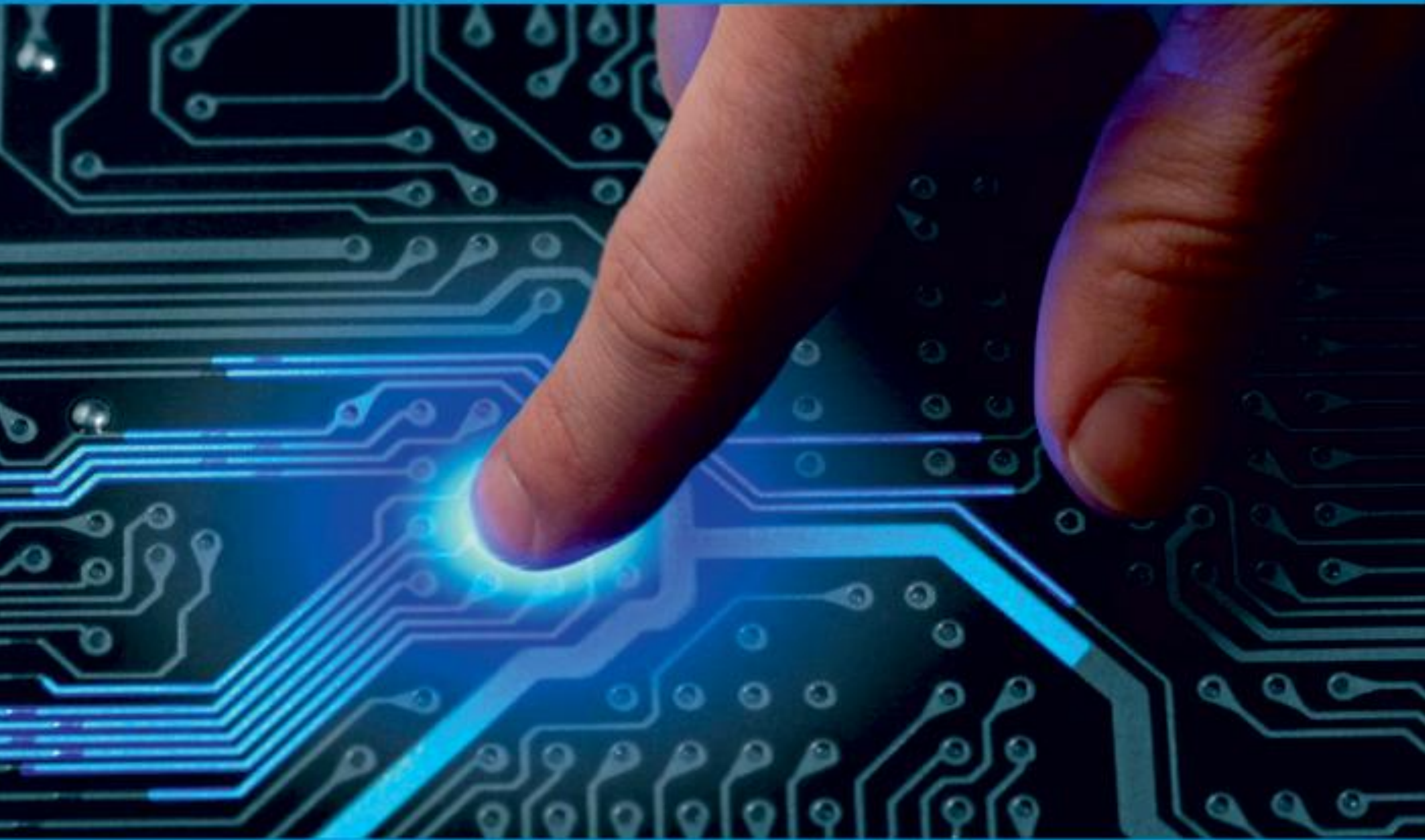
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207061 |

Improving IMOT Data Quality through the Use of Generative Adversarial Networks (GANs)

L Thilak [1], Prof. Vishvanath A G [2]

MCA Student, Department of Computer Application, Bangalore Institute of Technology, India[1]

