****

**Unlocking Automatic Diagnostic Potential: AI-Generated Data for Enhanced Chronic Kidney Disease Classification with ML Models**

Abdelrhman Salah1, Ahmed Sayed1, Muhammad Megahed1, Youssef ElNaggar1, Ahmed Elshemy1, and Nermin Negied1,2

*1School of Information Technology and Computer Science, Nile University, Giza, Egypt*

*2Ministry of Communication and Information Technology, Giza, Egypt*

\* Corresponding author’s Email: [nnegied@nu.edu.eg](mailto:nnegied@nu.edu.eg)

**Abstract:** chronic kidney disease (CKD) is a progressive condition characterized by the gradual loss of kidney function, leading to significant health issues and mortality worldwide. Early and accurate classification of CKD is essential for effective intervention and management. The existence of medical data became very crucial in the digital era to facilitate the early diagnosis, classification and even prediction of disease, however, the availability of medical data is always rare because of many reasons. Amongst those reasons are the confidentiality of the patient’s data, some challenges in recording and keeping accurate data records, and the lack of support of physicians and medical institutions in gathering and providing data. This research explores the potential of advanced AI models, specifically ChatGPT-3.5, GPT-4, and GPT-4o, in generating synthetic datasets that closely resemble real-world CKD data. By comparing the data generation capabilities of these models, we aim to identify the most accurate and realistic datasets for training machine learning (ML) models in CKD classification. Leveraging AI to create high-quality synthetic data addresses the scarcity of accessible medical datasets due to stringent privacy regulations. Results demonstrate that GPT-4, the latest iteration, exhibits superior performance in synthesizing data closely resembling real-world CKD datasets, thereby significantly enhancing the performance of various ML models, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). These findings underscore the transformative potential of AI-generated data in revolutionizing medical diagnostics, with implications for improving patient outcomes and reducing the global healthcare burden of CKD.

**Keywords:** chronic kidney disease, Machine Learning, Generative AI, Synthetic Data, Tabular Data Generation.

1. **Introduction**

Chronic kidney disease (CKD) is a significant global health issue, affecting millions of people worldwide. According to the World Health Organization (WHO), CKD is the 12th leading cause of death and the 17th cause of disability globally [1]. (CKD) is a long-term condition characterized by a gradual loss of kidney function over time. The kidneys, which are vital organs responsible for filtering waste and excess fluids from the blood, become damaged and lose their ability to function effectively. This deterioration can lead to a buildup of waste products in the blood, causing various health issues, including high blood pressure, anaemia, weak bones, poor nutritional health, and nerve damage. The disease progresses through five stages, with the final stage, end-stage renal disease (ESRD), requiring dialysis or a kidney transplant for patient survival. Early detection and treatment are crucial to slow progression and improve quality of life, making it imperative to develop efficient methods for diagnosing and monitoring CKD. Despite advances in medical science, CKD often remains undiagnosed until it reaches an advanced stage, underscoring the need for better diagnostic tools and greater public awareness of the disease.

Early and precise classification of CKD can lead to better patient outcomes through timely intervention, personalized treatment plans, and improved disease management. Furthermore, understanding the disease's progression and risk factors can help in the development of preventive measures, reducing the overall burden of CKD on healthcare systems.

Machine Learning (ML) has emerged as a powerful tool in the medical field, offering significant potential in diagnosing and classifying many diseases including chronic conditions like CKD [2 - 5]. By analysing large datasets, ML algorithms can identify patterns and correlations that may not be apparent through traditional methods. However, in the medical field, the collection of large, high-quality tabular data is challenging due to stringent security and privacy regulations protecting patient information. This scarcity of accessible medical datasets hinders the training and validation of ML models. AI's potential to generate synthetic data resembling real patient data can help bridge this gap, enabling researchers and practitioners to develop and refine ML algorithms without compromising patient privacy. By using AI-generated data, researchers can simulate various clinical scenarios, conduct extensive testing, and improve the robustness of their models. This approach not only addresses the data scarcity issue but also enhances the overall quality and applicability of ML solutions in healthcare.

