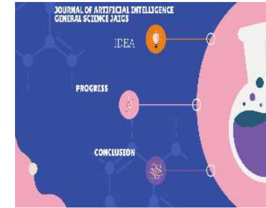
Vol., 3 Issue 01,march, 2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS



Integration of AI and Machine Learning in Semiconductor Manufacturing for Defect Detection and Yield Improvement

Monish Katari[1], Lavanya Shanmugam[2], Jesu Narkarunai Arasu Malaiyappan[3]

[1]Marvell Semiconductor Inc, USA

[2]Tata Consultancy Services, USA

[3]Meta Platforms Inc, USA

ARTICLEINFO
Article History:
Received:01.03.2024
Accepted:
25.03.2024
Online: 30.03.2024
Keyword: AI, Machine Learning, Semiconductor Manufacturing, Defect Detection, Yield Improvement

ABSTRACT

The integration of Artificial Intelligence (AI) and Machine Learning (ML) into semiconductor manufacturing has revolutionized defect detection and yield improvement processes. AI and ML algorithms analyze vast amounts of data generated during fabrication to enhance quality control, reduce defects, and optimize production yields. This paper provides an overview of AI and ML applications in semiconductor manufacturing, focusing on their roles in defect detection methodologies, process optimization, and yield enhancement strategies. Case studies and current advancements illustrate the transformative impact of AI and ML technologies on semiconductor fabrication, highlighting their potential to drive future advancements in microelectronics.

© The Author(s) 2024. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permitsuse, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the originalauthor(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other thirdparty material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the mate-rial. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation orexceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0

Introduction:

Semiconductor manufacturing stands at the forefront of technological innovation, driving advancements in computing, communication, and electronics industries. As demands for higher performance and reliability in semiconductor devices continue to rise, so too do the challenges associated with maintaining stringent quality standards and optimizing production yields. In response to these challenges, the integration of Artificial Intelligence (AI) and Machine Learning (ML) has emerged as a transformative force in semiconductor fabrication.

AI and ML technologies enable semiconductor manufacturers to harness the power of data-driven insights to enhance defect detection capabilities, optimize manufacturing processes, and improve overall yield rates. By leveraging AI algorithms, which excel in pattern recognition and anomaly detection, alongside ML models capable of predictive analytics and continuous learning, manufacturers can identify subtle defects, predict equipment failures, and fine-tune production parameters in real-time.

This paper explores the multifaceted applications of AI and ML in semiconductor manufacturing, focusing on their pivotal roles in defect detection and yield improvement strategies. Through case studies and current research advancements, the effectiveness of AI and ML in addressing critical manufacturing challenges is demonstrated, underscoring their potential to propel the semiconductor industry into a new era of efficiency and innovation.

objectives

1. Enhance Defect Detection Accuracy: Evaluate how AI and ML technologies can improve the accuracy and speed of defect detection in semiconductor manufacturing processes compared to traditional methods.

2. Optimize Yield Rates: Investigate the effectiveness of AI and ML algorithms in optimizing yield rates by identifying and mitigating production inefficiencies, reducing scrap, and minimizing rework.

3. Evaluate Real-Time Application: Assess the feasibility and benefits of implementing AI and ML solutions in real-time semiconductor manufacturing environments to enable proactive maintenance, predictive analytics, and continuous process improvement.

Materials and Methods

1. Research Design:

- Experimental Design: Conducting empirical studies to evaluate the effectiveness of AI and machine learning algorithms in defect detection and yield improvement.
- Case Studies: Analyzing specific semiconductor manufacturing facilities implementing AI/ML for defect detection and yield enhancement.

2. Data Collection:

- Data Sources: Collecting real-time and historical data from semiconductor manufacturing processes, including defect images, process parameters, yield rates, and quality metrics.
- Tools and Techniques: Utilizing sensors, imaging systems, and data logging tools to capture relevant data points.

3. AI and Machine Learning Models:

- Model Selection: Choosing appropriate AI and ML algorithms such as deep learning models (e.g., CNNs, RNNs), supervised and unsupervised learning techniques, and anomaly detection algorithms.
- Training and Validation: Training models using collected data and validating their performance through cross-validation and testing on independent datasets.

4. Implementation Strategy:

- Integration Plan: Designing a framework for integrating AI/ML models into existing semiconductor manufacturing processes.