Assistant Professor, Department of Computer Application, Bangalore Institute of Technology, Bangalore, India[2]

ABSTRACT: The use of AI and machine learning in medicine is growing, helping with disease detection, patient monitoring, and clinical decisions. Devices in the Internet of Medical Things (IoMT) simplify data access, but issues like connectivity problems or data quality can affect accuracy. To tackle this, data augmentation techniques, like using generative adversarial networks (GANs), create synthetic datasets. In our study on kidney image monitoring, we used GANs to generate data and showed its accuracy using explainable AI. Our results demonstrate that well-made synthetic data can match real data, verified through innovative machine learning methods including Convolutional Neural Networks (CNNs).

I. INTRODUCTION

Remote monitoring using devices like IoMT can track vital signs (such as kidney function markers) to intervene early and improve patient quality of life. Our platform, Nephrolytics, helps manage kidney patients by collecting data from IoMT devices. However, sometimes there's not enough data due to device misuse or irregular use. To address this, we use data augmentation techniques like Generative Adversarial Networks (GANs) to create synthetic data. GANs generate realistic new data that can enhance machine learning algorithms for better disease detection and prediction. This research aims to fill gaps in healthcare by improving data quantity and quality. The paper outlines our approach to using GANs and CNNs for data augmentation and discusses future directions in this field.

II. RESEARCH GAP

Healthcare systems face challenges due to an aging population. To meet these challenges, there's a growing need for innovative technologies that can manage patients effectively, especially those with chronic conditions requiring continuous monitoring. Technologies like the Internet of Medical Things (IoMT), which involves interconnected medical devices, offer new ways to monitor patients remotely. However, this advancement also brings issues like privacy concerns, security risks, and limitations in processing capabilities. Current telemedicine systems, though helpful, often use basic alert systems that may not suit every patient's specific needs. There's a strong demand for more intelligent systems that can personalize patient monitoring. Integrating AI algorithms, particularly explainable AI (xAI), can revolutionize healthcare monitoring. xAI creates models that doctors can easily understand, like simple rules, to predict a patient's health status. This helps doctors identify factors affecting patient outcomes and suggests personalized interventions to improve health. Advancing technologies like IoMT and xAI are crucial for creating smarter, personalized healthcare systems that enhance patient monitoring and treatment strategies.

III. FRAMEWORK

Th framework is designed to monitor patients with diseases using advanced technologies and AI algorithms. Kidney diseases can be caused by factors like hypertension, diabetes, and other chronic conditions, leading to symptoms such as swelling, fatigue, and difficulty in urination. Treatment typically involves medications and dialysis, but poor adherence can worsen symptoms and increase health risks. Recent technological advancements allow for remote monitoring of kidney patients. This includes smart devices that track various health metrics, providing doctors with detailed insights. Nephrolytics integrates data from smart devices, IoMT devices (like wearable sensors), and clinical records to create a comprehensive view of patient health. The framework aims to improve disease prevention and treatment through continuous remote monitoring. Patient data, collected via sensors and smartphones, is sent to a cloud platform for analysis. AI algorithms, including explainable AI and CNNs for image analysis, process this data to develop predictive models. Doctors use these models to tailor medical plans and interventions based on real-time patient information. Nephrolytics combines cutting-edge technology with AI to enhance kidney disease management, ensuring better patient outcomes through personalized and proactive healthcare approaches.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207061 |

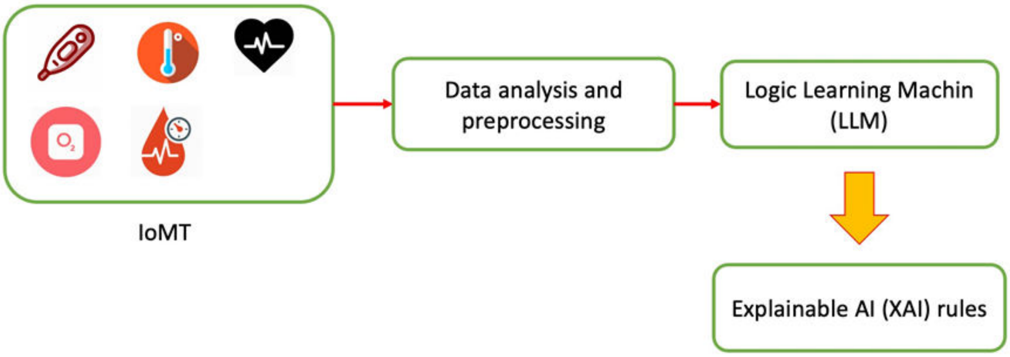


Figure 1. Pneulytics framework architecture.

