Kurdish Studies
December 2022
Volume: 10, No: 2, pp 917-933
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v10i2.3854

AI-Powered Fault Detection In Semiconductor Fabrication: A Data-Centric Perspective

Botlagunta Preethish Nandan*

*SAP Delivery Analytics, preethishnananbotlagunta@gmail.com, ORCID ID : 0009-0008-3617-8149

Abstract
This paper presents an AI algorithm to detect faults in semiconductor fabrication. The algorithm works in multiple stages. In the first stage, a signal classification approach is proposed. The segmentation of long time series signals is performed, and a one-class classification algorithm is proposed. The class is then created using the normal samples to identify the anomalies later in the diffusion process. In the second stage, a neural network-based tool is proposed to explain and filter the results from the first model.
Semiconductor fabrication is a highly complex and automated technological process that consists of hundreds of steps. As equipment becomes more advanced, the quality of the equipment and control of the process improve, making the detection of faults more difficult and therefore more significant. In semiconductor manufacturing, chip-level faults should be detected using equipment-level signals throughout the fabrication process.
Semiconductor fabrication is a combination of a series of unit processes with different themes that create 2D or 3D patterns on a silicon wafer. Each time a signal is collected, it is a time series consisting of hundreds of thousands of high frequency data points. In addition, these signals are affected by other variables such as the temperature of the manufacturing environment, recipe parameters, and noise. To understand such complex and high-frequency data for a specific unit process, the first approach is to characterize each signal using time intervals as a result of statistical information such as mean, median, maximum, minimum, kurtosis, and more. A machine learning approach is then used to detect anomalies, which impacts the normal operation of equipment and productivity [1].

Keywords : Data analytics, yield prediction, semiconductor manufacturing, machine learning, predictive modeling, process optimization, defect analysis, big data, statistical process control, anomaly detection, real-time monitoring, artificial intelligence, wafer-level data, equipment data, root cause analysis, pattern recognition, data mining, manufacturing intelligence, sensor data, quality control, production efficiency, regression analysis, classification models, predictive maintenance, deep learning, feature extraction, high-dimensional data, yield enhancement, data-driven decision-making, advanced analytics.

Introduction

Semiconductor chips with high device integration density and evolving sophisticated device structures increase the complexity of detecting defects. To assess the production quality and reliability of these chips, various characterization methods, including microwave power loss measurement, optical emission spectroscopy analysis, and electrical parametric testing, are utilized. These methods offer effective structural and functional defect detection, but they are inefficient in analyzing chips on a large scale. Therefore, unsupervised fault detection is proposed as an additional algorithm to significantly reduce the complexity of defect diagnostics [2].
Deep generative models, such as variational autoencoders and generative adversarial networks, have been the majority in defect visualization and classification in semiconductor failure analysis. However, the reconstruction-based defect detection algorithm fails in severe defect scenarios and delicate structure chips due to the lack of prior knowledge. Thus, a new generative defect detection algorithm by combining the statistical region proposal network and improvement of feature fusion method is proposed to complement the existing high-performance fault detection and classification algorithm. Simulation results on a located circuit-transistor failure illustration illustrate the increased robustness of the proposed unsupervised defect detection algorithm for enhancing defect detection performance [1].
Owing to increasingly challenging manufacturing environments and the sophistication of emerging technologies/products, detection of the potential faults occurring in manufacturing processes has become increasingly difficult in semiconductor fabrication. During the fabrication process, high-performance and proper fault detection and diagnostic systems should be employed to ensure efficient operations, reduce yield loss, and effectively control the complex chemical-process techniques.

