Journal of Scientific Insight

Vol. 1 (1), 2024
Page: 26-32

Transforming the Diabetes Mellitus Diagnosis and Treatment Using Data Technology: Comprehensive Analysis of Deep Learning and Machine Learning Methodologies

Dwi Anggriani [1], Syaiful Bachri Mustamin [1,*], Sahriani [1], Muhammad Atnang [1], Siti Fatmah [1], Nur Azaliah Mar [1], Nurhikmah Fajar [2]

[1] Departement of Information Technology, Faculty of Science Technology and Health, Institut Sains Teknologi dan Kesehatan 'Aisyiyah Kendari, Indonesia
[2] Departement of Computer Engineering, Faculty of Engineering, Institut Teknologi dan Sains Muhammadiyah Kolata Utara, Indonesia

*Email (corresponding author): sbm@istekaisyiyah.ac.id

Abstract

Recent research in health data analysis has transformed our understanding, prediction, and management of diabetes mellitus. This review explores various approaches used in related studies to enhance understanding and management strategies of diabetes through data analysis. Various data analysis methods, including machine learning such as neural networks, Gaussian Process Classification (GPC), and deep learning, have been used to enhance illness management and forecast accuracy. One of the included studies created customised care plans and used data to forecast the likelihood of complications in diabetes.. Another focused on comparative approaches for diabetes diagnosis using artificial intelligence, while others explored disease classification techniques using GPC algorithms. On the other hand, some studies utilized deep learning to identify diverse trajectories of type 2 diabetes from routine medical records, while others developed wide and deep learning models to predict diabetes onset. This review notes that data analysis approaches have significantly advanced accuracy in diagnosis, predictive modeling, and disease management of diabetes. Integrating these technologies allows for more personalized treatment approaches, where patient data can tailor individualized care strategies. Study findings indicate that machine learning and deep learning applications not only enhance prediction accuracy but also unlock new potentials in identifying risk factors, managing complications, and preventing diseases. Thus, this review provides profound insights into how data analysis has shifted paradigms in diabetes management, extending beyond diagnosis and treatment to encompass prevention and long-term management of chronic diseases. These studies lay a robust foundation for further research in developing more sophisticated and effective approaches in health data analysis, ultimately aiming to enhance the overall quality of life for patients with diabetes.

Keywords: Diabetes mellitus, health data analysis, machine learning, neural networks, GPC classification

1. Introduction

Diabetes type I, particularly type 2, is a chronic disease posing a global health challenge with increasing prevalence (World Health Organization, 2020). It is often associated with serious complications such as retinopathy, nephropathy, and heightened risks of heart attack, stroke, and lower extremity amputation (American Diabetes Association, 2021). Early diagnosis and proper management are therefore crucial.

Numerous studies have concentrated on creating methods for classifying and predicting diabetes mellitus using machine learning, deep learning, and electronic health data (Anwar et al., 2019; Maniruzzaman et al., 2017; Nguyen et al., 2019). These techniques aim to enhance diagnostic accuracy and prediction compared to traditional methods. One highlighted contribution is the use of wide and deep learning models, combining the strengths of general linear models and artificial neural networks to predict the onset of type 2 diabetes (Nguyen et al., 2019). Furthermore, efforts have been made to identify subgroups or clusters of diabetes progression using deep learning algorithms and clustering on longitudinal electronic health record data (Manzini et al., 2022). This aims to understand the heterogeneity of diabetes and provide insights into disease evolution, facilitating more precise and patient-centered management.

Overall, these studies aim to address gaps in managing chronic diseases like diabetes by leveraging advanced data analysis techniques (Mustamin et al., 2024). The results are anticipated to improve the quality of life for diabetic patients for patients, healthcare providers, and healthcare systems as a whole (Mustamin et al., 2024; Kavakiotis et al., 2017)..

2. Materials and Methods

Based on the literature review discussed, diabetes mellitus, especially type 2, is a chronic disease increasingly becoming a global health issue with rising prevalence (Chen et al., 2020). This disease is often associated with serious complications such as retinopathy, nephropathy, heart attacks, strokes, and lower extremity amputations, emphasizing the critical importance of early diagnosis and proper management (American Diabetes Association, 2021). Consequently, a great deal of research has gone into creating deep learning and machine learning-based methods for controlling, predicting, and classifying diabetes using electronic medical information.

