JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Ischemic Event Detection from Clinical Diagnosis of Diabetic-Cardiac Victims Through GAN

[1]Sathyasree C D, [2]Dr. K R Sumana

[1]PG Student, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka, India
[2]Faculty, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka, India

Abstract: Ischemic heart events represent a critical health risk, particularly among diabetic patients, as they often progress silently without early symptoms. This project presents an AI-driven diagnostic system that integrates Conditional Tabular GAN (CTGAN) for synthetic data generation with machine learning models (Logistic Regression, Random Forest, and K-Nearest Neighbors) and a deep learning Multilayer Perceptron (MLP) to enhance ischemic risk prediction. A curated diabetic-cardiac dataset, augmented through CTGAN and preprocessed for quality improvement, was used for training and evaluation. To ensure interpretability and transparency in decision-making, SHAP (SHapley Additive Explanations) and ALE (Accumulated Local Effects) techniques were applied for feature importance analysis. The system is deployed through a Flask-based web application, allowing clinicians to input patient data, obtain predictions, and interpret risk factors. The proposed model supports early diagnosis, improves clinical decision-making, and aims to reduce mortality in high-risk diabetic-cardiac patients.

Index Terms – Ischemic Heart Disease, CTGAN, Machine Learning, Deep Learning, SHAP, ALE, Risk Prediction, Flask

I. Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, with ischemic heart disease posing a major risk, especially among diabetic patients. Early detection of ischemic events is critical, yet traditional diagnostic methods are often invasive and limited. To overcome the scarcity of medical datasets, this study employs Conditional Tabular GAN (CTGAN) for generating synthetic diabetic-cardiac records, enhancing training diversity. Machine learning models such as Logistic Regression, Random Forest, and K-Nearest Neighbors, along with a Multilayer Perceptron, were evaluated for ischemic risk prediction. SHAP and ALE methods were integrated to ensure interpretability and highlight key predictive features. Finally, a Flask-based web application was developed to provide real-time predictions using with an interactive interface, bridging advanced AI with practical clinical decision-making

II. LITERATURE SURVEY

Several studies have explored the use of machine learning (ML) and deep learning (DL) techniques for cardiovascular disease (CVD) detection and prediction in diabetic patients. David Chushig et. al. [1] proposed an interpretable, data-driven approach using the T1D dataset with algorithms such as KNN, Decision Tree, Random Forest, and MLP, where MLP achieved the best performance (MAE = 0.0088, MRAE = 0.0817), identifying age, HbA1c, and albuminuria as key features. Wackers et. al. explores [2] investigated silent myocardial ischemia in type 2 diabetics using adenosine SPECT imaging, finding 22% prevalence and highlighting the potential of AI integration for improved risk prediction. Lamario J. Williams et. al. explores [3] reviewed mechanisms of diabetic cardiomyopathy, including insulin resistance, microRNAs, and ROS, providing a basis for DL-based mechanistic modeling. Javaid and Alrajeh et.al. explores [4] employed XGBoost with SHAP explainability on UCI and Kaggle datasets, achieving 88% accuracy and demonstrating the value of interpretable AI in healthcare.

Ayush Mishra and Bhaskar Agarwal et.al. [5] developed an SVM + KNN hybrid model using the Framingham dataset, achieving 83% precision, suggesting benefits of hybrid approaches. Dr. S. Nandini [6] et.al. explores compared ML algorithms (LR, SVM, RF, KNN, Naive Bayes) on UCI and Statlog datasets, with RF and SVM performing best (85–87% accuracy) but requiring further benchmarking. Zhang Y et.al. explores [7] applied CNN and RNN models to diabetes data, improving accuracy but noting interpretability challenges. Chen L. [8] et.al. explores proposed a DNN framework for early ischemic heart disease detection, outperforming traditional methods yet requiring broader validation. For sequential prediction, Kumar A et.al. explores [9] used LSTM and GRU models on cardiac event data, achieving high accuracy and showing promise for time-series health analytics. Lee J et.al. explores [10] evaluated ensemble learning techniques, finding higher accuracy than individual models and suggesting future exploration of hybrid ensemble strategies. Collectively, these studies demonstrate that ML/DL methods—especially interpretable, hybrid, and ensemble modelshold significant promise for accurate and early CVD risk prediction in diabetic populations, though broader validation and interpretability remain essential for clinical adoption.

