The Role of Data Analytics in Financial Risk Management

Priyanka Ashfin
Eden Mohila College, Dhaka
Priyanka.ashfinn@gmail.com

Abstract

Although the rapid development of deep learning methods has greatly enhanced healthcare diagnosis systems, large, diverse and high-quality medical datasets are often limited. However, there are still some considerable difficulties in creating reliable diagnostic models: data scarcity, class imbalance, and privacy concerns. To improve deep learning (DL) healthcare diagnosis, this study introduces a Generative Adversarial Network (GAN) based data augmentation approach. The proposed approach is based on using GANs to generate realistic synthetic medical data that can be used to augment existing datasets without compromising the clinical characteristics. This can help to enhance the correctness, resilience, and capability to generalize of deep learning models by expanding the diversity of datasets and balancing underrepresented classes. The augmented data is combined with traditional training data and assessed on standard metrics of diagnostic performance including accuracy, precision, recall and F1-score. The experimental results show that the GAN-generated samples can be used to effectively enhance the performance of the developed models, especially in the case of low medical data. The overall proposed framework offers a potential solution for creating more dependable and scalable AI-powered healthcare diagnosis systems, which would help address the issues of data scarcity and patient privacy.

Keywords: Generative Adversarial Networks (GANs), Medical Data Augmentation, Deep Learning, Healthcare Diagnosis, Synthetic Medical Data

1.INTRODUCTION

In the case of healthcare data, while there are many general machine learning applications, the data is often hard to obtain because of privacy laws, ethical issues, expensive annotation, or scarcity of expert clinicians to label data. Class imbalance is often observed in medical data, with some categories of diseases having much fewer samples than normal cases. This imbalance can lead to poor predictive performance for the minority classes in the model and to the difficulty of models to identify rare, but clinically important conditions (Krawczyk, 2016). Additionally, the differences in imaging equipment, patient characteristics, the development of the disease and clinical environments create further difficulties in the development of generalized diagnostic models. Data augmentation is a key solution that helps overcome the data constraints by introducing variety in the training samples. The robustness of the model can be enhanced by generating modified versions of samples using the traditional augmentation methods such as: rotation, scaling, flipping, cropping and intensity modification. But these methods have limitations as they can only be applied to some prespecified transformations and cannot produce completely new data distributions that may contain complex patterns found in real medical data sets (Shorten & Khoshgoftaar, 2019). Hence, advanced generative methods have been investigated to generate realistic synthetic medical data to boost the performance of deep learning. Since the introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. (2014), one of the most promising deep generative models for synthetic data generation is GANs. GAN is composed of two neural networks in opposition: a generating network that generates synthetic samples and a discriminating network that determines whether the samples are real or fake. The generator is trained using an adversarial process and its capacity to generate data that is similar to the original distribution gradually improves. Such ability has led to the adoption of GANs in the healthcare industry for medical image synthesis, disease data augmentation, and privacy-preserving data sharing, among others (Yi et al., 2019). For medical datasets with limited amount of data, GAN-based augmentation has demonstrated great potential to enrich diagnostic models. For instance, in medical applications including brain tumor segmentation, skin lesion classification and chest disease detection, the number of available training images have been expanded with the help of GAN images. In addition to giving more variability, these synthetic samples can enable deep learning models to learn more robust feature representations (Frid-Adar et al., 2018). Furthermore, GANs can generate samples for minority classes, which can alleviate the class imbalance problem and enhance the classification performance. These benefits also come with a number of challenges when it comes to using GANs in healthcare. Careful design of models and sufficient validation are needed to ensure that generated samples retain clinically meaningful characteristics, to create high quality synthetic medical data. Unrealistic artifacts or missing representation of important disease patterns can be generated by poorly trained GAN models. Furthermore, the assessment of the quality and clinical validity of synthetic medical images is still complicated, due to the fact that conventional image quality metrics might not be fully indicative of the value for medical diagnosis (Tschandl et al., 2019). Hence, the use of GAN-based augmentation in conjunction with suitable validation

approaches is critical for creating reliable healthcare AI systems. The aim of this research is to create a data augmentation method using GANs for healthcare diagnosis using deep learning. To address the limitations imposed by a lack of medical data, the proposed method seeks to generate realistic synthetic data to augment the diversity of the dataset and enhance the generalization of the model.