Using Gaussian Process Classification (GPC) approaches to categorise diabetes mellitus data is one strategy that has been highlighted. Its performance is compared with other classification methods, such as LDA, QDA, and NB (Maniruzzaman et al., 2017). Key findings indicate that GPC-based models achieve higher accuracy, sensitivity, specificity, and other performance metrics compared to other methods in diabetes data classification (Maniruzzaman et al., 2017). Additionally, efforts have been made to identify different subgroups or clusters of diabetes progression using deep learning algorithms and clustering on longitudinal electronic health record data (Manzini et al., 2022). This aims to understand the heterogeneity of diabetes and provide insights into disease evolution, aiding more precise and patient-centered management.

The wide and deep learning method is presented to combine feed-forward neural networks with generic linear models with several features to improve prediction accuracy in the context of forecasting the onset of type 2 diabetes (Nguyen et al., 2019). This approach is expected to address issues of imbalanced data and improve prediction accuracy compared to traditional methods (Nguyen et al., 2019). Overall, these studies aim to fill gaps in managing chronic diseases like diabetes by leveraging advanced data analysis techniques, as proposed by (Mustamin et al., 2024) in understanding patterns, predicting complication risks, and designing patient-centered disease management solutions.

The results are expected to improve the quality of life for individuals with diabetes, as well as healthcare providers and systems as a whole(Mustamin et al., 2024; Kavakiotis et al.,

2017). However, there are still some limitations and new research questions emerging, such as the effectiveness of recommended disease management solutions based on predictive models in clinical practice, and identifying factors that can strengthen predictive model capabilities in predicting diabetes complication risks (Mustamin et al., 2024).

Table 1. Research on predicting diabetes complication risks conducted and reported

Research Theme	Methods Used	Research Findings	References
Classification of diabetes mellitus data	Linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Gaussian Process Classification (GPC), and Naive Bayes (NB)	Compared to alternative approaches, the GPC-based model offers improved performance criteria such as sensitivity, specificity, accuracy, and other metrics..	Maniruzzaman et al., 2017
Clustering the development trajectories of Type 2 Diabetes Mellitus (T2DM)	Deep learning (Kernelized-AutoEncoder) and longitudinal clustering	Identifies various longitudinal trajectories of T2DM using routine electronic medical records.	Manzini et al., 2022
Prediction of Type 2 diabetes onset	Deep and broad learning (combining feedforward neural networks with general linear models)	Improves prediction accuracy of Type 2 diabetes onset using electronic medical records compared to traditional methods.	Nguyen et al., 2019
Data analysis for pattern understanding, risk prediction, and diabetes management solutions	PCA, SMOTE, Random Forest, SVM, Neural Network	The neural network model predicts the likelihood of diabetic complications with an AUC-ROC of 0.92.	Mustamin et al., 2023
Comparative analysis of diabetes mellitus diagnosis	Deep learning, machine learning, additional features (heart rate)	Deep learning and combined machine learning algorithms yield better results in diabetes diagnosis.	Anwar et al., 2020

Table 1 presented illustrates various recent studies in health data analysis, particularly focused on the management of diabetes mellitus. Each study has a unique research objective, such as ""Health Data Analysis for Comprehensive Understanding of Patterns, Prediction, and Disease Management: A Case Study on Diabetes Mellitus," which makes use of methods such as Synthetic Minority Over-sampling Technique (SMOTE), Principal Component Analysis

(PCA), as well as predictive models such as Random Forest and Neural Network to understand health patterns and predict diabetes complication risks. Additionally, The comparative accuracy of different methods utilising machine learning and deep learning for diabetes diagnosis is covered in the research "Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm". In the meantime, deep learning is used in the study "Longitudinal deep learning clustering of Type 2 Diabetes Mellitus trajectories using routinely collected health records" to group Type 2 diabetes development trajectories from routine medical records., On the other hand, "Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records" focuses on utilising deep learning models from electronic health records to predict the development of type 2 diabetes.. The provided references each support these research outcomes with complete citations related to the methods and main findings achieved in each literature review.