III. PROPOSED WORK

The proposed system is designed to enable end-to-end ischemic event risk prediction from clinical patient data through a modular and interpretable architecture. Clinical datasets containing diabetic-cardiac parameters such as age, blood pressure, cholesterol, and glucose are pre-processed using cleaning, normalization, and feature selection. The refined data is trained on XGBoost and MLP models to predict ischemic event risk, optimized using cross-validation and hyperparameter tuning. To enhance explainability, techniques like SHAP and ALE are integrated for feature importance and confidence visualization. A Flask-based web interface provides secure access, intuitive input forms, and interactive dashboards for prediction results. The system delivers high/low risk outcomes with confidence scores, feature impact analysis, and precautionary recommendations, ensuring clinical transparency and actionable insights.

IV. METHODOLOGY

4.1 Data Collection

The dataset used in this study consists of structured clinical records of patients diagnosed with diabetes and exhibiting cardiovascular risk factors. It includes a variety of patient health parameters such as age, gender, systolic and diastolic blood pressure, cholesterol and glucose levels, lifestyle-related indicators like smoking and alcohol consumption, as well as physical activity, weight, and height. These features were carefully selected to cover both physiological and behavioral attributes associated with ischemic heart risk. By incorporating such a diverse range of clinical variables, the dataset ensures that the predictive analysis is well generalized to real-world diabetic patients with cardiac complications.

4.2 Data Preprocessing

Before applying machine learning and deep learning algorithms, extensive preprocessing was performed to ensure data quality, consistency, and compatibility. Missing values in the dataset were imputed using statistical measures such as mean and median to avoid bias from incomplete records. Categorical variables like gender, smoking status, alcohol consumption, and physical activity were converted into numerical values through label encoding. Feature scaling was carried out using the StandardScaler technique to normalize the variables, ensuring uniform weightage and faster convergence of models. To address the class imbalance problem between low-risk and high-risk ischemic cases, additional synthetic samples were generated using CTGAN, thereby improving model robustness and prediction accuracy.

4.3 Synthetic Data Generation Using CTGAN

To strengthen the dataset and overcome limitations caused by class imbalance, Conditional Tabular GAN (CTGAN) was applied for synthetic data generation. CTGAN is specifically designed to generate realistic tabular data that mimics the statistical patterns of the original dataset. In this project, it was particularly useful in augmenting minority high-risk ischemic cases, which are often underrepresented in real-world clinical data. By enriching the dataset with synthetic yet clinically valid samples, the model was able to learn more robust decision boundaries and avoid bias toward majority low-risk patients. This approach ultimately enhanced the accuracy and reliability of ischemic risk prediction.

4.4 Model Selection and Training

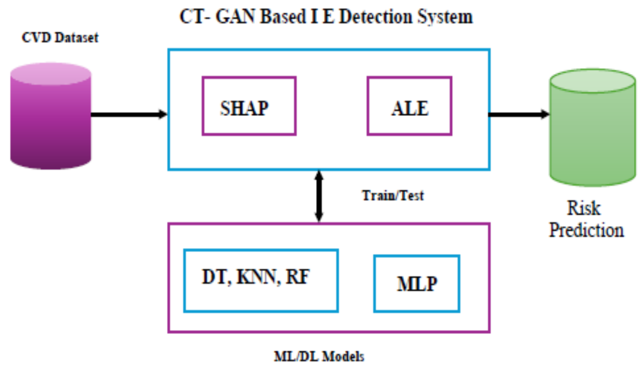
Both machine learning and deep learning techniques were employed to build predictive models for ischemic event detection. On the machine learning side, models such as Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and XGBoost Classifier were explored for their ability to identify nonlinear relationships in clinical features. XGBoost was particularly emphasized due to its gradient boosting mechanism that efficiently handles imbalanced data and improves predictive performance. In parallel, a Multilayer Perceptron (MLP) deep learning model was developed to capture complex feature interactions and non-linear dependencies. Model performance was validated using a train-test split, and evaluation metrics such as accuracy and precision were considered to ensure generalizability.