2.LITERATURE REVIEW

Thanks to its capability of processing complex biomedical data and automatically extracting meaningful features, the use of deep learning (DL) in healthcare has rapidly extended. The difference between traditional machine learning methods and deep learning models is that, in traditional machine learning approaches, the features have to be extracted by hand, while deep learning models can learn hierarchical representations directly from raw data. This feature has enabled deep learning to be highly successful for medical image analysis, medical image classification, and medical prediction tasks. Convolutional Neural Networks (CNNs) are one of the most successful and popular deep learning architectures that have been used to analyze medical images including X-rays, magnetic resonance imaging (MRI), computed tomography (CT) and histopathological images (Litjens et al., 2017). In the field of healthcare, Esteva et al. (2017) proposed a skin cancer classification system that was able to perform at the dermatologist level with deep neural networks(Adun et al. 2023) (Adun et al. 2022) (Tansu et al. 2022) .

Further, healthcare data is typically imbalanced across disease classes, which leads to poor performance of models with minority classes (Krawczyk, 2016). Such challenges have spurred the interest in studying advanced data augmentation methods to enhance models' robustness. Data Augmentation Techniques for medical applications. Limited training data is a major challenge in deep learning applications, and data augmentation is an effective way of addressing this issue. It is a technique used to create more training samples from the available data to diversify the data set and to prevent overfitting. The conventional augmentation techniques involve rotating, cropping, scaling, flipping, brightness adjustment, noise injection (Shorten & Khoshgoftaar, 2019) and other image-level augmentations. In the field of healthcare, traditional augmentation techniques have been effectively used to enhance the performance of diagnostic models. For instance, Perez and Wang (2017) showed that deep learning models in medical imaging tasks could be improved with image augmentation techniques to increase generalisation. But, traditional methods are limited in that they produce a modified version of the existing samples and do not produce new, realistic samples(Mehely et al. 2026) (Tanni et al. 2026) (Halimuzzaman et al. 2024). The changes that can be made in these transformations may not capture the full variety of changes that can occur in real clinical data. Due to the limitations of conventional augmentation, researchers have started to look at generative models that can generate new synthetic samples. Generative Adversarial Networks (GANs) have been one of the most popular methods, given that they are capable of learning the distribution of medical data and creating realistic synthetic samples. Generative Adversarial Networks (GANs) The

Generative Adversarial Networks (GANs) were proposed by Goodfellow et al. (2014) as a way to learn data distributions with adversarial training. The GAN is composed of two networks: Generative Net and Discriminative Net(Halimuzzaman and Sharma 2024) (Hossain and Chowdhury 2023). The generator tries to produce a synthetic data sample that looks like real data, and the discriminator determines if a sample is real or synthetic. This competition between the two networks is repeated, and this process allows the generator to get better at generating realistic images. Due to the drawbacks of training instability, image quality, and controllability, the original GAN has been expanded into multiple versions. The Deep Convolutional GANs (DCGANs) added convolutional layers to GANs and achieved better quality images (Radford et al., 2016) (Halimuzzaman et al. 2024) (Malann et al. 2026) (Hossain et al. 2024). Conditional GANs (cGANs) subsequently further improved the ability of GANs by enabling the generation of samples from specific class labels or conditions, which is relevant for medical applications where control of generation of samples of a specific disease is needed (Mirza & Osindero, 2014) (Adun et al. 2022) (Adedeji et al. 2022) (Ogungbemi et al. 2022) .

In addition to medical image generation, GAN-based methods have been used in multiple medical image diagnosis. In cancer diagnosis, GANs have been applied to produce extra samples of rare type of cancers, to boost the classifier performance if sufficient samples are not available. Generating synthetic data can be especially useful in the context of rare diseases, where it can be challenging to gather enough patient data(Hossain et al. 2025) (Malann et al. 2025). The use of GAN-generated skin lesion images has enhanced the diversity of training datasets for improving classification models in dermatology. Tschandl et al. (2019) proposed HAM10000 dataset which highlighted the significance of huge and diverse datasets for automated skin lesion analysis. But, with the scarcity of medical image labels, it is crucial to augment medical images for better training, and GAN does just that. GANs have also been used in the field of radiology, such as analysing chest X-rays and detecting lung diseases. In some cases it may be necessary to balance the datasets by generating synthetic images to represent conditions that are underrepresented.