This research aims to explore the efficacy of different AI models (ChatGPT-3.5, GPT-4, GPT-4o) in generating synthetic datasets for CKD that closely mimic real-world data. By comparing the performance of these models in terms of data generation, we seek to identify which AI version provides the most accurate and realistic datasets.

The rest of this paper is organized as follows; Section 2 surveys the related work done in literature, Section 3 demonstrates the proposed approaches with detailed explanation of architectures and methodologies. Section 4 discusses the findings and analyses results. Section 5 concludes the paper and suggests some future work directions.

1. **Related Work**

**2.1. ML in early detection of CKD**

Ekanayake and Damayanthi [6] focused on how domain knowledge significantly enhances the accuracy of ML classifiers like Extra Trees and Random Forest. Their research emphasized the importance of informed feature selection, showing that integrating clinical insights during data preprocessing can greatly enhance the model's ability to detect CKD at early stages.

Al Mustafa [7] investigated the efficacy of several ML classifiers, including J48 and K-Nearest Neighbors, on clinical datasets composed of blood and urine test results. The study demonstrated that advanced feature selection methods drastically improve diagnostic accuracy, with some classifiers achieving up to 99.75%. This finding suggests ML's potential to reduce the number of diagnostic tests needed, thereby lowering healthcare costs.

Pankaj Chittora et al. [8] analyzed the performance of several classifiers on a robust dataset from the UCI repository. They highlighted the Linear Support Vector Machine with an L2 penalty as particularly effective, achieving the highest prediction accuracy of 98.86%. The use of balancing techniques and sophisticated feature selection methods, like the Wrapper method and LASSO regression, was crucial in achieving these results.

Abdel-Fattah et al. [9] utilized hybrid ML techniques along with Apache Spark to enhance the prediction accuracy of CKD. They applied feature selection methods like Relief-F and chi-squared with various classifiers to achieve perfect accuracy scores, showcasing how big data technologies can effectively handle large healthcare datasets.

Islam et al. [10] employed 12 different ML classifiers, including XGBoost, within a supervised learning framework to predict CKD. The study began with a comprehensive dataset, optimizing it through intensive preprocessing and feature selection to achieve an accuracy of 98.3%. This emphasizes the importance of predictive modeling in improving the selection process for attributes that significantly enhance the accuracy of the ML models used.

Maurya et al. [11] focused on predicting CKD stages and providing dietary recommendations based on potassium levels. Their approach, using data from the UCI repository, not only predicted CKD stages with high accuracy but also offered actionable insights for disease management.

Debal and Sitote [12] applied various ML algorithms to predict CKD stages using data from St. Paulo’s Hospital. Their findings highlighted the Random Forest model, optimized with Recursive Feature Elimination, as particularly effective, showcasing the critical role of ML in enhancing early detection in public health settings.

Amirgaliyev et al. [13] and Charleonnan et al. [14] further validated the effectiveness of SVM and other classifiers in CKD prediction. These studies demonstrated over 93% accuracy, emphasizing the potential of non-invasive ML techniques in resource-limited settings.

Gangani Dharmarathne et al. [15] highlighted the use of explainable AI tools, like SHAP and PDP, to aid healthcare professionals in interpreting ML predictions, thereby improving diagnostic accuracy and supporting decision-making.

**2.2. Generative AI for generating medical data**

Elkholy et al. [16] focused on the security and privacy concerns in the development and deployment of generative AI. They highlighted how these technologies could expose vulnerabilities at various stages from data collection to clinical implementation, advocating for stringent security measures to ensure their responsible use.

Rahat et al. [17] examined the potential of generative AI to produce lifelike simulations for medical training, providing a risk-free environment for students and professionals. This study showcased generative AI's capability to create detailed models of various medical conditions, which could significantly aid in medical education and practice readiness.

Xanthis et al. [18] introduced a cost-effective method for creating training datasets for medical imaging using advanced MR simulations based on the 4D-XCAT model. This technique, which generated artificial MR images for training neural networks, significantly reduced the need for real patient data, achieving high Dice coefficients for automatic segmentation. The inclusion of a small percentage of real images further enhanced performance, demonstrating a scalable approach to data generation in medical imaging.