4.5 Web-Based Deployment Using Flask

To translate the research outcomes into a practical solution, a web-based application was developed using the Flask framework. The system was designed to be simple yet user-friendly, enabling healthcare professionals to access predictions in real time. Through secure login, users can input patient parameters and obtain ischemic risk predictions categorized as low or high risk. Additionally, the application integrates interpretability insights by highlighting the top five influential features contributing to each prediction. It also provides visualizations of dataset summaries, including distributions of age, cholesterol, glucose, BMI, and overall cardiovascular risk patterns, thus offering both prediction and interpretability in a single platform. The final system architecture successfully integrates backend processing logic with an interactive frontend interface. The frontend was designed using HTML, CSS, and Jinja templates to provide a responsive and engaging user experience. On the backend, Python scripts handled data preprocessing, model inference, and visualization tasks, ensuring accurate and efficient prediction. To make the system more intuitive, various charts and graphs were generated using Matplotlib and Seaborn, which were embedded into the web application via Base64 encoding for seamless display. This integration ensured smooth interaction between users and the system, bridging the gap between advanced AI models and practical healthcare usability. The developed application thus demonstrates the feasibility of deploying AI-driven ischemic risk prediction in real-world healthcare settings.

Table 1: Comparative

Model	Training Accuracy (%)	Validation Accuracy (%)	Training Time (sec)
Logistic Regression	72.07	70.15	1.2
KNN	68.42	66.80	0.9
Decision Tree	72.94	71.10	1.8
Random Forest	69.72	68.50	3.5
MLP Classifier	73.88	72.30	6.2

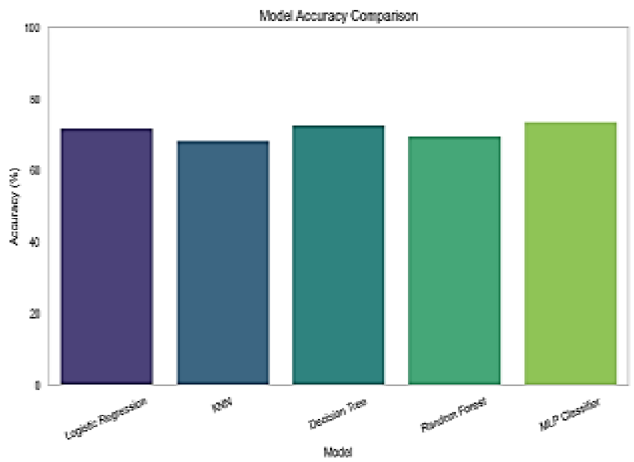


Analysis of Model Performance

Figure 1- Ischemic Risk Prediction System Design

V. RESULT ANALYSIS

Figure 2 depicts the performance comparison among the selected machine learning and deep learning models for ischemic event detection. Logistic Regression achieved a baseline accuracy of 72.07%, performing effectively on linearly separable cases. KNN recorded the lowest accuracy of 68.42%, reflecting its sensitivity to feature scaling and dataset noise. The Decision Tree classifier achieved 72.94%, slightly outperforming Logistic Regression, but suffered from overfitting tendencies. Random Forest, though robust, yielded 69.72%, showing stable performance but relatively lower generalization compared to the Decision Tree. The MLP Classifier achieved the highest accuracy of 73.88%, demonstrating its strength in capturing complex non-linear relationships in diabetic-cardiac patient data. From the analysis, it is evident that deep learning-based approaches such as the MLP Classifier outperform traditional machine learning models like Logistic Regression, KNN, and Random Forest in predicting ischemic event risks. While conventional models provide a baseline for comparison, their predictive power is limited due to their inability to capture complex nonlinear relationships in the dataset. On the other hand, neural architectures leverage multiple hidden layers and non-linear activation functions, enabling them to identify subtle clinical patterns that are often overlooked by simpler algorithms. This superior performance highlights the effectiveness of deep learning for handling high-dimensional medical data.