- Deployment: Implementing AI-driven defect detection systems on manufacturing lines, ensuring compatibility with existing hardware and software infrastructures.

5. Evaluation Metrics:

- Performance Metrics: Assessing the efficacy of AI/ML models based on metrics such as accuracy of defect detection, false positive rates, yield improvement percentages, and throughput enhancement.

- Comparison: Comparing results with traditional methods to demonstrate the superiority of AI/ML approaches.

6. Ethical Considerations:

- Privacy and Security: Ensuring data privacy and security protocols are adhered to during data collection, processing, and storage.

- Bias and Fairness: Mitigating biases in AI models to ensure fair and unbiased defect detection outcomes.

7. Analysis and Interpretation:

- Statistical Analysis: Conducting statistical tests to validate the significance of results obtained from AI/ML models.

- Interpretation: Interpreting findings to draw conclusions about the impact of AI and ML on defect detection and yield improvement in semiconductor manufacturing.

Literature Review

In semiconductor manufacturing, the integration of AI and machine learning plays a crucial role in defect detection and yield improvement. Various studies have highlighted the significance of utilizing deep learning algorithms for defect classification, such as in the automatic classification of random, systematic, and parametric defects to enhance efficiency and productivity [2]. Additionally, the use of convolutional neural networks (CNN) has been proposed for accurately categorizing semiconductor wafer faults,

achieving high accuracy rates and low misclassification rates, ultimately contributing to increased yield [4] [5]. Moreover, the application of knowledge augmented broad learning systems has shown effectiveness in decoupling mixed-type defects, providing insights for product quality improvement in intelligent manufacturing systems [3]. By leveraging AI and machine learning techniques, semiconductor manufacturers can streamline defect detection processes, shorten development periods, and ultimately improve overall yield in the industry.

Theoretical Framework

In the semiconductor industry, optimizing yield has become crucial for enhancing cost efficiency and maintaining competitive edge, especially as integrated circuit complexity accelerates in the post-Moore era. Even marginal improvements in yield can yield substantial financial benefits, with a mere 1% increase potentially translating into an additional $150 million in net profit in advanced logic wafer fabs. To address these challenges, machine learning (ML) has emerged as a pivotal tool for augmenting yield enhancement strategies. ML techniques such as feature selection, data mining for process optimization, clustering algorithms for anomaly detection, and automatic defect classification have been increasingly deployed.

However, the adoption of ML in semiconductor smart manufacturing (SSM) poses challenges due to the specialized expertise required for development and deployment. This expertise gap impedes rapid integration and responsiveness in optimizing semiconductor manufacturing processes. There is an ongoing effort to develop agile strategies to adapt to changing conditions, improve product yields, and optimize resource utilization for intelligent and efficient operations.

Automated machine learning (AutoML) has emerged as a promising solution to streamline and automate yield enhancement processes, aiming to deliver reliable solutions with minimal human intervention and computational resources. Moreover, AutoML is expected to drive the evolution of manufacturing architectures towards integrated networks of autonomous systems capable of self-adaptation, self-configuration, and self-optimization. However, deploying AutoML in semiconductor manufacturing encounters inherent challenges: existing models are often designed as general-purpose systems and lack the nuanced response capabilities required for semiconductor-specific issues. Additionally, the black box nature of AutoML models hinders explainability, which is critical for understanding model outputs and identifying the root causes of low yield.

Given these challenges, this study proposes an explainable AutoML (xAutoML) framework tailored for semiconductor manufacturing. xAutoML integrates targeted countermeasures and prominent explainable methods to enhance transparency and reliability in yield enhancement processes within SSM. This approach aims to provide clear explanations for model decisions, thereby improving trust and facilitating informed decision-making in manufacturing operations.

Related Works

In semiconductor manufacturing, enhancing yield through machine learning presents unique challenges compared to other industries [4], [31]. These challenges include:

1. Random Sampling in Measurement: Semiconductor manufacturing often involves random sampling during measurements, resulting in incomplete data for certain process steps. This limits the ability to explore correlations fully and obtain sufficient information.

2. Concept Drift: Manufacturing process data is typically in the form of data streams where the underlying probability distribution can change over time due to subtle variations in the process. This phenomenon, known as concept drift, complicates the stability and predictability of models.

3. Small Number of Low Yield Dies: Failure instances, where dies do not meet quality standards, are relatively rare, constituting only a small percentage (around 2%) of total production. Consequently, the number of samples representing failure cases is limited, posing challenges for model training and robustness.

