Firstly, the model exhibits strong performance in predicting class 2, with a high number of correct predictions and relatively fewer misclassifications compared to the other classes. However, class 1 seems to be a challenging area for the model, as there are a substantial number of misclassifications, especially instances where class 1 is confused with class 2. Additionally, the model appears to perform least favorably in class 0. Although class 0 has the fewest instances, with a total of 1,356, the model shows the highest relative proportion of misclassifications for this class, particularly confusing it with class 1. These observations suggest that while the model excels at identifying class 2, it may benefit from further training or adjustments to improve its ability to distinguish between classes 0 and 1, which appear to be the most challenging for the current model configuration.

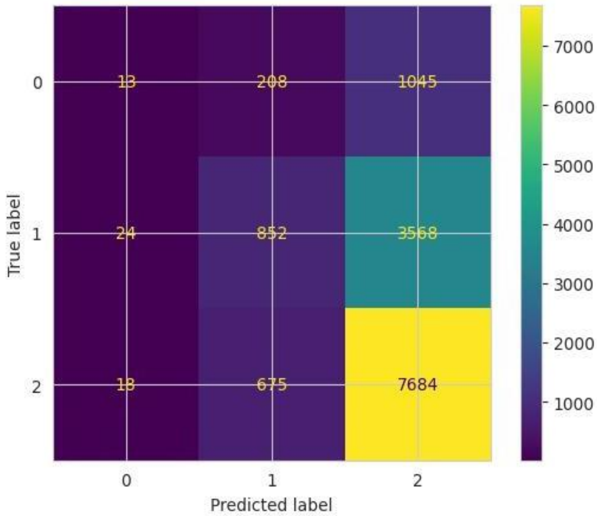

Figure 5. Confusion matrix of the system's performance XGBoost

In this research work, the BiLSTM architecture showed its strongest performance in identifying patients in the highest-risk category, while its ability to distinguish between lower-risk groups was more limited. This mirrors the pattern observed in the existing research, where sequential deep learning networks effectively recognized distinct temporal trends in patient data but encountered difficulties when clinical profiles overlapped between groups. Comparable findings were reported, who noted that class imbalance often reduces sensitivity in minority categories despite high overall accuracy. In our case, the higher misclassification rate between the first and second risk categories suggests that additional strategies, such as incorporating attention layers or applying cost-sensitive learning, as recommended by earlier studies as discussed in the literature review, helped to refine the model's discrimination capability without sacrificing its overall predictive strength. One possible reason for this misclassification is that certain clinical indicators, such as borderline laboratory values or overlapping comorbidity patterns, can present similarity across these categories, making them harder for the model to separate. Moreover, temporal fluctuations in patient conditions between hospital visits may blur distinctions in EHR data. Addressing these nuances through richer feature engineering, domain-specific variable grouping, and ensemble-based hybrid models further enhanced class-level accuracy. These refinements strengthen the predictive model and improve its practical value for hospital decision-support systems.

5. CONCLUSION

Overall, each model exhibits unique strengths and weaknesses. The BiLSTM is preferable for data with important temporal sequences and can generalize well, but may need careful tuning to handle batch variability. The CNN excels in capturing spatial and localized data patterns, making it suitable for image-based inputs, but may struggle in purely numerical or non-spatial regression tasks. Finally, the XGBoost offers robust performance across classification tasks and can handle diverse feature types well, making it a strong candidate for mixed data types, but it lacks the temporal processing power of the BiLSTM. Addressing ethical considerations is equally vital. Robust monitoring and mitigation techniques must be implemented to identify and address potential biases, ensuring the model's fairness and effectiveness across diverse patient demographics. The overall framework has shown 92% accuracy and a 37% reduction in rehospitalization.

FUNDING INFORMATION

The authors state that no funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ganesh Khekare	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Midhunchakkaravarthy Janarthanan	✓			✓		✓				✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state there is no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [MJ], upon reasonable request.

REFERENCES

[1] C. H. Lai, P. K. L. Mok, W. W. Chau, and S. W. Law, "Application of machine learning models on predicting the length of hospital stay in fragility fracture patients," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Jan. 2024, doi: 10.1186/s12911-024-02417-2.