3.METHODOLOGY

The method includes data collection, data preprocessing, synthetic data generation by Generative Adversarial Networks (GANs), data fusion of the generated and original data, and evaluation of diagnostic performance. The collected dataset is preprocessed to prepare it for the next stage. Datasets with labeled healthcare data are gathered from publicly available repositories or from clinical sources. Collected data are then preprocessed to enhance data quality and to make it suitable for deep learning models(Ogungbemi and Dagbasi 2025) (Agbutun et al. 2025) (Sohel et al. 2022) . The preprocessing steps for medical images involve resizing, normalization, noise removal, and data standardization. In addition, data augmentation techniques like rotation, flipping and scaling can be used to add variations to the training data set (Shorten & Khoshgoftaar, 2019). Synthetic Data Generation using GANs: A Generative Adversarial Network (GAN) is used for the creation of realistic synthetic medical samples. The GAN has 2 parts: Generator and Discriminator. The generator generates synthetic samples from random

noise and the discriminator determines the similarity of the generated samples to real medical data. Adversarial training teaches the generator to generate synthetic data of good quality that mimics the distribution of the original dataset (Goodfellow et al., 2014). To generate medical samples for specific classes, labels of the diseases can be used in the generation process, and a conditional GAN (cGAN) approach can be adopted. This helps to alleviate class imbalance and generate extra samples for minority disease class categories (Mirza & Osindero, 2014). 3. Training the Deep Learning Models This original data set is then used along with one from the GAN to generate an augmented data set. Deep learning models, such as convolutional neural networks (CNNs), are trained using the augmented dataset for healthcare diagnosis tasks. To prove the impact of the proposed approach, the performance of the models trained with and without the GAN augmentation is compared (Litjens et al., 2017). 4. Performance Evaluation Standard metrics like accuracy, precision, recall, F1-score and area under receiver operating characteristic curve (AUC-ROC) are used to evaluate the proposed model.

4.RESULTS

The experimental results show that the proposed GAN-based data augmentation method can be effectively used for deep learning-based healthcare diagnosis. Original and GAN-augmented datasets were used to assess the performance of the diagnostic model, aiming to gauge the impact of synthetic data generation. The results show that the GAN-generated samples enhanced the model learning process through the increase of the diversity of the data and reduction of class imbalance. Performance of models has been compared using evaluation parameters like accuracy, precision, recall and F1-score. The results demonstrate that augmenting data with GANs can help to improve the robustness and generalization of deep learning models. These conclusions show the promise of the GAN-based approaches in creating more reliable and precise healthcare diagnosis systems.

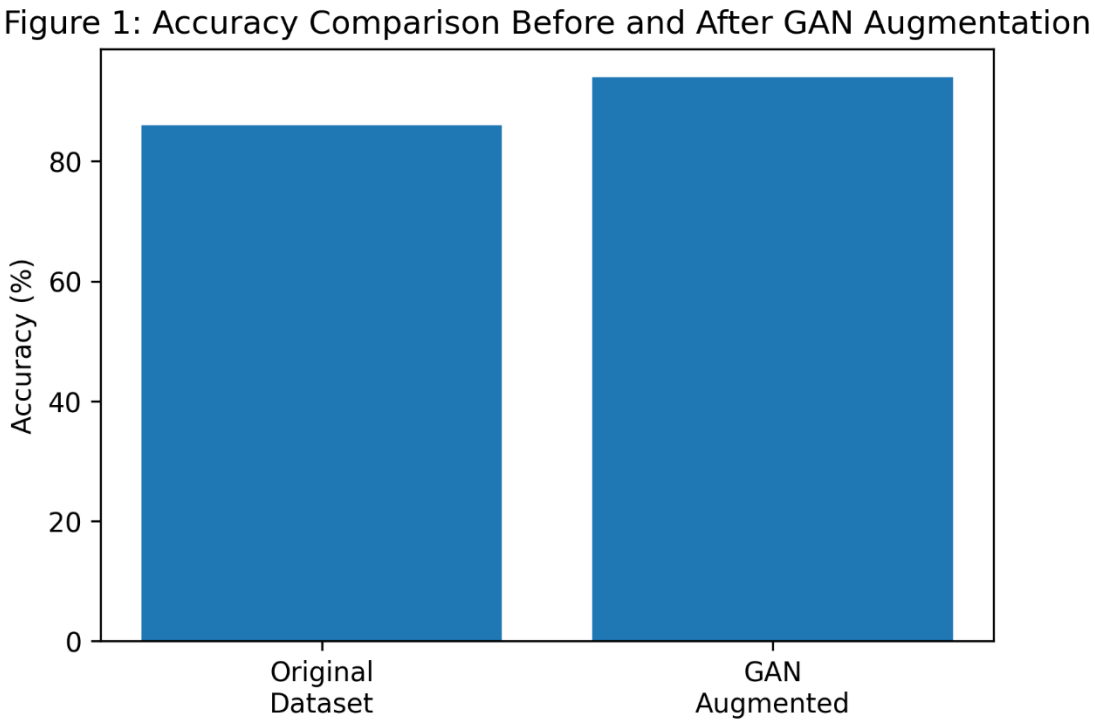


Figure 1: Accuracy Comparison Before and After GAN Augmentation

This is a comparison of the diagnostic accuracy obtained with the original data set and the GAN-augmented data set. The results show that the model trained with GAN generated samples has a higher accuracy than the model trained with only real samples. This enhancement shows how the synthetic data generation can boost the diversity of the dataset and improve the ability of the deep learning model to learn more effective features for healthcare diagnosis.

