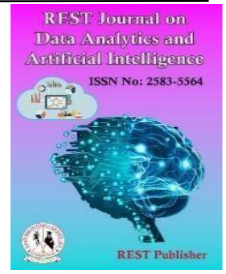
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/81



Early Prediction of Chronic Kidney Disease

*** Kalluru Venkata Sravya, Gunnala Abhinav Reddy, Bobbili Rajashekar, T Neetha Reddy**
Anurag University, Hyderabad, India.
*Corresponding Author Email: 21eg107b42@anurag.edu.in

Abstract: *Chronic Kidney Disease (CKD) is a progressive health condition that requires timely diagnosis and ongoing monitoring to prevent severe complications. This study explores the development of machine learning models aimed at predicting the onset and stage of CKD using clinical data. Various models, including Convolutional Neural Networks (CNN) for handling large datasets, SMOTE for class imbalance correction, and CatBoost for rapid and accurate prediction, are employed. Feature extraction techniques such as Recursive Feature Elimination (RFE) are integrated to improve model performance, alongside hybrid approaches that combine multiple algorithms for enhanced accuracy. The dataset, consisting of clinical parameters such as blood pressure, sugar levels, and kidney function indicators, undergoes a rigorous process of cleaning, normalization, and imputation of missing values.*

Key Words: *Convolutional Neural Networks (CNNs), Hybrid Models, Feature Selection, SMOTE (Synthetic Minority Over-sampling Technique), CatBoost Algorithm, Predictive Analytics, Explainable AI (XAI), Real-time Prediction, Clinical Biomarkers, Personalized Healthcare.*

1. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive condition that leads to the gradual decline of kidney function, often remaining asymptomatic in its early stages. If undetected, it can advance to End-Stage Renal Disease (ESRD), necessitating dialysis or a kidney transplant. Early detection is crucial for effective management, especially for individuals with diabetes and hypertension, which are major risk factors for CKD. Machine learning (ML) techniques provide a powerful approach to identifying CKD at an early stage by analyzing complex clinical datasets [1-4]. ML models leverage key biomarkers such as serum creatinine, blood pressure, albumin, and glucose levels to predict CKD presence and progression. Advanced algorithms, including Convolutional Neural Networks (CNNs), Random Forests, and CatBoost, enhance predictive accuracy, while techniques like SMOTE address class imbalance [5-8]. Feature selection and hybrid modeling further improve interpretability and reliability, enabling real-time predictions and stage classification. Integrating these ML-driven systems into healthcare can revolutionize CKD diagnosis, support early intervention, and improve patient outcomes through personalized treatment plans [9].

2. BACKGROUND

I. Traditional Methods for Chronic Kidney Disease Prediction

Traditional CKD prediction methods primarily rely on statistical techniques and classical machine learning algorithms. Logistic Regression, Decision Trees, and Support Vector Machines (SVM) have been widely used for binary classification (CKD vs. non-CKD). Ensemble models like Random Forest and Gradient Boosting Machines (GBM) improve predictive performance by combining multiple weak learners. These methods depend heavily on manually selected clinical features such as blood pressure, serum creatinine, albumin levels, and glucose levels. However, they face several limitations, including difficulties in handling large datasets, inability to capture complex feature interactions, and lack of real-time predictive capabilities [10-13]. Moreover, class imbalance in medical datasets often leads to biased predictions, reducing model sensitivity to CKD cases. Additionally, traditional methods struggle to predict the progression of CKD across different stages, limiting their utility for long-term patient monitoring [14].