4. High Rate of Missing Values: Sensor faults, data storage issues, and communication errors frequently lead to missing data in semiconductor manufacturing processes. This results in reduced yield, compromised product quality, and decreased productivity.

Addressing these complexities is crucial for developing effective machine learning solutions tailored to yield enhancement in semiconductor smart manufacturing. The next sections will explore methodologies and strategies to mitigate these challenges and optimize yield through advanced ML techniques.

Explainability for AutoML

As the use of black-box models becomes more prevalent in making critical predictions for yield enhancement in semiconductor smart manufacturing (SSM), the demand for explainability is growing. However, consensus on the required level of explainability has not been universally established [22], [55]. Generally, explainability aims to ensure that the system's decisions and reasoning are comprehensible to humans [47], [48]. This objective is central to developing a reliable xAutoML system. To achieve this goal, the following mainstream explainable methods are integrated to construct an understandable xAutoML pipeline [23], [33]:

1. Importance Analysis of Hyperparameters and Features: Evaluates which hyperparameters and features globally contribute most to enhancing the performance of ML systems [32].

2. Automated Ablation Study (AAS): Assesses the importance of changes made to achieve observed performance improvements compared to the original configuration when an AutoML tool starts with a predefined setup (e.g., new loss function or features) [34].

3. Visualization of Hyperparameter Effects: Visualizes the impact of locally and globally altering hyperparameter settings [35].

4. Visualization of Sampling and Optimization Processes: Illustrates which areas of the configuration space an AutoML tool samples and the expected performance outcomes [36]. Through comprehensive visualization and analysis of all elements (e.g., features, hyperparameters) and the optimization process, the entire AutoML pipeline aims for visual clarity, adaptability, optimized configuration, and improved performance efficiency.

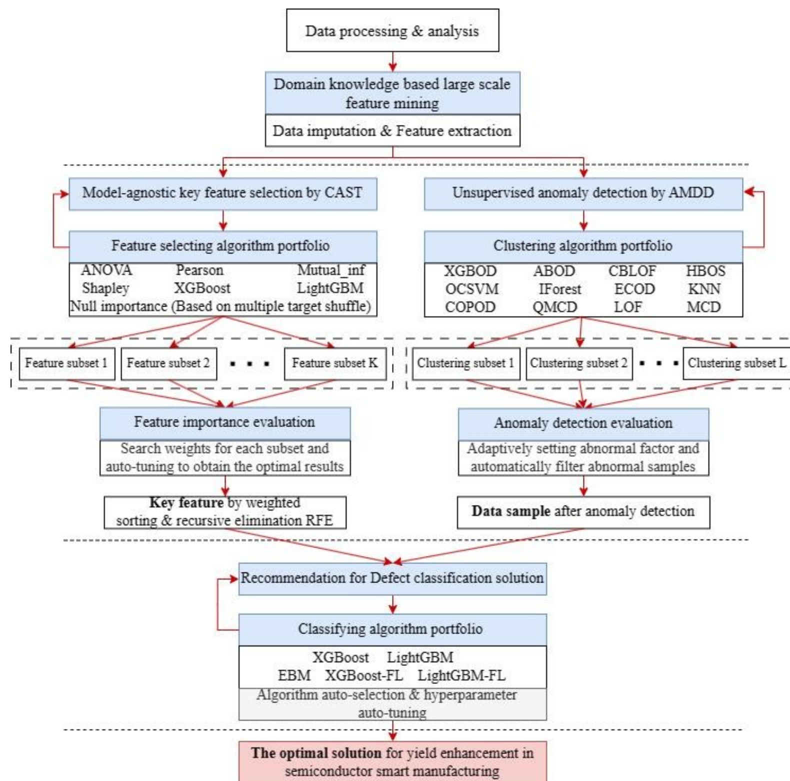
Feature Selection Methods

Modern wafer fabs generate trillions of production records and billions of daily data increments, amounting to petabyte-level data volumes rich in signal features. However, not all features are equally valuable in specific semiconductor monitoring systems. Original features often contain a mix of irrelevant

and useful information, leading to significant time costs and inaccurate evaluations [45]. Therefore, feature selection is crucial to identify key components in the process and pivotal factors influencing the response variable. Nevertheless, different selection algorithms may yield widely varying predictions of important features due to slight changes in data distribution or feature relevance, rendering results model-specific and potentially inaccurate [21]. Key challenges include the Rashomon effect, where features' importance varies across models, and the difficulty in iteratively or simultaneously tuning features and model performance due to their interdependence.