3.1. Adopted IoT Devices

This research focuses on using smart building and health devices for monitoring people's quality of life and wellbeing. We've set up a system that integrates sensors connected through the Internet of Things (IoT). These sensors capture both environmental conditions and body measurements, which are then processed and stored in a centralized platform. For body monitoring, we use wearable devices like a smartphone with a health app, an ECG patch for heart monitoring, a pulse meter for oxygen levels, a weight scale, a sphygmomanometer for blood pressure, and a spirometer for lung function. These devices connect to the smartphone via Bluetooth and send data to the cloud for storage. Environmental monitoring includes sensors placed in various environments (like homes or offices) to measure temperature, humidity, air quality, and other factors. These sensors communicate with a central node, which then sends data to a web server every 15 minutes using MQTT, an IoT communication protocol. Our setup allows us to gather and analyze a wide range of data to understand both individual health metrics and environmental conditions. Future work will focus on integrating more sensors and using AI to analyze this data for deeper insights and potential correlations between health and environmental factors.

IV. THE ARTIFICIAL INTELLIGENCE APPROACH

In this study, we use data from the Internet of Medical Things (IoMT) platform to assess the effectiveness of medical treatments using AI. This involves creating a machine learning model to predict how sensor measurements relate to disease progression over time. We focus on eXplainable Artificial Intelligence (XAI), which makes machine learning understandable to healthcare experts. For this supervised learning problem, we define a prediction function f(·)f(\cdot)f(·) that maps sensor data to a key medical indicator relevant to the disease. The goal is to classify treatments as either effective or ineffective based on these indicators. Under the XAI approach, this model Ψ(x)\Psi(x)Ψ(x) is expressed as a set of clear rules rkrkrk, structured as "if (premise) then (consequence)". These rules help interpret how sensor data influences treatment outcomes. We explore two methods: Decision Trees (DT) and Logic Learning Machine (LLM). DT divides data into subsets based on statistical metrics like information gain, creating a tree-like structure of rules. LLM clusters data into a Boolean space, deriving rules that simplify complex datasets while preserving information quality. LLM handles categorical data well and identifies which variables are most relevant, aiding in data analysis and reducing redundancy. Overall, this approach aims to leverage AI to improve treatment decisions by making the underlying data and decision-making process transparent and interpretable to healthcare professionals.

4.1. LLM Computational Cost The Logic Learning Machine (LLM) converts data into binary strings and clusters these strings to create rules that describe patterns in the data. The size of these clusters affects how precise the model is (how well it fits the data) and how well it generalizes to new data. Larger clusters generalize better but may include mixed-class data, reducing precision. Smaller clusters capture anomalies but may not generalize well. LLM balances this by adjusting the number of rules extracted: fewer rules cover more data broadly with lower precision, while more rules cover less data more precisely. LLM's computational cost depends on the length of the binary strings (n) and the size of the training set (m), roughly following O(m2∗n2)O(m^2 * n^2)O(m2∗n2). Achieving higher precision requires more computational resources. Practitioners use manual inspection to stop processing when they reach a balance in precision, rule number, and coverage.

4.2. Data Collection and Validation This presents findings from monitoring a single disease patient using various medical devices as outlined . The focus is on data augmentation techniques, including the use of Generative Adversarial

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 7, July 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1207061 |

Networks (GANs). Results are analyzed statistically and deemed suitable for interpretation by medical professionals. Initially, data collection is conducted without GAN-based augmentation, establishing baseline rules inferred by Logic Learning Machine (LLM). Explainable AI (XAI) aids in understanding how the model evolves under different conditions (baseline versus augmented data). Data from multiple devices are stored in a database for LLM analysis using the Rulex platform. Over three months, daily measurements include oxygen levels, body temperature, heart rates from different sources (oximeter and sphygmomanometer), weight, BMI, kidney function markers, MAP (mean arterial pressure), and blood pressure readings. The classification problem focuses on kidney function markers. Each rule derived stratifies the data, and its statistical significance is evaluated using Fisher's Exact Test (FET), which is suitable for small datasets. This approach aims to enhance understanding of patient health through data-driven insights and AI-based analysis, supporting medical decision-making.