2. Overview of Semiconductor Fabrication

Semiconductor fabrication is a crucial process in which various methods and equipment are utilized to fabricate integrated circuit (IC) chips. The fabrication process is divided into front-end-of-line (FEOL) and back-end-of-line (BEOL) processes. The FEOL process consists of wafer slicing, polishing, doping, and thin film deposition, among others, and involves the critical step of patterning circuits and structures on silicon wafers through photolithography. After the FEOL process, BEOL processes, including chemical vapor deposition, etch, chemical mechanical polish, and metal mask deposition, are employed to fabricate the wiring layers between transistors with a target process yield of over 99.8% [1]. The number of equipment pieces used in IC fabrication plants (fabs) is enormous, and the cost required to build a fab to manufacture the latest IC chips

reaches billions of dollars. After completing the main semiconductor fabrication processes, the wafers are sent to wafer probing to test the performance of the chips and detect faulty dies before packaging. A significant challenge faced in the IC industry is keeping the high yield, as any failure in the processes can cause a large yield loss. Furthermore, a new technology generation brings about degraded yield due to difficulty in integration and compatibility. Fab and equipment vendors put great effort into seeking possible reasons and identifying root causes for yield loss by analyzing mass technological and diagnostic data acquired from the processes .

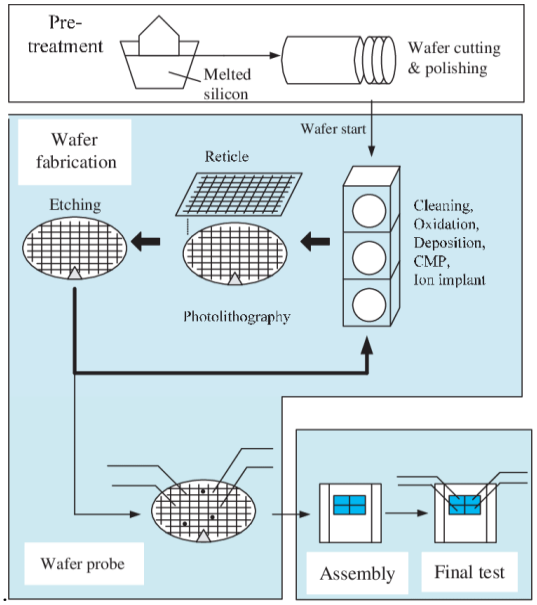

Fig 1: An Introduction to Semiconductor Foundries

The goal of fault detection is to automatically identify such abnormal events by modeling normal behavior, and recently, extensive efforts have been made to build machine learning (ML)-based fault detection models that do not rely on predefined assumptions. The benefits of early and accurate fault detection are most apparent, as they provide engineers with more time to avoid severe equipment malfunctions or changes to the production process to prevent faults from occurring downstream. In addition, it enables more effective collaboration within the fab by reviewing process issues and finding the best-fitting equipment, thus enhancing the quality of production and reducing the additional cost of manufacturing faulty chips. The significance of building ML-based fault detection systems are as follows: Early fault detection significantly reduces downtime and the cost of equipment maintenance; Early fault detection allows production instability and quality issues to be problem solved, improving production quality; Prompt diagnosis throughout the fabrication process dramatically reduces time spent per incident.

2.1. Process Steps in Semiconductor Manufacturing

In semiconductor fabrication, a large number of process steps are performed for manufacturing chips. As of 2021, about 160 steps have been performed for manufacturing chips using the 5nm and 3nm processes, and it is estimated that over 180 steps would be performed for the 2nm process [2]. In addition to process complexity, the fabrication cost of chips has increased to over $10B since more than a few hundred million euros were needed to build extreme ultraviolet lithography equipment. To attain high productivity and yield, online and real-time anomaly detection of the semiconductor fabrication processes has become critical [1].

Another issue in the fabrication of semiconductor chips is that random particle contaminations have become one of the major causes of yield and processing issues. Random particles can deteriorate device performance by producing shorts and leakage paths. Particles can also cause local geometry variations that lead to unacceptable edge steepness. Such defects arising from particle contaminations are difficult to detect due to their insignificant effects on local electrical parameters or variability of the effects on global electrical parameters. As a result, a particle that leads to catastrophic yield loss can be missed, while a harmless, outlier particle may lead to a gross error in chip performance metrics.

With increased device integration density and evolving sophisticated integrated circuits structures of semiconductor chips, detecting defects becomes much more elusive and complex. At the end of the manufacturing process of semiconductor chips, parametric testing is performed; the electrical parameters of fabricated chips are measured, and the chips are classified into "good" or "bad" states. However, with fast increases in chip complexity and the number of new types of defects, the occurrence of a new type of failure or distribution change of the data is common, but it demands retraining of the model. During the manufacturing process of semiconductor chips, on the other hand, the electrical parameters of a tested lot, consisting of thousands of chips, are not available. Only the detailed fabrication conditions and metrics of inline monitors are available for the process steps where the lot wafted are processed. This makes detecting defects in a single-pass online fashion much more challenging and is often favoured.