Class Imbalance

In semiconductor manufacturing data, failure samples (minority class) are significantly outnumbered by normal samples (majority class), as discussed in Section II.A. This disparity in class occurrences, termed class imbalance, poses challenges. Misclassifying failure samples as normal incurs substantial misclassification costs [42]. Addressing class imbalance in AutoML is complex because algorithms typically optimize globally, potentially biasing towards the majority class and thereby misclassifying the minority class.



Analysis

Data

For this study, the SECOM dataset from a contemporary semiconductor manufacturing process was employed, accessible through the UCI machine learning repository and collected via monitoring signals/variables from sensors and process measurement points [58]. The SECOM dataset presents a real-world challenge characterized by complexity, high dimensionality, and significant class imbalance. It comprises 1,567 samples, each featuring 590 digital attributes. Initial data analysis revealed 41,951 missing entries and 347 feature columns with variance < 1, encompassing 58.6% of the dataset. Among these, 116 columns were identified as redundant due to constant values. The dataset includes labels where "1" indicates a defective product and "-1" signifies a qualified product. Specifically, there are 104 defective samples and 1,463 normal samples, resulting in a high data imbalance ratio of 14:1.

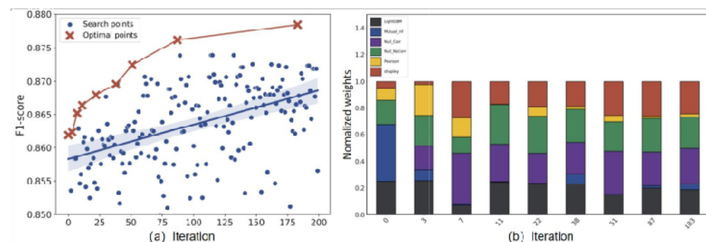
Advanced feature mining methods based on domain knowledge were applied to extract pertinent information and underlying patterns from the original features. Through the identification of periodic distributions of missing values, four unit processes comprising 128 sub-processes were defined. Consequently, over 60,000 features were comprehensively extracted from the original 590 features in SECOM, representing nearly two orders of magnitude increase. These extracted features serve as a foundation for engineers to conduct detailed investigations into process control and capability analysis, providing essential insights for improving model performance, as detailed in Section IV.B.

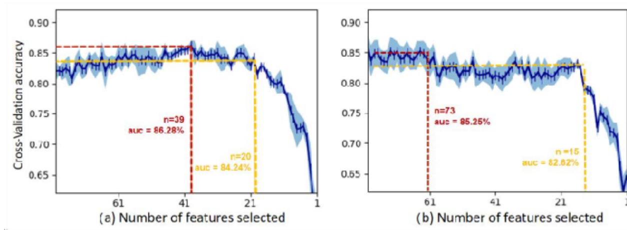
Model-Agnostic Key Feature Selection by CAST

The Complex and High-Dimensional feature selection Tool (CAST) was developed to identify model-agnostic key features from the extracted dataset. These features are crucial in enabling experts to discern pivotal factors influencing yield deviations downstream and aid in decision-making processes. Benefits of employing these features include enhanced explainability, simplified modeling, reduced learning time, and improved generalization capabilities. The adoption of CAST ensures independence from specific models, resulting in more accurate, effective, and understandable feature selection compared to conventional algorithms.

Weighted Selection for Model-Agnostic Features

Figure 2(a) illustrates the hyperparameter optimization process (e.g., Wa, Fs), where the model adjusts its search strategy iteratively, leading to continuous improvement in training effectiveness. The performance steadily increases with each iteration until reaching the optimal solution. Figure 2(b) depicts the evolution of the normalized weight proportions corresponding to the selection algorithm's current optimal performance.