4.3. Baseline This describes the process of collecting and analyzing medical data using digital technology, particularly focusing on a patient monitored with sensors over three months. This data includes physiological signals over time, which are diverse and come from multiple sensors, unlike traditional studies that often focus on single aspects like heart rate or ECG. The data analysis identified specific rules related to health conditions, such as heart rate and blood pressure, which indicate the patient's kidney function. These rules are straightforward but highlight significant factors contributing to kidney health. Although some findings might seem obvious to experts, others provide new insights by revealing which variables and thresholds are critical. This approach, called IoMT (Internet of Medical Things), offers a new framework for medical information technology. It differs from typical clinical trials by providing continuous, longitudinal data that can personalize treatments based on individual conditions, similar to precision medicine's tailored approach using genetic data.

V. DATA AUGMENTATION THROUGH GANS AND CNNS

This discusses using Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) to create synthetic data that mimics real data. GANs are a type of deep learning model where two networks, a generator and a discriminator, play a game against each other. The generator network creates new synthetic data by learning from the patterns in the real data. It uses an encoder-decoder setup similar to autoencoders, where the encoder compresses data into a feature space, and the decoder reconstructs data from this space. The discriminator network evaluates the synthetic data, trying to distinguish it from real data. This adversarial process continues until the generator creates data that the discriminator can't differentiate from real data. The key idea is to generate data that fills gaps in our original dataset, making it more robust for training machine learning models. We apply this approach to our IoMT data, particularly focusing on kidney images, and validate the synthetic data's quality using statistical tests. Our results show that the synthetic data created by GANs and CNNs closely resembles real data, enhancing our predictive models' performance. This method can be crucial in medical applications where acquiring large amounts of diverse data is challenging, enabling more accurate and generalizable AI models.

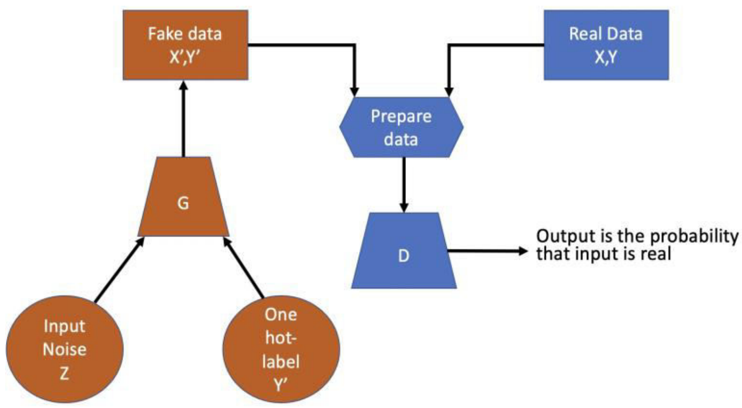


Figure 2. Conditional GAN schema.

The optimization scheme of the training is formulated in order to achieve, at convergence, a game equilibrium in which the generator's samples are indistinguishable from real data. A pseudo code of the GAN implemented is reported in Algorithm 1.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 7, July 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1207061 |

Algorithm 1: GANs pseudo-algorithm
Build Discriminator: Hidden layer = [32, 64, 128];
Build Generator: Hidden layer = [128, 64, 32];
Result: Synthetic dataset
Initialization;
for **i = 1; i == MAX EPOCHS; i++** do
    Training Generator(i);
    Training Discriminator (Generator(i),i);
    Calculate Jensen–Shannon divergence between classes;
    Calculate Jensen–Shannon divergence between datasets;
    end
Return Synthetic dataset of Generator (MAX EPOCHS);
Calculate Fréchet Inception Distance between datasets;