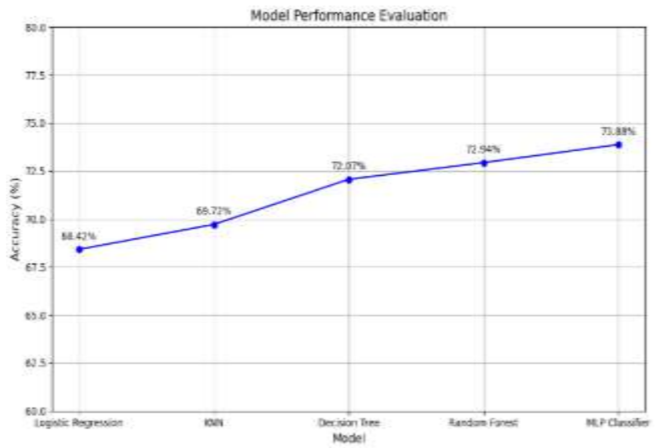


Figure 2(a) – ML/DL Model Accuracy Comparison. Figure 2(b) –ML/DL Model Performance Evaluation

Figure 2(a) presents a bar chart comparing the accuracy of different models, where the MLP Classifier achieved the highest performance. Figure 2(b) shows a line graph highlighting the trend in model performance, which demonstrates a steady improvement from Logistic Regression to MLP Classifier.

User Interface and Module Overview**:** The frontend of the system is designed with a structured and user-friendly layout that ensures smooth navigation across different modules. The interface provides secure access through a login process, followed by a central dashboard that connects users to key functionalities such as dataset exploration, prediction, and result analysis. An about section is included to describe the project's objectives, methodology, and scope, helping users gain a clear understanding of the system's purpose and the dataset module presents clinical information used for model training and evaluation, highlighting its role in supporting accurate ischemic event risk prediction. The interface not only enhances user engagement but also ensures reliability in accessing critical information. Each module has been carefully designed to deliver relevant outputs, making the system both practical for healthcare applications and adaptable for future extensions.

Figure 3(a) – Login Page of the System

Figure 3(b) – Home Page of the System

Figure 3(a) shows the login interface of the ischemic event prediction system, where users can securely enter their credentials to access the platform. Figure 3(b) presents the home page, which introduces the system's purpose, emphasizing the use of CTGAN for synthetic data generation and improved accuracy in ischemic event detection.

Figure 4(a) – System Introduction Page

Figure 4(b) – Project Metrics and Scope

Figure 4(a) displays the About Page, which provides an overview of the project, highlighting the role of GANs in ischemic event risk prediction. Figure 4(b) presents the Scope of the system, summarizing key achievements such as the number of patient records processed, clinical features considered, prediction accuracy, and potential for future deployment.

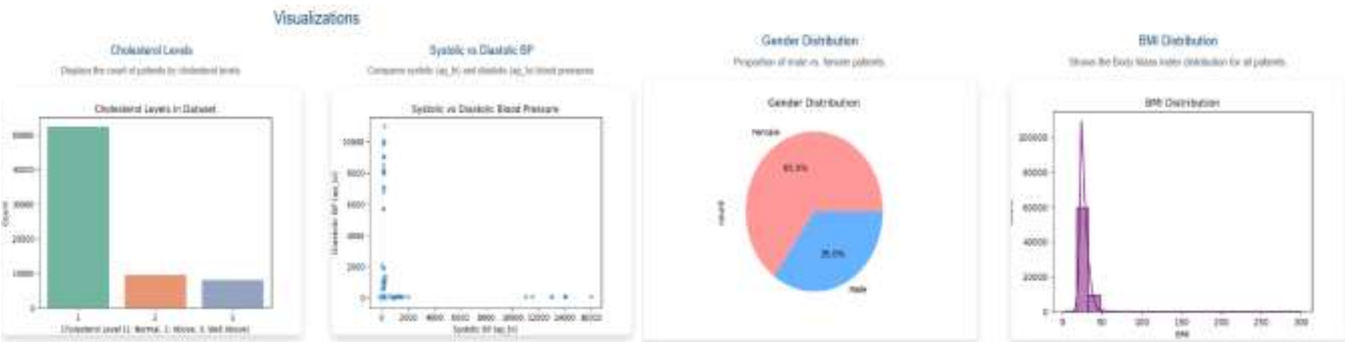


Figure 5 shows dataset visualizations of cholesterol, blood pressure, gender, and BMI

Figure 5 shows dataset distributions of cholesterol levels and the relationship between systolic and diastolic blood pressure among patients, illustrates the gender distribution of patients along with BMI variations, helping to identify population patterns relevant to ischemic event risk prediction.