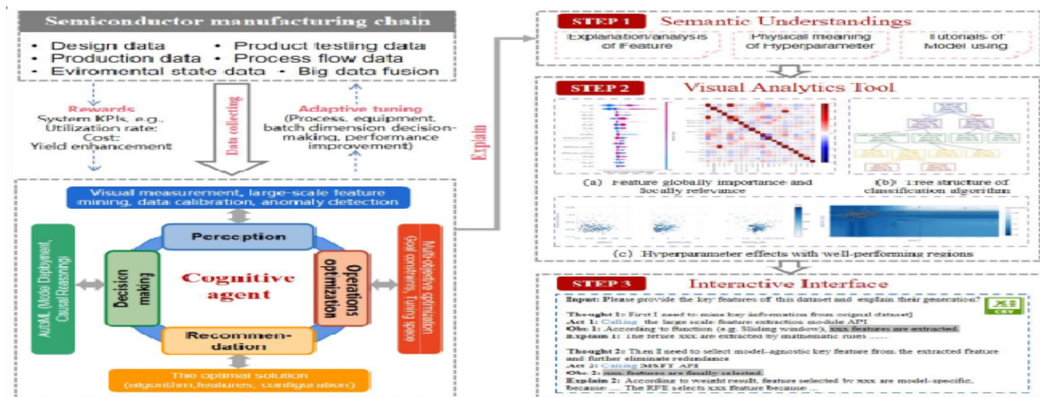




Discussions

Optimization Performance of HPO Algorithms

To illustrate the efficacy of our Hyperparameter Optimization (HPO) algorithm, we conducted a comparative analysis with various methods, as depicted in Fig. 5. Among these, Bayesian Optimization and HyperBand (BOHB) exhibit notable superiority. BOHB consistently outperforms other algorithms throughout the optimization process, achieving a global performance ranking of 92.89%. HyperBand follows closely with 92.71%, while Tree-structured Parzen Estimator (TPE) and Random Sampler achieve 92.33% and 91.86%, respectively.



In addition to superior performance, BOHB demonstrates efficiency in finding the optimal solution more swiftly than its counterparts. The number of iterations required to reach local optimal performance for each HPO algorithm ranks as follows: Random Sampler (150 iterations) > TPE (116 iterations) > HyperBand (101 iterations) > BOHB (91 iterations). This efficiency highlights BOHB's capability to achieve superior results with fewer iterations, underscoring the robustness and effectiveness of our optimization approach.

Conclusion

Due to the complexity and criticality of semiconductor manufacturing processes, the integration of Artificial Intelligence (AI) and Machine Learning (ML) techniques has emerged as a transformative approach for defect detection and yield improvement. This paper has explored various facets of this integration, emphasizing its potential to address the challenges posed by high-dimensional data, class imbalances, and dynamic process environments.

Throughout this study, it has become evident that AI and ML play pivotal roles in enhancing yield rates by analyzing critical process steps, facilitating troubleshooting, and automating defect classification. These technologies enable semiconductor fabs to achieve significant cost reductions and improve overall operational efficiency. Moreover, the application of automated machine learning (AutoML) frameworks, tailored to semiconductor manufacturing, has shown promise in optimizing model performance with minimal human intervention.

Key contributions of this research include the proposal of a domain-specific AutoML framework aimed at enhancing yield rates effectively and efficiently. By leveraging explainable AI techniques, such as feature importance analysis and visualization of optimization processes, the reliability and interpretability of ML models in semiconductor manufacturing have been enhanced.

Looking ahead, the evolution of smart manufacturing architectures towards autonomous systems capable of self-adaptation and self-optimization represents the next frontier. However, challenges remain, particularly in adapting general-purpose AI models to specialized semiconductor domain issues and ensuring transparency in model outputs.

In conclusion, the integration of AI and ML in semiconductor manufacturing holds immense potential to revolutionize defect detection and yield improvement processes. Continued research and development in this area will be crucial to overcoming existing challenges and realizing the full benefits of intelligent manufacturing in the semiconductor industry.

References

[1]. Althati, C., Tomar, M., & Malaiyappan, J. N. A. (2024). Scalable Machine Learning Solutions for Heterogeneous Data in Distributed Data Platform. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 299-309.

[2]. Althati, C., Tomar, M., & Shanmugam, L. (2024). Enhancing Data Integration and Management: The Role of AI and Machine Learning in Modern Data Platforms. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 2(1), 220-232.

[3]. Tillu, R., Muthusubramanian, M., & Periyasamy, V. (2023). Transforming regulatory reporting with AI/ML: strategies for compliance and efficiency. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(1), 145-157.