2.2. Common Faults in Fabrication

The fabrication of integrated circuits (ICs) involves several processes and it may be affected by various faults which may be detected at different levels. This work focuses on die attachment process monitoring at DAF level. Die attachment plays a vital role in the reliability and performance of any IC. Some common defects in the die attachment include die lost, die offset, die tilt, batch variation, and not attached die [4].

D.1.1. Die Lost: During the assembly of an IC, the chip is glued to a ceramic or plastic substrate, and in the die lost case, there is no IC on the substrate.

D.1.2. Die Offset: During DAF, the die might not be aligned to its pad which may in turn increases the likelihood of short circuit with adjacent pads.

D.1.3. Die Tilt: It is associated with a change in Z-position of the die that may lead to misalignment of the PAD.

D.1.4. Batch Variation: Due to the variations in the chip fabrication, there might be a different shape of the die.

D.1.5. Not Attached Die: In some cases, the adhesive is either missing or insufficient which leads to unreliable dies.

In the context of die lost and incorrect die offset, a modified DAF routine is presented to detect these faults. This modified DAF relies on the state changes of some specified pixels among the pixel values on and around the die region. To be specific, the pixel values in the surrounding region below the bonding pad of the die will be chosen.

The proposed algorithm comprises two parts: two rules that will detect die lost and incorrect offset die respectively, and a post-processing step. The proposed method has been tested on real-world data obtained from the fabrication facility of a large semiconductor company. The die lost detection algorithm can detect lost dies with a speed of a few milliseconds per die using a common personal computer.

3. The Role of AI in Fault Detection

The emergence of AI can be described at two ends, one being a revolutionary tool by which representing and manipulating knowledge is moving closer to humans and approximating them, impacting a broad set of industries, and another being that of an unsustainable technology fuelling the unequally shared wealth generation and threatening our environment [3]. While both are views have their merits, there is no denying that AI is already impacting how corporations strategize, implementing intelligent process automation spanning robots to tools. The semiconductor industry has deployed AI in different domains across the design and manufacturing of ICs or chips, such as circuit design, characterization, packaging, and fabrication. Taking the latter as an instance, fab machine logs contain N limited fields and multiple periodic records found in M unique processes such that N*M samples produce cumulative insights into fault sources.

Fault detection is a process where logs are checked before the production of wafers and chips so as to determine whether incidents matching the signature of defective goods have happened. Per the rule-based methods available today, domain experts manually define rules that match the point, kind, and case association properties of logs, falling short on changes and repairs such as super-criteria or joint recycling. There is plenty of room for improvement and effort for AI-enabled methods on fault detection which automates such tasks and expresses knowledge in models instead of rules. A dominant class of methods is to build a supervised classifier that learns the representation of a fabric process and its ideal abnormality over cleaned data.

Using records for data augmentation, techniques such as post-fabric tortoise shell disambiguation matching improve both loss and running time. Recently, the use of knowledge graph path matching expands the approach beyond supervised learning, allowing normal logs alone in training. Queries are first parsed into sub-paths and parameterized paths. Paths are matched before and after parsing via types and entities, extending to probabilistic aggregation paths. The increasing data volume to process, initiate, transfer, post- or triggered by AI might cause growth in other attributes that change, overall detouring from being lean-on, nimble-on, or green-on.

3.1. Machine Learning Techniques

Computing technology has permeated every aspect of work and life and continues to evolve at a breakneck pace, from giant mainframe computers, supercomputers, and workstations to personal computers (PCs) and laptops, and now the rapidly ubiquitous emergence of mobile computing devices and smart phones [5]. Even so, the industry is continuously striving for further miniaturization of the semiconductor components that are at the core of these computing devices. In the last few decades, Moore's Law has brought continuous improvement of the semiconductor fabrication process with architectural scaling, performance enhancement, and reduction of power consumption. In order to realize such a challenge with larger and larger die sizes, new materials and structures, such as High-k/Metal gate, FinFET, or Gate-All-Around FET devices, were introduced in modern semiconductors. At the same time, the new increasing complexity of the processes requires more innovation in the process control and monitoring technology [2].