The Ischemic Risk Prediction page serves as an interactive tool that enables patients or clinicians to assess the likelihood of ischemic heart risk based on multiple health parameters. The user inputs details such as age, gender, height, weight, systolic and diastolic blood pressure, cholesterol, glucose levels, lifestyle factors (smoking, alcohol use, and physical exercise), and other clinical measures. Once submitted, the model processes this data and generates a prediction indicating whether the patient falls into a High Risk or Low Risk category. Along with the prediction, the page displays model confidence levels, showing the probability distribution between risk categories, thereby enhancing transparency of the AI's decision-making process. To further improve interpretability, the system highlights the Top 5 most important features influencing the prediction through a bar chart. For instance, systolic blood pressure and cholesterol levels appear as highly influential factors in determining ischemic risk.

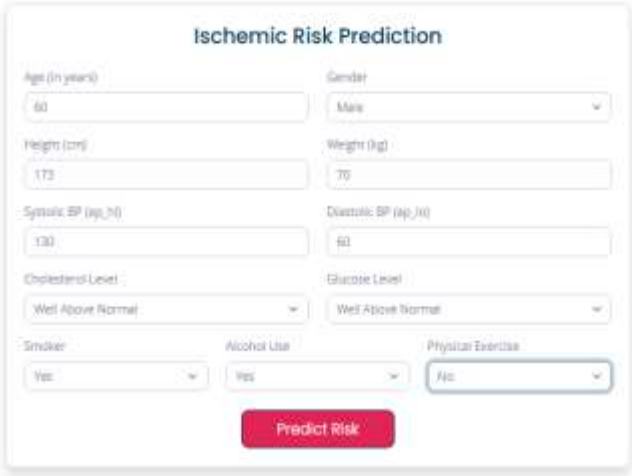


Figure 6(a) – Prediction Page (Input Section) Figure 6(b) – Prediction Page (Output Section)

Figure 6(a**)** shows the input section where patient details such as age, vitals, and lifestyle factors are entered for risk prediction. Figure 6(b) illustrates the output section displaying the predicted ischemic event risk along with key contributing features for interpretability.

VI. CONCLUSION

This project titled "Ischemic Event Detection from Clinical Diagnosis of Diabetic-Cardiac Victims through GAN" was developed with the objective of creating a predictive framework capable of identifying ischemic heart event risks at an early stage using both real and GAN-generated synthetic patient data. The methodology integrated data preprocessing, synthetic data augmentation, and advanced machine learning and deep learning models to improve prediction accuracy and address the limitation of small, imbalanced medical datasets. The study demonstrated that synthetic data generation using CTGAN can effectively expand limited datasets while preserving realistic feature relationships. Models such as MLP and XGBoost showed strong predictive capabilities for ischemic risk detection, and the importance of preprocessing and feature selection was evident in improving accuracy and model stability. Visualizations such as correlation heatmaps and demographic plots proved useful in understanding feature interactions and dataset trends, enabling better model interpretation. The final model achieved promising accuracy, indicating its potential to assist healthcare professionals in early diagnosis and preventive care planning. Additionally, the use of synthetic data addressed patient privacy concerns while enhancing model robustness without compromising data utility.

VII. ACKNOWLEDGEMENT

I would especially want to thank Dr. K R Sumana, my guide, for her constant support, and The National Institute of Engineering, Mysuru's dedicated staff. I truly owe my friends and classmates, as well as my ever-supportive parents, for their priceless encouragement. Finally, I want to express my sincere gratitude to everyone who helped me finish my project successfully, either directly or indirectly.

References

[1] D. Chushig-Muzo, H. Calero-Díaz, F. J. Lara-Abelenda, V. Gómez-Martínez, C. Granja and C. Soguero-Ruiz, "Interpretable Data-Driven Approach Based on Feature Selection Methods and GAN-Based Models for Cardiovascular Risk Prediction in Diabetic Patients," in IEEE Access, vol. 12, pp. 84292-84305, 2024, doi: 10.1109/ACCESS.2024.3412789.