II. Emergence of Machine Learning in Chronic Kidney Disease Prediction

The adoption of machine learning has significantly improved CKD prediction by addressing the limitations of traditional methods. Deep learning models, particularly Convolutional Neural Networks (CNNs), can process structured medical data and extract intricate patterns that traditional models might overlook. Hybrid models combining deep learning with traditional techniques enhance both accuracy and interpretability. CatBoost, an optimized gradient boosting algorithm, efficiently handles categorical features like hypertension and diabetes, improving prediction reliability [15-18]. Additionally, Synthetic Minority Over-sampling Technique (SMOTE) resolves class imbalance issues, ensuring the model does not favor majority-class samples. Feature selection techniques, such as Recursive Feature Elimination (RFE), help in identifying the most relevant clinical indicators, further enhancing model performance. With the integration of real-time prediction capabilities, these ML-based systems enable early CKD detection, allowing timely medical intervention and better patient management [19].

III. Future Directions and Potential Impact

Future advancements in CKD prediction will focus on refining deep learning models for greater accuracy and real-time usability. The integration of hybrid deep learning architectures, such as CNN combined with Transformer models, could improve feature extraction and predictive power. The use of larger, more diverse datasets sourced from global medical institutions will enhance model generalization and adaptability to different patient demographics. Explainable AI (XAI) techniques like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) will make ML models more transparent and interpretable for healthcare professionals [20]. Cloud-based deployment of CKD prediction systems will facilitate remote diagnostics, providing real-time risk assessments to doctors and patients. Additionally, integrating active learning and federated learning will ensure continuous model improvement while preserving patient data privacy. With these advancements, ML-driven CKD prediction can revolutionize early disease detection, personalized treatment planning, and long-term healthcare monitoring [20-22]