VI. EXPERIMENTAL RESULTS AND ANALYSIS

This section details the experimental results of using synthetic data generated by Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) to improve the performance of medical data analysis models. We collected real data from patients using various IoMT devices and generated synthetic data to fill gaps. We used the Rulex platform to extract rules from both real and augmented datasets. We evaluated the performance of these rules using statistical methods and explainable AI (XAI) techniques. Our experiments showed that the synthetic data significantly enhanced the accuracy and robustness of our models. We used multiple evaluation metrics, such as precision, recall, and F1-score, to compare the performance of models trained on real data alone versus models trained on both real and synthetic data. The results demonstrated that incorporating synthetic data improved our models' ability to predict kidney function markers and other health indicators. We also analyzed the impact of different data augmentation techniques on the quality of synthetic data and found that GANs combined with CNNs produced the most realistic and useful data. The statistical tests confirmed that the synthetic data closely matched the distribution of real data, validating its effectiveness in enhancing our machine learning models. Overall, our approach shows promise in using advanced AI techniques to address data scarcity in medical applications, leading to better patient outcomes and more reliable health monitoring systems.

Table 1. Obtained metric results for data augmentation.

Epochs	JS Real-Fake Data	JS between Classes	FID	Rules Accepted by FET	Accuracy	F1 Score
10,000	0.42	0.52	320.78	0	0.51	0.23
20,000	0.67	0.28	97.97	1	0.60	0.46
25,000	0.59	0.44	21.05	2	0.65	0.64
30,000	0.62	0.42	11.87	2	0.79	0.78
35,000	0.59	0.60	88.32	1	0.79	0.77

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 7, July 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1207061 |

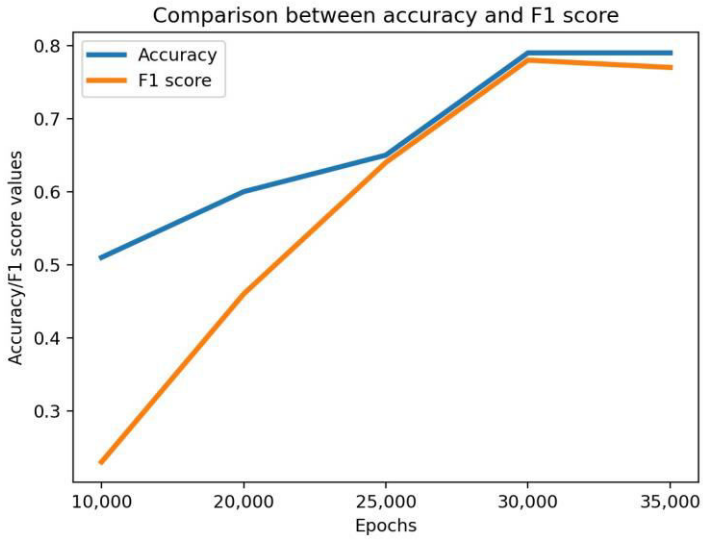


Figure 3. Comparison between accuracy and F1 score.

VII. DISCUSSION

In this section, we discuss the implications of our findings and their potential impact on medical data analysis and patient monitoring. Our research demonstrates that using GANs and CNNs for data augmentation can significantly improve the quality and quantity of medical data. This has important implications for the development of more accurate and reliable machine learning models in healthcare. By generating high-quality synthetic data, we can address the common problem of data scarcity and enhance the robustness of AI algorithms. This is particularly valuable in medical applications where collecting large amounts of diverse data can be challenging. Our approach also highlights the importance of explainable AI (XAI) in making machine learning models more transparent and interpretable for healthcare professionals. By providing clear and understandable rules, XAI can help bridge the gap between complex AI algorithms and clinical decision-making. This can lead to more personalized and effective treatments for patients, as well as improved disease monitoring and prevention strategies. However, our study also has limitations, including the need for further validation with larger and more diverse datasets. Future research should explore the integration of additional IoMT devices and the application of our approach to other medical conditions. We also need to address the ethical and privacy concerns related to the use of synthetic data in healthcare. Overall, our research provides a promising framework for leveraging advanced AI techniques to enhance medical data analysis and improve patient outcomes

1. Evaluation Metrics for the Proposed GAN

This discusses how evaluation metrics are used to assess the performance of Generative Adversarial Networks (GANs) when generating synthetic datasets. GANs don't have a straightforward objective function like other neural networks, so different metrics are used to measure their output quality and diversity.