Chen and Esmaeilzadeh [19] explored the broader applications and implications of generative AI across various healthcare domains including medical diagnostics and clinical decision support. They emphasized the technology's impact on improving accuracy and efficiency but also highlighted the privacy and security risks associated with generative AI systems. Their study proposed robust security measures such as encryption and differential privacy to mitigate these risks, underscoring the importance of comprehensive security in the deployment of generative AI technologies.

Singh [20] discussed the transformative applications and challenges of generative AI in healthcare, particularly in medical education and clinical diagnostics. The study pointed out the potential of generative AI, using techniques like GANs and VAEs, to enhance medical training with realistic simulations and improve early disease detection. However, challenges related to data quality, integration, and ethical considerations were noted as significant hurdles.

1. **MATERIALS & METHODS**

This section explains the proposed approach, how and why it has been selected, the dataset used, and how it was prepared and processed. The following subsections demonstrates the prosed work in this paper step by step.

* 1. **Study Population and Inclusion Criteria**

The study population comprises individuals suspected of CKD, with inclusion criteria extending across all age groups. Prior to participation, individuals provided informed consent. Notably, the study maintained stringent exclusion criteria to ensure the integrity of the dataset, specifically excluding individuals with incomplete medical records.

**3.2. Model Architecture**

First, input data used in this model were gathered from two different sources, real data from Kaggle [23], and artificially generated data for every GPT [24]. Secondly, Data preprocessing is done in 4 states: EDA, Data Cleaning, Ordinal Encoding, and Normalization to enhance the quality of the data before splitting. Then data splitting step to split the dataset into training set and testing set. Afterwards ML Classification models such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multiple Layer perceptron (MLP) were built for disease classification. Finally, Models evaluation was done using Precision, Recall, F1-Score, and accuracy. Many validation techniques were used to validate results, such as Stratified K-Fold, Leave-One-Out, and Shuffle Split. Fig.1 represents the detailed steps architecture of the proposed work. The following subsections furtherly explain the details of every step.Fig.1 represents the detailed steps architecture of the proposed work. The following subsections furtherly explain the details of every step.

A diagram of a data processing process

Description automatically generated

Figure.1. The architecture and steps of the proposed work

**3.2.1. Data Collection and Sources**

A comprehensive data collection strategy was implemented, gathering information from a variety of sources such as Electronic Health Records (EHR), laboratory tests, patient surveys, and clinical assessments. This multifaceted approach ensured the collection of a wide range of data, including demographic details, clinical metrics, lifestyle factors, and detailed medical histories. For the research, real data was sourced from the "[kidney\_disease.csv](https://www.kaggle.com/code/youssef22ashraf/kidney-disease-with-chatgpt3-5-dataset/input)" dataset on Kaggle [23], complemented by artificially generated data for every Generative Pre-trained Transformer (GPT), GPT-3.5, GPT-4.0, and GPT-4o were used in the study [24]. This combination of real and synthetic data enhanced the robustness and depth of the analysis.

**3.2.2. Data Preprocessing**

Prior to analysis, meticulous preprocessing steps were done to optimize data quality and compatibility with ML algorithms. These steps included imputation techniques to handle missing data, normalization of continuous variables to enhance comparability, and categorical encoding to facilitate seamless integration into machine learning frameworks. The machine learning algorithms applied in the analysis included Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-Layer Perceptrons (MLP).

**3.2.3 Machine Learning Algorithms**

We utilized five machine learning models for classification tasks using Scikit-learn: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). Logistic Regression operates as a linear model for binary classification, with max\_iter=100 set by default. Decision Tree employs recursive partitioning based on feature values. Random Forest aggregates results from multiple decision trees, utilizing default settings. SVM seeks optimal hyperplanes with probability=True and defaults for other parameters. MLP employs a feedforward neural network with max\_iter=100 as the default setting. Data was divided into training and testing sets (70:30) and evaluated using precision, recall, F1-score, accuracy, and AUC metrics across Stratified K-Fold, Leave-One-Out, and Shuffle-Split cross-validation methods. Random Forest generally demonstrated the highest accuracy and F1-score across these validation techniques.