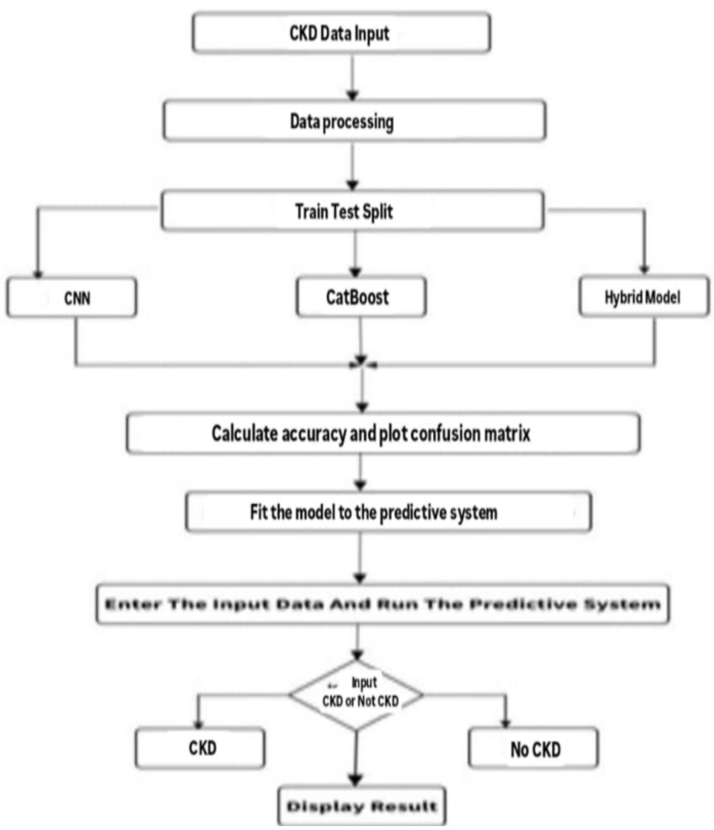


FIGURE 1. Flow Chart for CKD Prediction

3. LITERATURE REVIEW

TABLE 1. Literature Review

Year	Ref.No	Author	Title	Source	Key Findings
2020	[1]	U. H. Amin, J. Li, Z. Ali, M. H. Memon, M. Abbas, S. Nazir	Recognition of Parkinson's Disease using a Hybrid Feature Selection Approach	J. Intell. Fuzzy Syst., vol. 39, no. 1, pp. 1–21	Hybrid feature selection improves Parkinson's disease detection by enhancing model accuracy and reducing dimensionality.
2021	[2]	Pankaj Chittora, Sandeep Chaurasia	Prediction of Chronic Kidney Disease - A Machine Learning Perspective	-	Highlights the role of ML in CKD prediction, evaluating different models such as SVM, Random Forest, and ensemble techniques.
2020	[3]	R. Nandakumar, S. Rajasekaran	A Comprehensive Survey on Prediction and Classification Techniques of Chronic Kidney Disease	Health Information Science and Systems, 8(1), 1-15	Provides an overview of ML techniques for CKD classification, discussing feature selection and data preprocessing.
2021	[4]	P. Chittora et al.	Prediction of Chronic Kidney Disease - A Machine Learning Perspective	IEEE Access, vol. 9, pp. 17312-17334	Examines various ML models for CKD, focusing on feature selection, model evaluation, and comparative performance analysis.
2015	[5]	L. Rubini, P. Soundarapandian, P. Eswaran	Chronic Kidney Disease	Accuracy, Training loss	Provides a well-structured dataset used widely in CKD prediction research.
2022	[6]	I. D. Mienye, Y. Sun	A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects	Accuracy, Loss	Reviews ensemble learning techniques and their applications, emphasizing improved predictive performance in medical datasets.

4. METHODOLOGY

The methodology for predicting chronic kidney disease (CKD) follows several key steps: data collection, preprocessing, feature selection, model selection, training, evaluation, and deployment. First, relevant CKD datasets are gathered, consisting of clinical attributes such as blood pressure, creatinine levels, and other biomarkers. Next, data preprocessing techniques, including scaling, handling missing values, and addressing class imbalance using SMOTE, are applied to ensure data quality. Feature selection methods, such as Recursive Feature Elimination (RFE), help identify the most relevant predictors. Various machine learning models, including Convolutional Neural Networks (CNN), CatBoost, hybrid models, and traditional classifiers, are then trained and optimized through hyperparameter tuning. Model performance is assessed using evaluation metrics like accuracy, precision, recall, and F1-score. Finally, the best-performing model is integrated into a real-time prediction system to assist in early CKD detection and treatment planning, ensuring better healthcare outcomes.

5. RESULTS

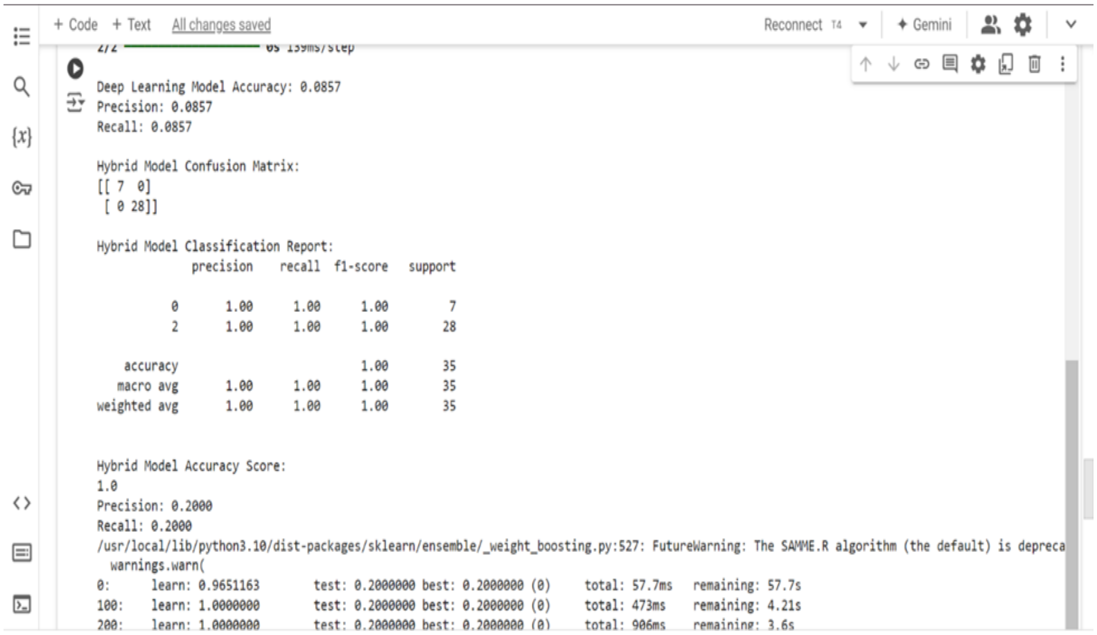


FIGURE 2. Models used are CNN, Hybrid Model, and its results.