3. Results and Discussion

3.1. Utilising Deep Learning and Machine Learning to Forecast Diabetes Mellitus

Studies focusing on diabetes mellitus data analysis showcase diverse approaches, all aiming to enhance understanding and management of the disease through information technology and contemporary scientific methods. Mustamin et al. (2023) conducted a case study using direct data from health information systems, patient interviews, and open datasets to develop predictive models. They used machine learning methods including Random Forest, Support Vector Machine, and Neural Network for additional analysis after implementing PCA for feature extraction.

Anwar et al. (2020) complemented this approach with a literature review encompassing various methods in classifying diabetes mellitus using AI. Although their focus was more on the literature, this study provides a comprehensive overview of the technological evolution in diabetes diagnosis. Additionally, Maniruzzaman et al. (2017) added a comparative aspect by utilizing the Pima Indian Diabetes dataset, comparing several classification algorithms to improve accuracy and effectiveness in disease management. Manzini et al. (2022) presented a more advanced approach by developing clustering methods using deep learning from routinely collected electronic medical records. They demonstrated that integrating Kernelized-AutoEncoder techniques for latent space analysis and classical clustering algorithms could uncover complex patterns in the trajectories of type 2 diabetes.

Nguyen et al. (2019) closed the loop by employing extensive and deep learning from electronic health information to forecast the onset of type 2 diabetes. By combining linear regression models and artificial neural networks, this study highlighted the potential of technology to support early diagnosis and timely intervention. Overall, these studies reflect the evolution in scientific approaches to diabetes mellitus, from in-depth data analysis to the application of cutting-edge technology in disease prediction and management. The integration of information technology and medical science enriches our understanding of diabetes mellitus, paving the way for the development of more effective diagnostic and intervention strategies in the future.

Table 2. Methods used for predicting diabetes mellitus

Research	Method	Object of Study	Research Findings
A Case Study on Diabetes Mellitus Using Health Data Analysis to Gain a Comprehensive Understanding of Patterns, Predictions, and Disease Management (Mustamin et al., 2023)	Direct data extraction, patient interviews, PCA, Random Forest, SVM, Neural Network	Diabetes Mellitus	Patterns, prediction, and management of diabetes mellitus
Comparative methods for diabetes mellitus data classification: the machine learning paradigm (Anwar et al., 2020)	Literature review, comparison of classification algorithms	Diabetes Diagnosis	Comparison of classification algorithm accuracy
Methods for classifying diabetes mellitus data that are comparable: the machine learning paradigm (Maniruzzaman et al., 2017)	Classification with GPC, LDA, QDA, NB	Pima Indian Diabetes Data	Strengths and weaknesses of various diabetes diagnosis approaches
Using routinely gathered health records, longitudinal deep learning clustering of Type 2 Diabetes Mellitus trajectories was conducted. (Manzini et al., 2022)	Kernelized-AutoEncoder, classical clustering	Type 2 Diabetes Mellitus Trajectories	Identification of type 2 diabetes trajectories
utilising extensive and deep learning in conjunction with electronic health information to predict the onset of type 2 diabetes (Nguyen et al., 2019)	Wide and deep learning, linear regression, Artificial Neural Network	Onset of Type 2 Diabetes	Prediction of type 2 diabetes onset from electronic health records

Table 2 above presents various methods used in research on diabetes mellitus, the objects studied, the main findings of each study, and references for further information. These studies reflect efforts to effectively utilize data in the management and diabetes mellitus prediction employing a variety of contemporary data analysis methods.

Conclusions

Based on a comprehensive review of various data analysis studies related to diabetes mellitus, it can be concluded that these approaches significantly contribute to understanding, predicting, and managing the disease. The study by Mustamin et al. (2023) shows how data analysis methods like Principal Component Analysis (PCA) and artificial neural networks can improve the prediction of diabetic complication risks.. Additionally, research by Anwar et al. (2020) identifies advancements in artificial intelligence approaches for diabetes diagnosis,

while Maniruzzaman et al. (2017) and Manzini et al. (2022) highlight the effectiveness of classification and clustering techniques in grouping and understanding diabetes trajectories based on medical record data. Recent research by Nguyen et al. (2019) contributions as well by creating deep and broad learning models to forecast the onset of type 2 diabetes.. Overall, the integration of information technology and medical science in data analysis provides deeper insights into health patterns and lays a strong foundation for developing more effective diagnosis and management strategies in the future.