**3.2.4. Data Augmentation and Synthetic Data Generation**

To mitigate the potential risk of overfitting arising from the limited size of the real dataset, innovative techniques for data augmentation and synthetic data generation were employed. Leveraging advanced language models such as GPT-3.5, GPT-4, and GPT-4o, synthetic data was meticulously crafted to closely emulate the statistical properties of the original dataset, thereby preserving fidelity to the underlying data distribution. The architecture workflow of every AI generation tool that standardize the process for every GPT version from GPT3.5, GPT-4, and GPT-4o as shown in figures 2.

A diagram of data processing

Description automatically generated

Figure.2. The architecture workflow of the dataset generated from every GPT version.

**3.2.5. Model Training and Evaluation**

The combined datasets were judiciously partitioned into training and testing subsets to facilitate robust model training and evaluation. Model performance was rigorously assessed using k-fold cross-validation, allowing for a comprehensive evaluation of metrics including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). This comprehensive approach provided valuable insights into the effectiveness of the developed predictive models.

1. **RESULTS AND DISCUSSION**

This study explores the efficacy of using AI models, specifically GPT-3.5, GPT-4.0, and GPT-4o, in generating synthetic data to enhance the performance of various machine learning models in diagnosing kidney disease. Each model's performance was assessed using three different cross-validation techniques: Stratified K-Fold, Leave-One-Out, and Shuffle-Split.

The results section presents a comprehensive comparison of the average AUC scores obtained from the original dataset and those generated by the afore mentioned AI models. The following tables measure performance improvements across various machine learning models, highlighting the advantages of integrating AI-generated data into the training process.

As can be seen in Table 2 among the GPT-generated datasets, GPT-4 yields the results most like those obtained with the original dataset. The GPT-4 dataset shows high performance across all models, particularly with Logistic Regression (LR) F1(original:0.9349,GPT-4:0.9327) and MLP model F1Score(original:0.8851,GPT-4:0.8252),which both achieved perfect scores in all evaluation metrics (Precision, Recall, F1-Score, Accuracy).The second closest dataset is GPT-4o, which also performs well, particularly with Logistic Regression (LR) F1(original:0.9349,GPT-4O:0.8424)and MLP F1(original:0.8851,GPT4O:0.7721)demonstrating high precision, recall, F1-score, and accuracy values close to those of the original dataset. GPT-3.5 is the third closest, showing less performance lower compared to GPT-4 and GPT-4o, especially with Logistic Regression (LR) F1(original:0.9349,GPT-3.5:0.8038)MLP F1(original:0.8851,GPT3.5:0.7433. Based on these findings, we suggest using the GPT-4 dataset for generating chronic kidney disease (CKD) data when there is a lack of sufficient data instances, as it closely mimics the characteristics of the original dataset.

TABLE I. Cross Validation across Datasets for every ML models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Data Source | Stratified K-Fold | Leave-One-Out | Shuffle-Split |
| LR | Original | 0.94 | Nan | 0.98 |
|  | GPT-3.5 | 0.91 | 0.91 | 0.91 |
|  | GPT-4.0 | 0.96 | 0.96 | 0.96 |
|  | GPT-4o | 0.98 | 0.98 | 0.98 |
| DT | Original | 0.93 | Nan | 0.97 |
|  | GPT-3.5 | 0.99 | 0.99 | 0.99 |
|  | GPT-4.0 | 0.98 | 0.98 | 0.98 |
|  | GPT-4o | 1.0 | 1.0 | **1.0** |
| RF | Original | 1.0 | Nan | **1.0** |
|  | GPT-3.5 | 1.0 | 1.0 | **1.0** |
|  | GPT-4.0 | 0.99 | 0.99 | 0.99 |
|  | GPT-4o | 1.0 | 1.0 | **1.0** |
| SVM | Original | 0.69 | Nan | 0.53 |
|  | GPT-3.5 | 0.89 | 0.89 | 0.89 |
|  | GPT-4.0 | 0.82 | 0.82 | 0.82 |
|  | GPT-4o | 0.63 | 0.63 | 0.63 |
| MLP | Original | 0.53 | Nan | 0.60 |
|  | GPT-3.5 | 0.89 | 0.89 | 0.89 |
|  | GPT-4.0 | 0.96 | 0.96 | 0.96 |
|  | GPT-4o | 0.96 | 0.96 | 0.96 |