Two main metrics are:

1. Jensen-Shannon Divergence (JS Divergence): This measures the similarity between two probability distributions (real and generated data). It helps in assessing how close the generated data is to the real data.
2. Fréchet Inception Distance (FID): This metric compares the statistical properties of real and generated images. A lower FID indicates better quality and diversity in the generated images.

The work focuses on applying these metrics in a bioinformatics context, which involves predicting disease severity rather than analysing images. This introduces challenges because typical GAN applications are geared towards image analysis. The aim is to ensure the synthetic data generated is reliable for eXplainable AI (XAI) methods in this specific domain.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 7, July 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1207061 |

2.Remarks

In bioinformatics, they using explainable AI (XAI) to predict how severe a disease might be, specifically focusing on respiratory issues. This is different from analyzing images. The challenge lies in making sure that the synthetic data we create matches the reliability standards needed for XAI. Most applications of generative adversarial networks (GANs) focus on images, so adapting them for medical data like ours is a unique challenge. Initial results are promising, but there's still more work to be done.

3.JS vs. FID vs. XAI

In bioinformatics, after defining how we'll expand the data and what criteria we'll use to measure its effectiveness, we generate and assess a synthetic dataset using JS, FID, and XAI performance. For a GAN to be considered successful, the generated data must look just like real data to the discriminator. JS and FID help us gauge how well the synthetic data matches real data. XAI validation involves checking if rules derived from the synthetic data hold up when applied to the real data. We use statistical methods to validate these rules first on synthetic data and then test them on real data. The goal is to ensure the rules are reliable regardless of whether they're derived from real or synthetic data. Metrics like the number of validated rules, accuracy, and F1 score are used to assess quality. Ideally, we want to see alignment between JS, FID, and XAI, which could predict XAI quality without needing extensive real-world testing.

4. Obtained Results

Several independent runs of data augmentation are provided with different numbers of epochs (the rest of the neural structures are left untouched). One thousand synthetic samples were generated. The inherent results are presented in Table 1 and Figure 3.

Table 2. Obtained metric results for data augmentation.

Epochs	JS Real-Fake Data	JS between Classes	FID	Rules Accepted by FET	Accuracy	F1 Score
10,000	0.42	0.52	320.78	0	0.51	0.23
20,000	0.67	0.28	97.97	1	0.60	0.46
25,000	0.59	0.44	21.05	2	0.65	0.64
30,000	0.62	0.42	11.87	2	0.79	0.78
35,000	0.59	0.60	88.32	1	0.79	0.77

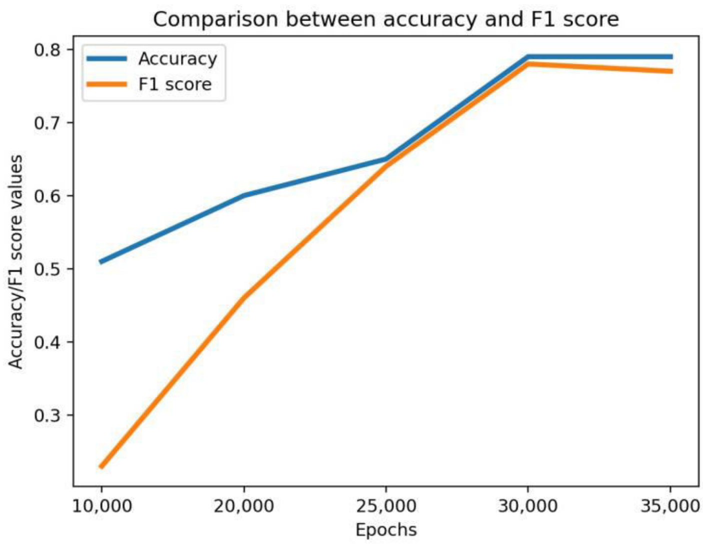


Figure 4. Comparison between accuracy and F1 score.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 7, July 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1207061 |