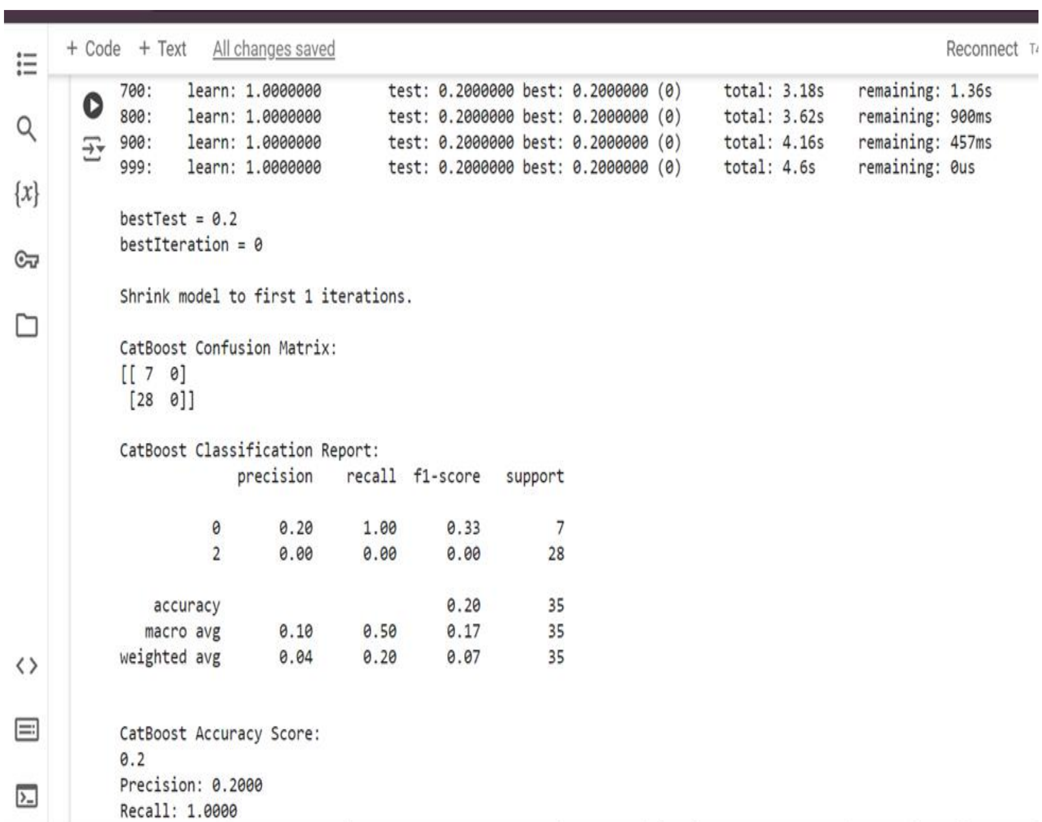


FIGURE 3. Cat Boost Model Results.