TABLE II. Evaluation Metrices among every Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data source | Model | Evaluation Metrices | | | |
| Precision | Recall | F1-Score | Accuracy |
| Original | LR | 0.94 | 0.92 | 0.93 | 0.94 |
| DT | 1.0 | 0.97 | 0.98 | 0.98 |
| RF | 1.0 | 0.98 | 0.99 | 0.99 |
| SVM | 0.93 | 1.0 | 0.96 | 0.97 |
| MLP | 0.86 | 0.90 | 0.88 | 0.90 |
| GPT-3.5 | LR | 0.90 | 0.72 | 0.80 | 0.90 |
| DT | 0.98 | 0.93 | 0.95 | 0.97 |
| RF | 1.0 | 0.97 | 0.98 | 0.99 |
| SVM | 0.98 | 0.94 | 0.96 | 0.98 |
| MLP | 0.66 | 0.83 | 0.74 | 0.84 |
| GPT-4 | LR | 0.91 | 0.95 | 0.93 | 0.92 |
| DT | 1.0 | 1.0 | 1.0 | 1.0 |
| RF | 1.0 | 1.0 | 1.0 | 1.0 |
| SVM | 0.98 | 0.98 | 0.98 | 0.97 |
| MLP | 0.87 | 0.78 | 0.82 | 0.81 |
| GPT-4o | LR | 0.87 | 0.81 | 0.84 | 0.89 |
| DT | 0.97 | 0.97 | 0.97 | 0.98 |
| RF | 1.0 | 0.97 | 0.98 | 0.99 |
| SVM | 0.97 | 0.98 | 0.98 | 0.98 |
| MLP | 0.97 | 0.63 | 0.77 | 0.87 |

Furthermore, all the previous papers used in the same dataset with same percentage of splitting, also with almost the same ML models. Table 3 compares the results of the proposed work with the two datasets (Original, GPT-4 generated data) with the results of the previous work in literature.

TABLE III. Comparison between our real dataset and best GPT generated dataset evaluation results with some of previous studies dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| Data source | Model | Evaluation Metrices | |
| F1-Score | Accuracy |
| [kidney\_disease.csv](https://www.kaggle.com/code/youssef22ashraf/kidney-disease-with-chatgpt3-5-dataset/input) (Original dataset) | L | 0.93 | 0.94 |
| DT | 0.98 | 0.98 |
| RF | 0.99 | 0.99 |
| SVM | 0.96 | 0.97 |
| MLP | 0.88 | 0.90 |
| GPT-4 (Best generated dataset) | LR | 0.93 | 0.92 |
| DT | **1.0** | **1.0** |
| RF | **1.0** | **1.0** |
| SVM | 0.98 | 0.97 |
| MLP | 0.82 | 0.81 |
| Ekanayake & Herath [6] | LR | 0.96 | 0.95 |
| DT | 1.0 | 1.0 |
| RF | 1.0 | 1.0 |
| SVC (rbf) | 0.96 | 0.95 |
| Classical NN | 0.98 | 0.97 |
| Almustafa [7] | KNN | 0.96 | 0.95 |
| DT | 0.99 | 0.99 |
| RF | 0.95 | 0.95 |
| NB | 0.95 | 0.95 |
| SGD | 0.98 | 0.98 |
| Chittora et al. [8] | LR | 0.71 | 0.71 |
| KNN | 0.73 | 0.64 |
| RF | 0.88 | 0.90 |
| LSVM | 0.92 | 0.94 |
| ANN | 0.92 | 0.94 |
| Abdel-Fattah et al. [9] | LR | 0.99 | 0.99 |
| DT | 0.95 | 0.95 |
| RF | 1.0 | 1.0 |
| SVM | 1.0 | 1.0 |
| GBT | 0.95 | 0.95 |
| Islam et al. [10] | KNN | 0.59 | 0.59 |
| DT | 0.97 | 0.97 |
| RF | 0.97 | 0.97 |
| SVM | 0.97 | 0.96 |
| ANN | 0.45 | 0.60 |
| Debal & Sitote[12] | DT | 0.76 | 0.77 |
| RF | 0.77 | 0.78 |
| SVM | 0.59 | 0.63 |
| Charleonnan et al. [14] | LR | - | 0.96 |
| DT | - | 0.94 |
| KNN | - | 0.98 |
| SVM | - | 0.98 |
| Dharmarathne et al. [15] | KNN | 0.82 | 0.80 |
| DT | 0.98 | 0.97 |
| RF | 0.98 | 0.97 |
| SVM | 0.85 | 0.84 |
| ANN | 0.97 | 0.96 |
| Rahat et al. [17] | LR | 0.92 | 0.90 |
| XGBoost | 0.95 | 0.94 |
| RF | 0.91 | 0.93 |
| AdaBoost | 0.94 | 0.91 |
| Hybrid Model | 0.96 | 0.95 |