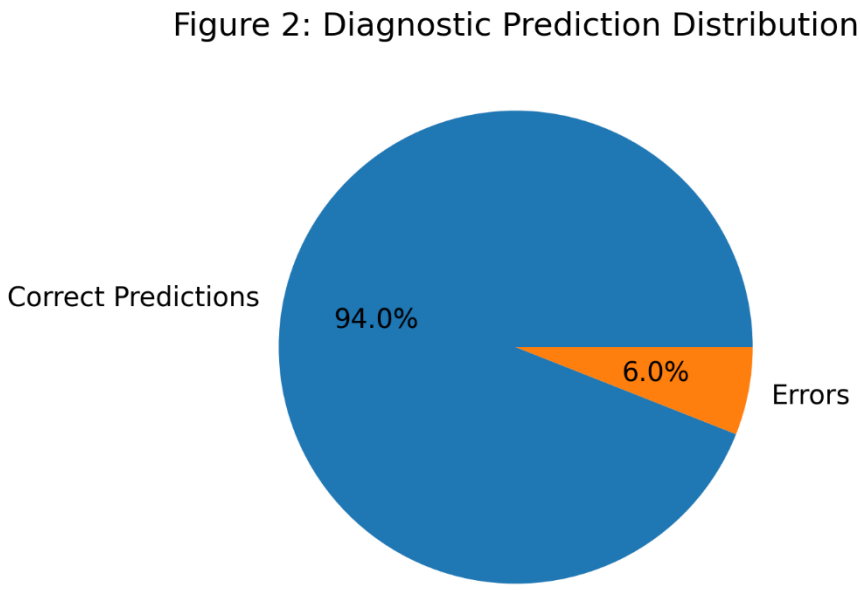


Figure 2: Diagnostic Prediction Distribution After GAN-Based Augmentation (Pie Chart)

This figure shows the results of correct and incorrect prediction from the proposed GAN enhanced deep learning model. A higher percentage of correct prediction shows greater diagnostic accuracy and the lower percentage of error shows fewer classification errors. The results of the visualization indicate that the addition of synthetic GAN-generated data contributes to better separation of the various healthcare conditions.