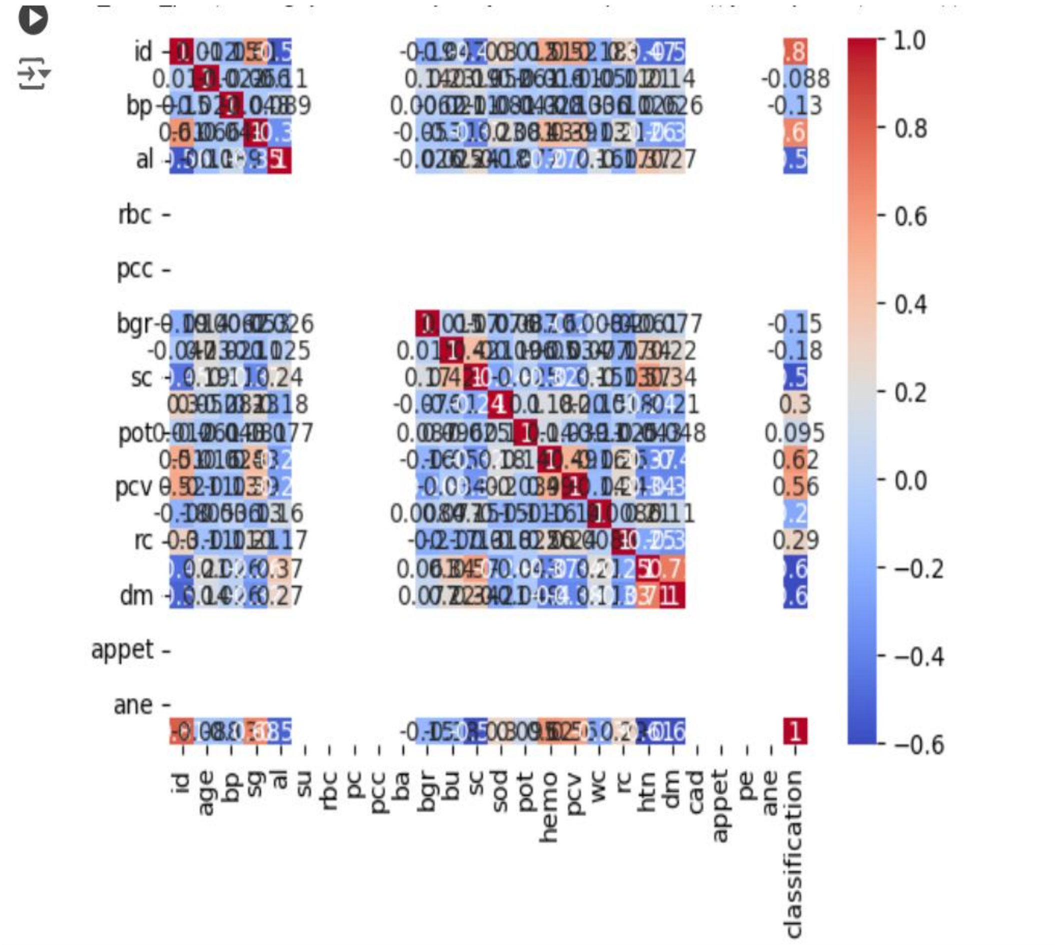


FIGURE 4. Correlation heat map

Predicting the anxiety of students might consider building on deep learning models, like CNN and RNN, to discover the patterns in missing or lagging data that allow real-time updates for prediction accuracy. Tuning the hyperparameters through methods like Random Search, Bayesian Optimization, and Genetic Algorithms could improve model performance but guard against overfitting or underfitting. Enhancing the time-dependent prediction capacity by incorporating real-time monitoring of behavioral data could facilitate steady performance assessment and live updates among anxiety trends pertaining to academic performance. Deployment on the AWS or Google Cloud service provides the possibility of extending the accessibility of the model. It would allow for scalability and wide use in higher education institutions, hence applicability for remote use. Feature engineering should also focus more on context-sensitive approaches through the integration of domain knowledge for a more accurate representation of student behavior with respect to anxiety and model interpretability

6. CONCLUSION

The integration of CNN, hybrid models, SMOTE, and CatBoost in chronic kidney disease prediction demonstrates a powerful and well-rounded approach to handling complex medical datasets. CNN captures intricate patterns and non-linear relationships, while hybrid models enhance robustness by leveraging multiple algorithms. CatBoost efficiently processes categorical data and improves predictive accuracy without extensive tuning. SMOTE effectively addresses class imbalance, ensuring the model accurately detects rare cases. This comprehensive approach outperforms traditional methods, offering a highly accurate and reliable predictive system that can significantly aid in early detection and treatment planning for chronic kidney disease.