From table 3, GPT-4's results are generally close to those found in previous work done in this area. GPT-4 achieved best performance in kidney disease classification using Decision Tree (DT) and Random Forest (RF). For other models, such as Logistic Regression (LR) and Support Vector Machine (SVM), GPT-4's performance is also competitive, often matching or slightly exceeding the results in state-of-the-art studies. Overall, it can be concluded that GPT-4 has the highest potential to generate trusted dataset which is very similar to real datasets.

Moreover, from Table III the original results are better and more generalized compared to many previous studies. They consistently show high performance across various models, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP), with strong metrics in F1-score, and accuracy. For instance, the original DT and RF models achieve nearly perfect scores, which are among the highest seen in the literature. This consistency contrasts with the variability in previous studies, where some report perfect scores for certain models while others show significantly lower performance. The robustness and high performance of the original results across multiple models indicate superior generalization compared to the often-varied results found in previous research.

The outcomes of this investigation offer insights into the effectiveness of AI-generated data in augmenting the performance of machine learning models for kidney disease diagnosis. The ensuing discussion addresses several pertinent points and their implications. Performance Discrepancies a thorough comparison between original and AI-generated datasets unveiled discernible performance variations across different machine learning models and AI architectures. While certain AI-generated datasets highlighted enhancements over the original data, others exhibited varied degrees of performance degradation. Effectiveness of AI Models. GPT-4o consistently outperformed GPT-3.5 and GPT-4.0 in synthesizing data that closely mirrored the original dataset's performance. This underscores the significance of AI advancements in generating high-fidelity synthetic data. Model-Specific Performance Some machine learning models, such as Logistic Regression and Decision Tree, demonstrated resilience to dataset source variations, maintaining commendable performance across all AI-generated datasets.

Conversely, models like Support Vector Machine and Multi-Layer Perceptron displayed sensitivity to the quality of synthetic data, manifesting performance fluctuations across different AI models. Clinical Implications the incorporation of AI-generated data holds promise in augmenting medical diagnostic datasets, facilitating the development of more precise predictive models for kidney disease diagnosis. By leveraging AI models for data synthesis, healthcare practitioners can address challenges associated with data scarcity and enhance the generalization capabilities of machine learning models. Ethical Considerations Despite the potential benefits, ethical considerations concerning data privacy, bias mitigation, and model interpretability warrant attention. Ensuring transparency and accountability in data generation and utilization is imperative to address concerns pertaining to algorithmic bias and inadvertent consequences in clinical decision-making.