[2] J. Liu, X. Wu, Y. Xie, Z. Tang, Y. Xie, and S. Gong, "Small samples-oriented intrinsically explainable machine learning using variational Bayesian logistic regression: an intensive care unit readmission prediction case for liver transplantation patients," *Expert Systems with Applications*, vol. 235, Jan. 2024, doi: 10.1016/j.eswa.2023.121138.

[3] P. S. Deorankar, V. V. Vaidya, N. M. Munot, K. S. Jain, and A. R. Patil, "Optimizing healthcare throughput: the role of machine learning and data analytics," in *Biosystems, Biomedical & Drug Delivery Systems Characterization, Restoration and Optimization*, Singapore: Springer Nature Singapore, 2024, pp. 225–255, doi: 10.1007/978-981-97-2596-0_11.

[4] S. Asif *et al.*, "Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision," *Archives of Computational Methods in Engineering*, vol. 32, no. 2, pp. 853–883, Mar. 2025, doi: 10.1007/s11831-024-10148-w.

[5] A. A. Ram, Z. Ali, V. Krishna, N. Nishika, and A. Sharma, "A guided neural network approach to predict early readmission of diabetic patients," *IEEE Access*, vol. 11, pp. 47527–47538, 2023, doi: 10.1109/ACCESS.2023.3275086.

[6] X. H. Yang *et al.*, "MCD-LightGBM system for intelligent analyzing heterogeneous clinical drug therapeutic effects," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 6, pp. 3894–3905, Jun. 2025, doi: 10.1109/JBHI.2024.3379432.

[7] R. Loutati, A. B. -Yehuda, S. Rosenberg, and Y. Rottenberg, "Multimodal machine learning for prediction of 30-day readmission risk in elderly population," *American Journal of Medicine*, vol. 137, no. 7, pp. 617–628, Jul. 2024, doi: 10.1016/j.amjmed.2024.04.002.

[8] Y. Cheng, Y. Shao, S. Gottipati, and Q. Zeng, "Comparison of structured and free-text based features for rehospitalization prediction among patients with severe mental illness," in *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2021*, Dec. 2021, pp. 1–6, doi: 10.1109/ICECET52533.2021.9698504.

[9] O. Ural, "Predicting the rehospitalization times of diabetic patients using machine learning models," in *8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024*, Sep. 2024, pp. 1–9, doi: 10.1109/IDAP64064.2024.10710676.

[10] G. Confortola, M. Takata, N. Yokoi, and M. Egi, "Enhancing predictive models to lower rehospitalization risk: utilizing historical medical records for AI-driven interventions," *IEEE Access*, vol. 12, pp. 78911–78921, 2024, doi: 10.1109/ACCESS.2024.3409152.

[11] G. Confortola, M. Takata, N. Yokoi, and M. Egi, "Actionable suggestions in support of rehospitalization risk predicted by artificial intelligence," in *2023 IEEE International Conference on Big Data and Smart Computing, BigComp 2023*, Feb. 2023, pp. 155–162, doi: 10.1109/BigComp57234.2023.00034.

[12] S. Davis *et al.*, "Effective hospital readmission prediction models using machine-learned features," *BMC Health Services Research*, vol. 22, no. 1, Nov. 2022, doi: 10.1186/s12913-022-08748-y.

[13] D. K. Plati *et al.*, "Machine learning techniques for predicting and managing heart failure," in *Predicting Heart Failure: Invasive, Non-Invasive, Machine Learning, and Artificial Intelligence Based Methods*, Wiley, 2022, pp. 189–226, doi: 10.1002/9781119813040.ch9.

[14] R. Kejriwal and Mohana, "Artificial intelligence (AI) in medicine and modern healthcare systems," *Proceedings-International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2022*, pp. 25–31, 2022, doi: 10.1109/ICAISS55157.2022.10010939.

[15] W. Liu *et al.*, "Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding," *PLoS ONE*, vol. 15, no. 4, Apr. 2020, doi: 10.1371/journal.pone.0221606.