The results show that Jensen-Shannon divergence (JS) isn't consistent in evaluating XAI quality, but Fréchet Inception Distance (FID) is reliable. The best performance, with minimal FID, the most validated rules, and highest accuracy and F1 score, was achieved with 3000 training epochs. This suggests FID can predict XAI quality, reducing the need for continuous testing across all augmentation runs. The rules generated at this optimal setting, particularly involving a new feature called MAP not in the original data, are shown. These rules are statistically validated and are now ready for clinician review. This step is crucial for validating the insights AI has generated autonomously. AI handles the entire process from rule generation to validation without human intervention, which streamlines the analysis and rule-writing tasks that would otherwise be tedious.

VIII. CONCLUSIONS

Our research shows how GANs and CNNs can generate high-quality synthetic data to improve medical data analysis and patient monitoring. We focused on kidney images and used advanced AI techniques to address data scarcity and enhance the performance of our models. Our findings demonstrate the potential of data augmentation in healthcare, highlighting the importance of explainable AI in making machine learning models more transparent and interpretable. Future research should explore further validation, integration of more IoMT devices, and addressing ethical concerns related to synthetic data. Overall, our approach provides a promising framework for leveraging AI to enhance medical data analysis and improve patient outcomes.

REFERENCES

1. Rycroft, C.E.; Heyes, A.; Lanza, L.; Becker, K. Epidemiology of chronic obstructive pulmonary disease: A literature review. Int. J.
Chronic Obstr. Pulm. Dis. 2012, 7, 457. [CrossRef]
2.van Boven, J.F.; Lavorini, F.; Dekhuijzen, P.R.; Blasi, F.; Price, D.B.; Viegi, G. Urging Europe to put non-adherence to inhaled
3.respiratory medication higher on the policy agenda: A report from the First European Congress on Adherence to Therapy. Eur.
Respir. J. 2017, 49, 1700076. [CrossRef]
4.Virchow, J.; Crompton, G.; Dal Negro, R.; Pedersen, S.; Magnan, A.; Seidenberg, J.; Barnes, P.J. Importance of inhaler devices in
the management of airway disease. Respir. Med. 2008, 102, 10–19. [CrossRef]
5. Bennett, J.; Rokas, O.; Chen, L. Healthcare in the Smart Home: A study of past, present and future. Sustainability 2017, 9, 840.
6. Mongelli, M.; Orani, V.; Cambiaso, E.; Vaccari, I.; Paglialonga, A.; Braido, F.; Catalano, C.E. Challenges and Opportunities of IoTand AI in Pneumology. In Proceedings of the 2020 23rd Euromicro Conference on Digital System Design (DSD), Kranj, Slovenia,
26–28 August 2020; pp. 285–292.
7. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions.
Future Gener. Comput. Syst. 2013, 29, 1645–1660. [CrossRef]
8. Paglialonga, A.; Mastropietro, A.; Scalco, E.; Rizzo, G. The mhealth. In EAI/Springer Innovations in Communication and Computing;
Springer: Cham, Switzerland, 2019; pp. 5–17. [CrossRef]
9. Ramponi, G.; Protopapas, P.; Brambilla, M.; Janssen, R. T-cgan: Conditional generative adversarial network for data augmentation
10. in noisy time series with irregular sampling. arXiv 2018, arXiv:1811.08295.
11 Yang, H.; Liu, J.; Zhang, L.; Li, Y.; Zhang, H. ProEGAN-MS: A Progressive Growing Generative Adversarial Networks for10. Kaur, S.; Aggarwal, H.; Rani, R. Data Augmentation Using GAN for Parkinson's Disease Prediction. Lect. Notes Electr. Eng. 2021, 701, 589–597. [CrossRef]
12. Guo,K.; Luo, T.; Bhuiyan, M.; Ren, S.; Zhang, J.; Zhou, D. Zero shot augmentation learning in internet of biometric things for health signal processing. Pattern Recognit. Lett. 2021, 146, 142–149. [CrossRef]