The analysis of the results indicates that the original models outperform those reported in many previous studies in terms of F1-score, and accuracy. The original models, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP), consistently demonstrate high performance. For instance, the original DT and RF models achieve nearly perfect scores, which are among the best seen in the literature. This suggests a high level of robustness and generalization in the original results. Compared to the previous studies, which show a wide range of performance metrics—with some achieving perfect scores and others reporting significantly lower outcomes—the original results exhibit greater consistency and reliability across different models. GPT-4’s results also show competitive performance, especially with perfect scores in DT and RF models, aligning well with the highest-performing previous studies. However, the original results maintain a slight edge in terms of overall generalization and consistency, indicating their robustness across diverse datasets and scenarios. This comprehensive performance underscores the effectiveness and reliability of the original models compared to those documented in the existing body of research.

1. **Conclusion**

This study reveals the transformative potential of AI-generated data in advancing medical diagnostics, particularly in the classification of chronic kidney disease (CKD). Through rigorous evaluation of AI models like GPT-4o for data synthesis, significant enhancements in diagnostic accuracy are demonstrated, addressing challenges posed by limited data availability in healthcare. However, ethical considerations regarding data privacy, bias mitigation, and interpretability are paramount for the ethical implementation of AI-generated data in clinical settings. Collaborative efforts among AI researchers, healthcare professionals, and policymakers are crucial for ensuring transparency, accountability, and the ethical use of AI technologies in healthcare. Moving forward, continued exploration and validation of AI-generated data across diverse medical contexts will be essential to fully realize its potential in improving medical diagnostics and patient outcomes on a global scale.

1. **Future Work**

Future research could explore newer AI generation tools and models, including advanced versions of ChatGPT and emerging AI technologies, to synthesize datasets for diverse medical conditions like CKD. This approach could help address ongoing challenges related to data scarcity and privacy concerns in healthcare. By applying similar methodologies and AI-generated data techniques, researchers may enhance diagnostic and predictive models across various diseases, potentially improving patient care and outcomes. Continued collaboration among AI researchers, healthcare professionals, and policymakers will be crucial for the ethical and effective implementation of these advancements in healthcare settings.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Nour Zawawi and Nermin Negied; methodology, Nour Zawawi; software, Nour Zawawi; validation, Nour Zawawi and Nermin Negied; formal analysis, Nermin Negied; investigation, Nermin Negied; data curation, Nour Zawawi; writing—original draft preparation, Nour Zawawi; writing—review and editing, Nermin Negied; visualization, Nour Zawawi; supervision, Nermin Negied; project administration, Nermin Negied.