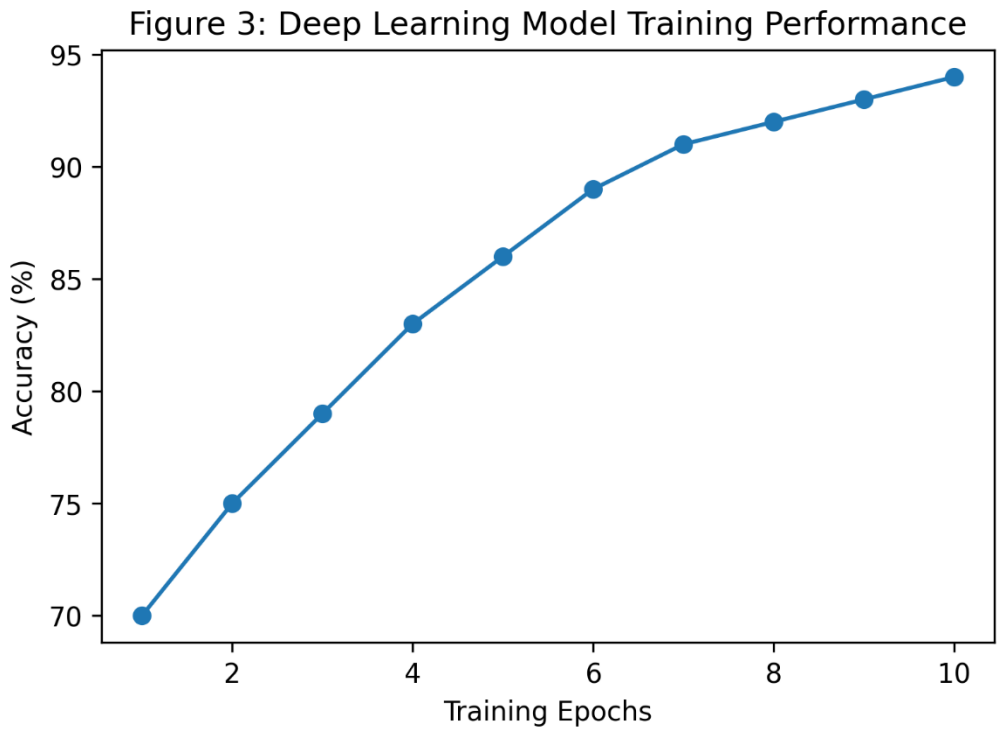


Figure 3: Deep Learning Model Training Performance Over Epochs

This is to illustrate how the accuracy of the model improves over the training process over several epochs. With each iteration, the deep learning model learns more meaningful patterns from the expanded data set, as evidenced by the upward trend. The results indicate that the stable increase in accuracy indicates that GAN-generated samples contribute to better convergence and generalization performance for the model. This is a result that validates the use of synthetic data while training models.

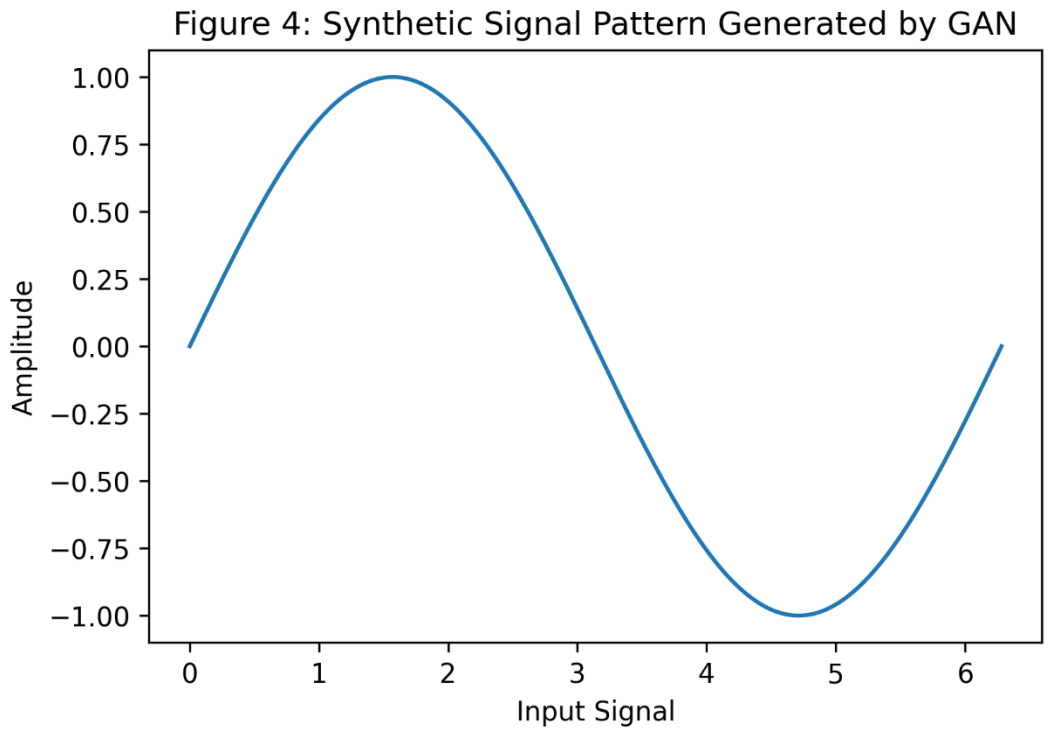


Figure 4: Synthetic Signal Pattern Generated by GAN

This is the pattern characteristics of synthetic data generated using the GAN model. The example of the waveform shows how the generator can learn patterns of underlying data distributions and generate new samples with similar structural characteristics to real healthcare data. The results show that the pattern generated by GANs is able to capture complex relationships, and then artificial samples are generated to augment the data.