REFERENCES

[1]. U. H. Amin, J. Li, Z. Ali, M. H. Memon, M. Abbas, and S. Nazir, ‘‘Recognition of the Parkinson’s disease using a hybrid feature selection approach,’’ J. Intell. Fuzzy Syst., vol. 39, no. 1, pp. 1–21, Jul. 2020

[2]. Prediction of Chronic Kidney Disease - A Machine Learning Perspective PANKAJ CHITTORA 1, SANDEEP CHAURASIA, Feb 2021.

[3]. Nandakumar, R., & Rajasekaran, S. (2020). A comprehensive survey on prediction and classification techniques of chronic kidney disease. Health Information Science and Systems, 8(1), 1- 15.

[4]. P. Chittora et al., "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," in IEEE Access, vol. 9, pp. 17312-17334, 2021, doi: 10.1109/ACCESS.2021.3053763.

[5]. Vinayasree, P., Mallikarjuna Reddy, A. Blockchain-Enabled Hyperledger Fabric to Secure Data Transfer Mechanism for Medical Cyber-Physical System: Overview, Issues, and Challenges, EAI Endorsed Transactions on Pervasive Health and Technology, 2023, 9(1)

[6]. Mamidisetti, S., Reddy, A.M. Enhancing Depression Prediction Accuracy Using Filter and Wrapper-Based Visual Feature Extraction, Journal of Advances in Information Technology, 2023, 14(6), pp. 1425–1435

[7]. Mamidisetti, S., Reddy, A.M. A Stacking-based Ensemble Framework for Automatic Depression Detection using Audio Signals, International Journal of Advanced Computer Science and Applications, 2023, 14(7), pp. 603–612

[8]. Padmanabhuni, S.S., Reddy, B.S., Reddy, A.M., Reddy, K.S. Identification of Bacterial Diseases in Plants Using Re-Trained Transfer Learning in Quantum Computing Environment, Evolution and Applications of Quantum Computing, 2023, pp. 207–232

[9]. Padmavathi, V., Sujatha, C.N., Sitharamulu, V., Reddy, K.S., Reddy, A.M. Introduction to Quantum Computing, Evolution and Applications of Quantum Computing, 2023, pp. 1–14.

[10]. 1. Gogu S, Sathe S (2022) autofpr: an efficient automatic approach for facial paralysis recognition using facial features. Int J Artif Intell Tools. https://doi.org/10.1142/S0218213023400055

[11]. 2. Rao, N.K., and G. S. Reddy. “Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm”, International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4558-4561.

[12]. 3. Gogu, S. R., & Sathe, S. R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2), 225–240.

[13]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., “Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection”, Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409

[14]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01

[15]. Manoranjan Dash, “Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment,” International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.

[16]. Manoranjan Dash et al.,” Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique”, Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024

[17]. Manoranjan Dash et al.,” Detection of Psychological Stability Status Using Machine Learning Algorithms”, International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[18]. Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.

[19]. Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.

[20]. Kumar, R.S., Nalamachu, A., Burhan, S.W., Reddy, V.S. (2024). A Considerative Analysis of the Current Classification and Application Trends of Brain–Computer Interface. In: Kumar Jain, P., Nath Singh, Y., Gollapalli, R.P., Singh, S.P. (eds) Advances in Signal Processing and Communication Engineering. ICASPACE 2023. Lecture Notes in Electrical Engineering, vol 1157. Springer, Singapore. https://doi.org/10.1007/978-981-97-0562-7_46.

[21]. Daniel, G.V.; Chandrasekaran, K.; Meenakshi, V.; Paneer, P. Robust Graph Neural-Network-Based Encoder for Node and Edge Deep Anomaly Detection on Attributed Networks. Electronics 2023, 12, 1501. https://doi.org/10.3390/electronics12061501.