.

Funding

This study is financed using personal funds and does not include any external sources of financial support.

Acknowledgments

We express our sincere gratitude to the Institute of Science Technology and Health 'Aisyiyah Kendari and the Institute of Technology and Sciences Muhammadiyah North Kolaka for their invaluable support. Their assistance has been tremendously meaningful to us and has played a pivotal role in accomplishing our common objectives. This collaboration not only enhances institutional partnerships but also fosters mutual advancements in the fields of science and technology. We are deeply thankful for every contribution made.

Conflicts of Interest

This research is conducted independently, without any personal, financial, or other interests that could potentially influence the interpretation or representation of the research findings. The study is free from external influences or funding that could affect the design, data collection, analysis, interpretation, or decisions about publication. The funders did not take role in the design of the study, data collection, analysis, or interpretation, or manuscript writing or decision to publish the results.

References

American Diabetes Association. (2021). Diabetes care: A journal of the American Diabetes Association. American Diabetes Association.

Anwar, F., Qurat-Ul-Ain, Ejaz, M.Y., & Mosavi, A. (2019). A comparative analysis on diagnosis of diabetes mellitus using different approaches – A survey. Informatics in Medicine Unlocked, 16, 100204. https://doi.org/10.1016/j.imu.2019.100204

Bellinger, C., Mohomed Jabbar, M. S., Zaïane, O., & Osornio-Vargas, A. (2020). A systematic review of data mining and machine learning for air pollution epidemiology. BMC Public Health, 20(1), 1-28. https://doi.org/10.1186/s12889-019-8139-3

Chawla, N. V. (2005). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Hernes, M. I., & Tan, J. X. (2019). Utilizing machine learning for patient risk stratification in diabetes management. Journal of Diabetes Science and Technology, 13(2), 238-244. https://doi.org/10.1177/1932296818820038

Jothi, N., & Husain, W. (2015). Data mining in healthcare – A review. Procedia Computer Science, 72, 306-313. https://doi.org/10.1016/j.procs.2015.12.145

Kaggle. (2023). Anomaly Detection: A Comprehensive Guide. https://www.kaggle.com/code/pabloinsente/anomaly-detection-a-comprehensive-guide

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. Computational and Structural Biotechnology Journal, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8-17. https://doi.org/10.1016/j.csbj.2014.11.005

Maniruzzaman, M., Kumar, N., Abedin, M.M., Islam, M.S., Suri, H.S., El-Baz, A.S., & Suri, J.S. (2017). Comparative Approaches for Classification of Diabetes Mellitus Data: Machine Learning Paradigm. Computer Methods and Programs in Biomedicine, 152, 23-34. https://doi.org/10.1016/j.cmpb.2017.09.004

Manzini, E., Vlacho, B., Franch-Nadal, J., Escudero, J., Génova, A., Reixach, E., Andrés, E., Pizarro, I., Portero, J.-L., Mauricio, D., & Perera-Lluna, A. (2022). Longitudinal deep learning clustering of Type 2 Diabetes Mellitus trajectories using routinely collected health records. Journal of Biomedical Informatics, 135, 104218. https://doi.org/10.1016/j.jbi.2022.104218

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236-1246. https://doi.org/10.1093/bib/bbx044

Mustamin, S. B., Muhammad Atnang, Sahriani, B. Sulham, Samsidar. (2023). Health Data Analysis for In-Depth Understanding of Patterns, Prediction, and Disease Management: A Case Study on Diabetes Mellitus. International Journal of Science and Technology, 1(1), Vol.2(1), 2024, 27-34.

Nguyen, B.P., Pham, H.N., Tran, H., Nghiem, N., Nguyen, Q.H., Do, T.T.T., ... & Simpson, C.R. (2019). Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Computer methods and programs in biomedicine, 182, 105055.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances on deep learning techniques for electronic health record (EHR) analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589-1604. https://doi.org/10.1109/JBHI.2017.2767063

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

World Health Organization. (2020). Global report on diabetes. World Health Organization.