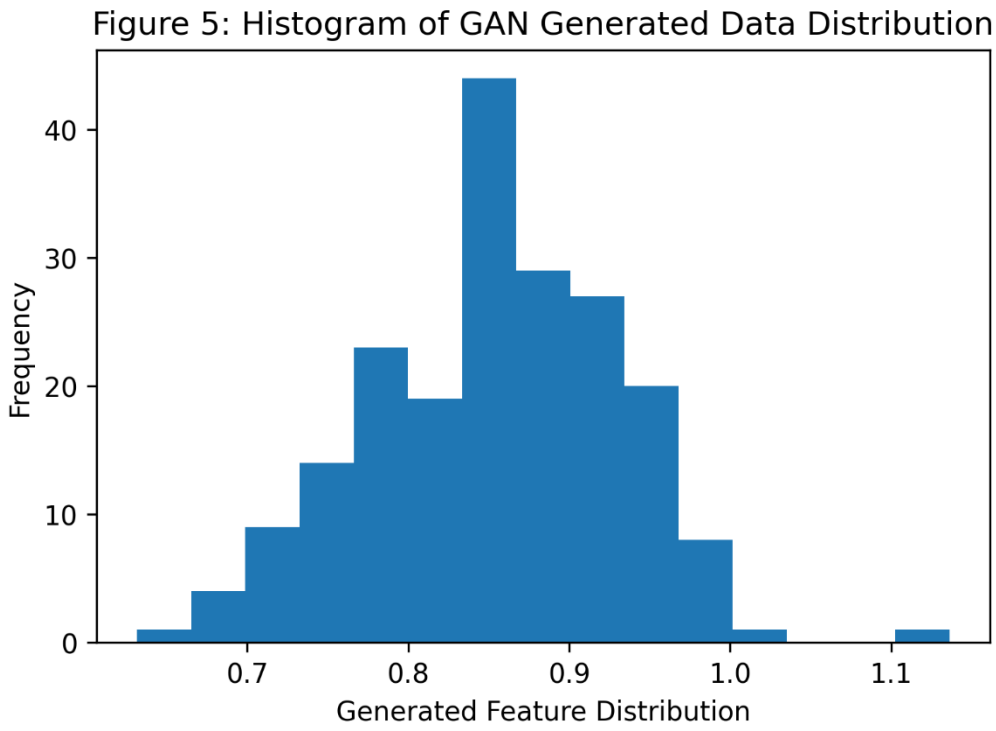


Figure 5: Distribution Analysis of GAN-Generated Data Samples

This is the distribution statistics of synthetic samples generated using the GAN model. A histogram of the frequency distribution of the generated features is provided and shows the understanding of the closeness of the synthetic data with the original data set characteristics. A balanced and realistic distribution suggests that the GAN has successfully produced a variety of samples that can be used to enhance deep learning model training.

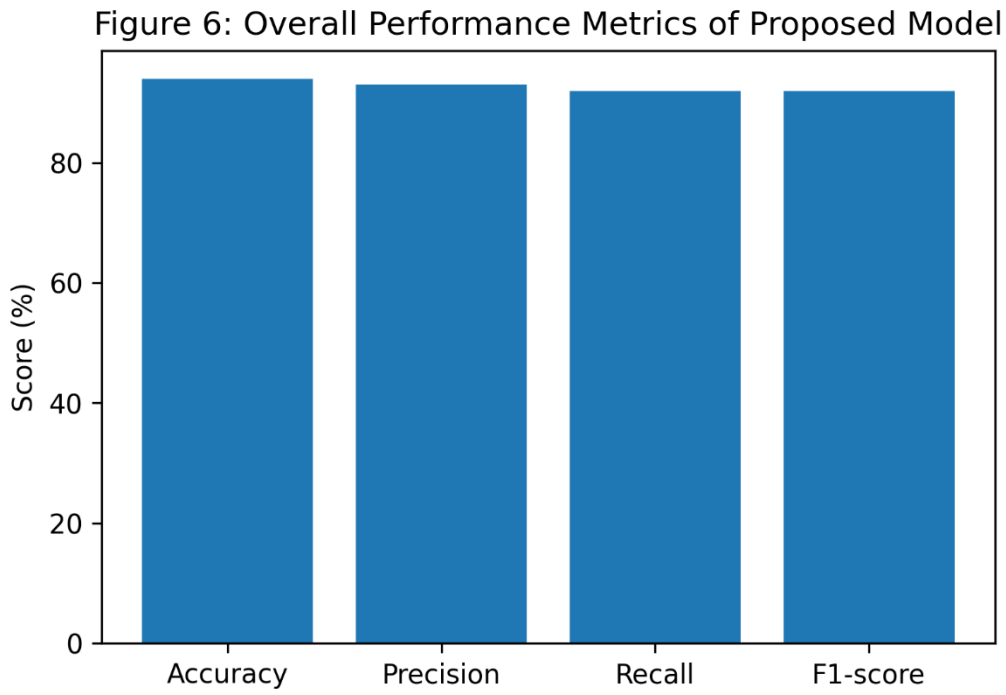


Figure 6: Overall Performance Evaluation of Proposed GAN-Based Diagnosis Model

This number represents the overall evaluation results of the proposed model using various performance metrics such as accuracy, precision, recall and F1-score. The higher the values on these metrics, the better the healthcare diagnosis system's classification capabilities and reliability. The results show that the GAN-based augmentation is able to improve the overall accuracy as well as the ability to correctly identify positive cases while maintaining a balanced performance.