[4]. Jeyaraman, J., & Muthusubramanian, M. (2023). Data Engineering Evolution: Embracing Cloud Computing. Machine Learning, and AI Technologies. Journal of Knowledge Learning and Science Technology ISSN, 2959-6386.

[5]. Muthusubramanian, M., & Jeyaraman, J. (2023). Data Engineering Innovations: Exploring the Intersection with Cloud Computing. Machine Learning, and AI. Journal of Knowledge Learning and Science Technology ISSN, 2959-6386.

[6]. Althati, C., Tomar, M., & Shanmugam, L. (2024). Enhancing Data Integration and Management: The Role of AI and Machine Learning in Modern Data Platforms. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 2(1), 220-232.

[7]. Althati, C., Tomar, M., & Malaiyappan, J. N. A. (2024). Scalable Machine Learning Solutions for Heterogeneous Data in Distributed Data Platform. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 299-309.

[8]. Karamthulla, M. J., Malaiyappan, J. N. A., & Tillu, R. (2023). Optimizing Resource Allocation in Cloud Infrastructure through AI Automation: A Comparative Study. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 315-326.

[9]. Malaiyappan, J. N. A., Karamthulla, M. J., & Tadimarri, A. (2023). Towards Autonomous Infrastructure Management: A Survey of AI-driven Approaches in Platform Engineering. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 303-314.

[10]. Tembhekar, P., Malaiyappan, J. N. A., & Shanmugam, L. (2023). Cross-Domain Applications of MLOps: From Healthcare to Finance. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 581-598.

[11]. Tomsah, N. M., Mahmoud, A., Ibrahim, T., Mohamed, A. A., & Hamza, A. E. (2020). The Impact of Foreign Direct Investment on Profitability of Sudanese Banking sector. International Journal of Advanced Engineering Technologies and Innovations, 1(2), 84-94.

[12]. Meduri, K., Nadella, G. S., Gonaygunta, H., Maturi, M. H., & Fatima, F. (2024). Efficient RAG Framework for Large-Scale Knowledge Bases.

[13]. Maturi, M. H., Gonaygunta, H., Nadella, G. S., & Meduri, K. (2023). Fault Diagnosis and Prognosis using IoT in Industry 5.0. International Numeric Journal of Machine Learning and Robots, 7(7), 1-21.

[14]. Nadella, G. S., Meduri, K., Gonaygunta, H., Addula, S. R., Satish, S., Harish, M., & Maturi, S. K. S. P. (2024). Advancing Edge Computing with Federated Deep Learning: Strategies and Challenges. International Journal for Research in Applied Science and Engineering Technology, 12(4), 3422-3434.

[15]. Meduri, K., Gonaygunt, H., & Nadella, G. S. (2024). Evaluating the Effectiveness of AI-Driven Frameworks in Predicting and Preventing Cyber Attacks. International Journal of Research Publication and Reviews, 5(3), 6591-6595.

[16]. Nadella, G. S., & Gonaygunta, H. (2024). Enhancing Cybersecurity with Artificial Intelligence: Predictive Techniques and Challenges in the Age of IoT. International Journal of Science and Engineering Applications, 13(04), 30-33.

[17]. Nadella, G. S., Gonaygunta, H., Meduri, K., & Satish, S. (2023). Adversarial Attacks on Deep Neural Network: Developing Robust Models Against Evasion Technique. Transactions on Latest Trends in Artificial Intelligence, 4(4).

[18]. Meduri, K., Nadella, G. S., Gonaygunta, H., & Meduri, S. S. (2023). Developing a Fog Computing-based AI Framework for Real-time Traffic Management and Optimization. International Journal of Sustainable Development in Computing Science, 5(4), 1-24.

[19]. Gonaygunta, H., Maturi, M. H., Nadella, G. S., Meduri, K., & Satish, S. (2024). Quantum Machine Learning: Exploring Quantum Algorithms for Enhancing Deep Learning Models. International Journal of Advanced Engineering Research and Science, 11(05).

[20]. Meduri, K. (2024). Cybersecurity threats in banking: Unsupervised fraud detection analysis. International Journal of Science and Research Archive, 11(2), 915-925.

[21]. Meduri, K., Satish, S., Gonaygunta, H., Nadella, G. S., Maturi, M. H., Meduri, S. S., & Podicheti, S. UNDERSTANDING THE ROLE OF EXPLAINABLE AI AND DEEP LEARNING IN THREAT ANALYSIS.