[2] M. Bhagawati and S. Paul, "Generative Adversarial Network-based Deep Learning Framework for Cardiovascular Disease Risk Prediction," 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2024, pp. 1-4, doi: 10.1109/ICITIIT61487.2024.10580562..

[3] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, ''Secure and robust machine learning for healthcare: A survey,'' IEEE Rev. Biomed. Eng., vol. 14, pp. 156–180, 2021.

[4] S. Chua, V. Sia and P. N. E. Nohuddin, "Comparing Machine Learning Models for Heart Disease Prediction," 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2022, pp. 1-5, doi: 10.1109/IICAIET55139.2022.9936861

[5] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, ''IDF diabetes atlas: Global estimates of the prevalence of diabetes for 2011 and 2030,'' Diabetes Res. Clin. Pract., vol. 94, no. 3, pp. 311–321, Dec. 2011.

[6] O. M. Moushi, N. Ara, M. Helaluddin and H. S. Mondal, "Enhancing the Accuracy and Explainability of Heart Disease Prediction Models through Interpretable Machine Learning Techniques," 2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ICICT4SD59951.2023.10303572.

[7] M. A. Ahmad, A. Teredesai, and C. Eckert, ‘‘Interpretable machine learning in healthcare,’’ in Proc. IEEE Int. Conf. Healthcare Informat. (ICHI), Jun. 2018, pp. 1–447.

[8] J. Zhong, D. Wei, Y. Zhuo, M. Yin, Y. Kang and W. Shi, "Feasibility Study and Practice of Machine Learning-Based Heart Disease Prediction," 2025 8th International Conference on Electronics Technology (ICET), Chengdu, China, 2025, pp. 492-496, doi: 10.1109/ICET64964.2025.11103110.

[9] R. Muthalagu, R. R, A. T, A. PH and J. A. A, "Pattern Recognition and Modelling in Electrocardiogram Signals: Early Detection of Myocardial Ischemia and Infraction," 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 1035-1041, doi: 10.1109/ICECAA58104.2023.10212133

[10] C. -W. Chang, K. -M. Liao, Y. -C. Chen, S. -H. Wang, M. -Y. Jan and G. -C. Wang, "Radial Pulse Spectrum Analysis as Risk Markers to Improve the Risk Stratification of Silent Myocardial Ischemia in Type 2 Diabetic Patients," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1-9, 2018, Art no. 1900509, doi: 10.1109/JTEHM.2018.2869091.

[11] H. Calero-Diaz, D. Chushig-Muzo, H. Fabelo, I. Mora-Jiménez, C. Granja, and C. Soguero-Ruiz, ‘‘Data-driven cardiovascular risk prediction and prognosis factor identification in diabetic patients,’’ in Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat., Sep. 2022, pp. 1–4.

[12] Y.-Y. Chen, Y.-J. Lin, E. Chong, P.-C. Chen, T.-F. Chao, S.-A. Chen, and K.-L. Chien, ‘‘The impact of diabetes mellitus and corresponding HbA1c levels on the future risks of cardiovascular disease and mortality: A representative cohort study in Taiwan,’’ PLoS ONE, vol. 10, no. 4, Apr. 2015, Art. no. e0123116.

[13] J. M. Lachin, ‘‘Risk factors for cardiovascular disease in type 1 diabetes,’’Diabetes, vol. 65, no. 5, p. 1370, 2016.

[14] M. A. Ahmad, A. Teredesai, and C. Eckert, ‘‘Interpretable machine learning in healthcare,’’ in Proc. IEEE Int. Conf. Healthcare Informat. (ICHI), Jun. 2018, pp. 1–447.

[15] M. Jurado-Camino, D. Chushig-Muzo, C. Soguero-Ruiz, P. Bohoyo, and I. Mora-Jiménez, ‘‘On the use of generative adversarial networks to predict health status among chronic patients,’’ in Proc. 16th Int. Joint Conf. Biomed. Eng. Syst. Technol., 2023, pp. 167–178.