References

1. Francis, A., Harhay, M.N., Ong, A.C.M. et al. Chronic kidney disease and the global public health agenda: an international consensus. Nat Rev Nephrol 20, 473–485 (2024). <https://doi.org/10.1038/s41581-024-00820-6>
2. [Negied, Nermin, Ahmed SeragEldin, “Automatic Detection of Alzheimer Disease from 3D MRI Images using Deep CNNs”, International Journal of Advanced Computer Science and Applications, Vol. 13, Issue 12, 2022. DOI:10.14569/IJACSA.2022.0131258.](https://www.proquest.com/openview/44421bad128919ebe4b89e41c125d94e/1?pq-origsite=gscholar&cbl=5444811)
3. Mamatha, B & Terdal, Sujatha. (2024). Early-Stage Chronic Kidney Disease Detection using Machine Learning with Bigdata. 10.1201/9781003529231-58.
4. M. Abdullah and N. Negied, "Detection and prediction of Future Mental disorder from Social Media Data using Machine Learning, Ensemble Learning, and Large Language Models.," in IEEE Access*,* doi: 10.1109/ACCESS.2024.3406469.
5. [Nermin K. Negied](https://www.worldscientific.com/doi/abs/10.1142/S0218001419570027), “Infrared Thermography-Based Breast Cancer Detection — Comprehensive Investigation”, [International Journal of Pattern Recognition and Artificial Intelligence](https://www.worldscientific.com/worldscinet/ijprai)[Vol. 33, No. 06, 1957002 (2019)](https://www.worldscientific.com/toc/ijprai/33/06), [doi.org/10.1142/S0218001419570027](https://doi.org/10.1142/S0218001419570027)
6. U. Ekanayake and D. Herath, “Chronic kidney disease prediction using machine learning methods,” Moratuwa Engineering Research Conference (MERCon), Jul, 2020. doi:10.1109/mercon50084.2020.9185249
7. K. M. Almustafa, “Prediction of chronic kidney disease using different classification algorithms,” *Informatics in Medicine Unlocked*, vol. 24, p. 100631, 2021. doi:10.1016/j.imu.2021.100631
8. P. Chittora *et al.*, “Prediction of chronic kidney disease - A machine learning perspective,” *IEEE Access*, vol. 9, pp. 17312–17334, 2021. doi:10.1109/access.2021.3053763
9. M. A. Abdel-Fattah, N. A. Othman, and N. Goher, “Predicting chronic kidney disease using hybrid machine learning based on Apache Spark,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–12, Feb. 2022. doi:10.1155/2022/9898831
10. Md. A. Islam, Md. Z. Majumder, and Md. A. Hussein, “Chronic kidney disease prediction based on machine learning algorithms,” *Journal of Pathology Informatics*, vol. 14, p. 100189, 2023. doi:10.1016/j.jpi.2023.100189
11. A. Maurya *et al.*, “Chronic kidney disease prediction and recommendation of suitable diet plan by using machine learning,” *2019 International Conference on Nascent Technologies in Engineering (ICNTE)*, Jan. 2019. doi:10.1109/icnte44896.2019.8946029
12. D. A. Debal and T. M. Sitote, “Chronic kidney disease prediction using machine learning techniques,” *Journal of Big Data*, vol. 9, no. 1, Nov. 2022. doi:10.1186/s40537-022-00657-5
13. Y. Amirgaliyev, S. Shamiluulu, and A. Serek, “Analysis of chronic kidney disease dataset by applying machine learning methods,” *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, Oct. 2018. doi:10.1109/icaict.2018.8747140
14. A. Charleonnan et al., “Predictive analytics for chronic kidney disease using machine learning techniques,” 2016 Management and Innovation Technology International Conference (MITicon), Oct. 2016. doi:10.1109/miticon.2016.8025242
15. G. Dharmarathne, M. Bogahawaththa, M. McAfee, U. Rathnayake, and D. P. P. Meddage, "On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence," Intelligent Systems with Applications, vol. 200397, 2024. doi: <https://doi.org/10.1016/j.iswa.2024.200397>
16. S. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Enhanced Optimized Classification Model of Chronic Kidney Disease," International Journal of Advanced Computer Science and Applications, vol. 14, no. 2, 2023. doi: 10.14569/IJACSA.2023.0140239
17. M. A. R. Rahat, M. T. Islam, D. M. Cao, M. Tayaba, B. P. Ghosh, E. H. Ayon, and M. S. Bhuiyan, "Comparing Machine Learning Techniques for Detecting Chronic Kidney Disease in Early Stage," Journal of Computer Science and Technology Studies, vol. 6, no. 1, pp. 20-32, 2024. doi: 10.32996/jcsts
18. C. G. Xanthis, D. Filos, K. Haris, and A. H. Aletras, “Simulator-generated training datasets as an alternative to using patient data for Machine Learning: An example in myocardial segmentation with MRI,” *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105817, Jan. 2021. doi:10.1016/j.cmpb.2020.105817
19. Y. Chen and P. Esmaeilzadeh, "Generative AI in medical practice: in-depth exploration of privacy and security challenges," Journal of Medical Internet Research, vol. 26, p. e53008, 2024. doi: 10.2196/53008
20. J. P. Singh, "The Impacts and Challenges of Generative Artificial Intelligence in Medical Education, Clinical Diagnostics, Administrative Efficiency, and Data Generation," International Journal of Applied Health Care Analytics, vol. 8, no. 5, pp. 37-46, 2023.
21. [chronic kidney disease (kaggle.com)](https://www.kaggle.com/datasets/abhia1999/chronic-kidney-disease)
22. [Kidney disease with ChatGPT versions datasets (kaggle.com)](https://www.kaggle.com/code/youssef22ashraf/kidney-disease-with-chatgpt-versions-datasets/edit)