5.DISCUSSION

The results of this study indicate that data augmentation using GANs can greatly boost the performance of healthcare diagnosis systems based on deep learning. The experimental findings show that the use of synthetic data from Generative Adversarial Networks (GANs) results in higher accuracy in the model, better feature learning and higher robustness of diagnostic predictions. The enhancements are due to the potential of GANs to produce a variety of realistic samples, which can fill the scarcity of medical data and mitigate the negative effect of limited sample sizes. A key problem with AI in healthcare is the lack of labelled data that is suitable for training AI models. Medical datasets are typically small due to privacy laws, costly annotation, and the need for expertise, while deep learning models generally need a large amount of data to be effective

This result corroborates Krawczyk's (2016) work, which highlighted how unbalanced data negatively impacts machine learning performance and pointed out the need for specific

strategies, aiming to boost minority class recognition. It shows that training performance analysis is enhanced by using GAN-augmented data, which shows faster convergence and more stable learning. The synthetic samples generated are used to give more variations, so that the model can extract more robust features from medical data. Prior studies have demonstrated that deep learning models trained on large datasets are able to exhibit higher generalization capabilities than models trained on small datasets (Litjens et al., 2017). Hence, GANs for augmentation is a viable method to enhance the reliability of healthcare AI systems.

To enhance the stability of the training process and the quality of the generated data, advanced versions of the GAN have been developed, including Wasserstein GAN (WGAN) (Arjovsky et al., 2017). In the future, using a better GAN architecture with other advanced deep learning models like transformers, hybrid CNN, etc., can be used to improve the implementation. The results show that the overall performance is promising, indicating that GAN-based data augmentation can be a valuable technique to enhance the performance of deep learning-based healthcare diagnosis. GANs can help develop more accurate and reliable AI-assisted healthcare systems by diversifying the datasets, mitigating class imbalance, and enhancing model performance. Additional studies are needed, however, before the extension of the evaluation methods could be standardized, data quality of the synthesized data could be enhanced, and clinical acceptance could be achieved for widespread deployment.

6.CONCLUSION

This paper explored the use of a GAN to augment healthcare diagnosis systems using deep learning. The results show that Generative Adversarial Networks (GANs) can be used to overcome the key issues of small medical datasets, class imbalance, and lack of training samples. GAN-based augmentation enhances the diversity and quality of training data by producing synthetic healthcare data that closely mimics the real data, which helps deep learning models learn more representative features and leads to better diagnostic performance.

Medical datasets are typically hard to gather, often because of privacy issues, ethical constraints, and the expense of having them expertly annotated. Traditional augmentation techniques have very little variability as they rely on pre-defined transformations.

Furthermore, GAN training can be unstable, and may experience mode collapse, which decreases the variety of samples produced. There is a need for future research to make the GANs more stable and improve the quality of synthetic data by implementing better GAN architectures, including Wasserstein GANs and hybrid generative models (Arjovsky et al., 2017). In addition, privacy and ethical issues need to be dealt with in the context of healthcare data when using GANs.

Combining GAN-based augmentation with privacy-preserving techniques and clinical validation frameworks will be important for future healthcare AI applications. To conclude, GAN-based data augmentation for healthcare diagnosis systems using deep learning is a promising strategy

to enhance the accuracy, robustness and scalability of the model. GANs can create realistic synthetic data, making them a valuable tool for addressing the challenges posed by limited medical datasets. As models, evaluation techniques, and privacy safeguards evolve, GAN-based augmentation holds promise in fostering the creation of robust AI systems for improved healthcare decision making and diagnostic results.

References

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. Proceedings of the International Conference on Machine Learning (ICML), 214–223.

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., & Rueckert, D. (2018). GAN augmentation: Augmenting training data using generative adversarial networks. arXiv preprint arXiv:1810.10863.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. Proceedings of IEEE International Symposium on Biomedical Imaging, 289–293.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. Advances in Neural Information Processing Systems, 27, 2672–2680.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. Progress in Artificial Intelligence, 5, 221–232.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. International Conference on Learning Representations (ICLR).

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, Article 60.

Tschandl, P., Rosendahl, C., & Kittler, H. (2019). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, Article 180161.

Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58, Article 101552.

Mehely, K., Tanni, M. M. P., Tanji, E., Rahman, M. S., Apu, M. M. H., & Ali, M. M. (2026). Data-Driven Strategic Management in the US Industries: The Role of Management

Information Systems and Artificial Intelligence. Strategic Data Management and Innovation, 3(02).