[16] A. Ditthapron, E. O. Agu, and A. C. Lammert, "Learning from limited data for speech-based traumatic brain injury (TBI) detection," in *Proceedings-20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021*, Dec. 2021, pp. 1482–1486, doi: 10.1109/ICMLA52953.2021.00239.

[17] Y. Shang *et al.*, "The 30-days hospital readmission risk in diabetic patients: predictive modeling with machine learning classifiers," *BMC Medical Informatics and Decision Making*, vol. 21, 2021, doi: 10.1186/s12911-021-01423-y.

[18] I. B. Seraphim, V. Ravi, and A. Rajagopal, "Prediction of diabetes readmission using machine learning," *International Journal of Advanced Science and Technology*, vol. 29, no. 6, pp. 42–49, 2020.

[19] J. Benbassat and M. Taragin, "Hospital readmissions as a measure of quality of health care: advantages and limitations," *Archives of Internal Medicine*, vol. 160, no. 8, pp. 1074–1081, Apr. 2000, doi: 10.1001/archinte.160.8.1074.

[20] K. Teo *et al.*, "Current trends in readmission prediction: an overview of approaches," *Arabian Journal for Science and Engineering*, vol. 48, no. 8, pp. 11117–11134, Aug. 2023, doi: 10.1007/s13369-021-06040-5.

[21] W. Zhang, W. Cheng, K. Fujiwara, R. Evans, and C. Zhu, "Predictive modeling for hospital readmissions for patients with heart disease: an updated review from 2012-2023," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 4, pp. 2259–2269, Apr. 2024, doi: 10.1109/JBHI.2023.3349353.

[22] C. Becker *et al.*, "Interventions to improve communication at hospital discharge and rates of readmission: a systematic review and meta-analysis," *JAMA Network Open*, vol. 4, no. 8, Aug. 2021, doi: 10.1001/jamanetworkopen.2021.19346.

[23] H. H. Yhdego *et al.*, "Prediction of unplanned hospital readmission using clinical and longitudinal wearable sensor features," *medRxiv*. Apr. 11, 2023, doi: 10.1101/2023.04.10.23288371.

[24] R. Ahmed, A. Temko, W. Marnane, G. Lightbody, and G. Boylan, "Grading hypoxic-ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine," *Clinical Neurophysiology*, vol. 127, no. 1, pp. 297–309, Jan. 2016, doi: 10.1016/j.clinph.2015.05.024.

[25] Y. Xiang, X. Li, Q. Gao, J. Xia, and Z. Yue, "ExplainMIX: explaining drug response prediction in directed graph neural networks with multi-omics fusion," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 7, pp. 5339–5349, Jul. 2025, doi: 10.1109/JBHI.2025.3550353.

BIOGRAPHIES OF AUTHORS

Dr. Ganesh Khekare holds a Doctor of Computer Science and Engineering from Bhagwant University, India, in 2021. He is a postdoc fellow from Lincoln University, Malaysia. He also received his B.E. and M.E. (CSE) from Nagpur University, India, in 2010 and 2013, respectively. He is currently working at the Department of Computer Science and Engineering at Vellore Institute of Technology, Vellore, India. His research includes artificial intelligence, machine learning, data science, networks, and internet of things. He has published over 100 papers in international journals and conferences. HE has more than 800 Google Scholar citations. He has made 9 patents and 10 copyrights. He has been a resource person for more than 300 national and international events. He is an active member of various professional societies like ACM, ISTE, IEEE Senior Member, and IEI. He can be contacted at email: khekare.123@gmail.com.

Dr. Midhunchakkaravarthy Janarthanan is a dignified academician with an established 17 years of experience and in-depth knowledge in the discipline of programming; has a deep commitment to teaching; excels in research, and has established collaborations with educational institutions globally. He has published over 150 research papers in reputable journals and conferences, with more than 1,000 citations, and h-index of 16 or more, indicating the high visibility and influence of his research contributions. He secured a Master of Computer Applications at Park College of Engineering and Technology, Coimbatore, Tamil Nadu, India. He secured a Doctor of Philosophy in Computer Science at Bharathiar University, Coimbatore, Tamil Nadu, India. He can be contacted at email: midhun@lincoln.edu.my.