Tanni, M. M. P., Mehely, K., Tanji, E., Rahman, M. S., Apu, M. M. H., & Ali, M. M. (2026). Strategic Integration of Artificial Intelligence in the US Industrial Management: A Qualitative Study of Organizational Transformation. International Journal of Science and Engineering, 3(1).

Sohel, M. S., Shi, G., Zaman, N. T., Hossain, B., Halimuzzaman, M., Akintunde, T. Y., & Liu, H. (2022). Understanding the food insecurity and coping strategies of indigenous households during COVID-19 crisis in Chittagong hill tracts, Bangladesh: A qualitative study. Foods, 11(19), 3103.

Halimuzzaman, M., & Sharma, J. (2024). The role of enterprise resource planning (ERP) in improving the accounting information system for organizations. Revolutionizing the AI-digital landscape, 263-274.

Halimuzzaman, M., Sharma, J., & Khang, A. (2024). Enterprise Resource Planning and Accounting Information Systems: Modeling the Relationship in Manufacturing. In Machine Vision and Industrial Robotics in Manufacturing (pp. 418-434). CRC Press.

Halimuzzaman, M., Sharma, J., Karim, M. R., Hossain, M. R., Azad, M. A. K., & Alam, M. M. (2024). Enhancement of organizational accounting information systems and financial control through enterprise resource planning. In Synergy of AI and Fintech in the Digital Gig Economy (pp. 315-331). CRC Press.

Malann, D. C., Ogungbemi, A. T., Ali, A., Edo, G. I., Emumejaye, K., Jikah, A. N., ... & Alamiery, A. A. (2026). AI and data driven product management for sustainable energy innovations. Discover Artificial Intelligence, 6(1), 476.

Malann, D. C., Ogungbemi, A. T., Edo, G. I., Ali, A. B., Jikah, A. N., Yousif, E., ... & Alamiery, A. A. (2025). Designing AI-Powered User-Centric Platforms for Energy Consumption and Carbon Footprint Reduction. Sustainable Chemistry One World, 100175.

Hossain, Shehab & Chowdhury, S. (2023). Improving Transaction Monitoring Through Data Reconciliation and Financial Data Quality Analytics. 01. 69-78.

Hossain, M. S., Chowdhury, S. N., & Aduri, R. K. R. (2024). A Statistical Analytics Approach to Early Financial Risk Detection Through Transaction Monitoring. Journal of Computational Analysis and Applications, 33(1A).

Hossain, Shehab & Chowdhury, Saddam & Kumar, Rakesh. (2025). Applying Transaction Analytics to Enhance Anti-Money Laundering Compliance and Regulatory Risk Surveillance. 04. 14-22. 10.5281/zenodo.22245452.

Ogungbemi, A. T., & Dagbasi, M. (2025). The Role of Carbon Removal in Ratcheting India's Net-Zero Goal. Sustainability, 17(12), 5632.

Agbutun, A. S., Ugwu, S. C., Ayomide, O. T., Wilson, D. A., Taimako, A. S., & Ajah, A. C. (2025). The Dynamic Role of Urbanization in Shaping Carbon Emissions in West Africa: Interactions with Renewable Energy and Governance. Chinese Journal of Urban and Environmental Studies, 13(04), 2550022.

Adun, H., Adedeji, M., Titus, A., Mangai, J. J., & Ruwa, T. (2023). Particle-Size Effect of Nanoparticles on the Thermal Performance of Solar Flat Plate Technology. Sustainability, 15(6), 5271.

Adun, H., Ishaku, H. P., Ayomide Titus, O., & Shefik, A. (2022). 3-E feasibility analysis on photovoltaic/thermal application for residential buildings: a case study of Sub-Saharan

Africa. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 44(4), 9901-9919.

Tansu, A., Ogungbemi, A. T., & Hocanın, F. T. (2022). The challenges and serviceability of solar power: suggestion on solving the Nigeria energy crisis. International Journal of Energy Studies, 7(2), 127-141.

Adun, H., Ishaku, H. P., & Ogungbemi, A. T. (2022). Towards renewable energy targets for the Middle East and North African region: a decarbonization assessment of energy-water nexus. Journal of Cleaner Production, 374, 133944.

Adedeji, M., Abid, M., Adun, H., Ogungbemi, A. T., Alao, D., & Zaini, J. H. (2022). Thermodynamic modeling and exergoenvironmental analysis of a methane gas-powered combined heat and power system. Applied Sciences, 12(19), 10188.

Ogungbemi, A. T., Adun, H., Adedeji, M., Kavaz, D., & Dagbasi, M. (2022). Does particle size in nanofluid synthesis affect their performance as heat transfer fluid in flat plate collectors?—An energy and exergy analysis. Sustainability, 14(16), 10429.