Techniques, methodologies, and tools to prod both wafer fabrication and device manufacture were developed for fault detection, e.g., Statistical Process Control, in-line Wafer Inspection or Post-Path-Finding fault localization and recovery, Wafer Quality Estimation, Failure Analysis, Yield Learning/Ranking/Prediction, Areal Defect Density Estimation, and Defect Figure Annotation. With the show of success on various domains for medical imaging, image editing, image generation, etc., a flurry of new research activities has arisen to extend these cutting-edge AI technologies to the semiconductor industry and have been proven successful in several specific applications.

Nonetheless, in most of these challenges, accessibility to high-quality labelled imaging defect samples is a roadblock inhibiting the development and deployment of AI solutions for better productions. In order to tackle this, unsupervised machine learning methodologies are being actively applied to semiconductor industries to admit the economic application of suspension defects in wafer fabrication. Wafer defects become a major concern since recently introduced Advanced High-k/Metal Gate technology significantly reduces device operating voltage levels, which causes enlarged impact on the yield performance of

CMOS devices. Therefore, approach towards a machine intelligible defect pattern recognition and classification technique is warranted to reduce the load for defect annotation.

3.2. Deep Learning Approaches

Automated visual inspection (AVI) along the continuous flow of production is a key instrument in ensuring the quality of manufactured semiconductor chips. Ideally, an AVI system continuously monitors the process over time. Early detection and removal of immediately visible defects allows for increased production yield and thus large cost savings. By tracking the progress of defect patterns over time throughout the production cycle, future defect classes can be anticipated and countermeasures can be taken early on. Semiconductor manufacturing is a very complex process chain consisting of many steps and controls, offering a lot of potential for defects introduced to the manufacturing system.

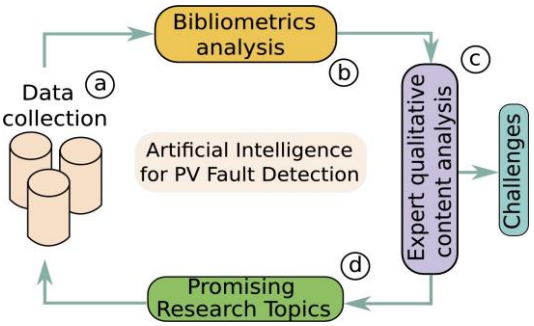

Fig 2 : Fault diagnosis of photovoltaic systems using artificial intelligence

Typically, defects are recorded in complaint tickets. Each complaint ticket is analyzed in a fault detection process to match the defect pattern with a fault model. Both the fault models and tickets are stored in databases. During the fault detection process, the software runs through a stage-wise process. First, the ticket information is checked and pre-processed in view of the faults to be analyzed. After that, all fault models are transformed from their normal representation into one that matches the ticket format. Finally, identified candidates are inspected.

A novel Visual Fault Detection and Classification (VFDC) system for semiconductor manufacturing is proposed, based on Stacked Hybrid Convolutional Neural Networks. The system aims to detect and classify different types of defects in binary defect maps, which highlight the defect pattern of interest, against a reference map, which is a cleaned defect map. It consists of two parallel sequences, where both extracts global image features to classify the raw input map while learning either the multi-scale features.

Machine Learning approaches can replace this early stage of the fault detection process and run the technology transfer approach completely. The task of machine learning is to detect defect patterns that are visually similar to previously analyzed ticket patterns. However, the variety of different fault classes leads to highly unbalanced data sets. In ML-based detection, this can result in models which classify defects as a majority class, while neglecting proper classification of important minority classes, depending on the approach used. For such use cases, images with sufficient examples for both majority and minority classes initially need to be collected. In the semiconductor industry, where chip layouts are designed in complex Block Diagrams, BS with potential defects and subsequently collected images can take up valued resources.

4. Data-Centric Approach to Fault Detection

In semiconductor wafer fabrication, 2D images are acquired for evaluation and quality control, and specific panes within these images contain useful information to facilitate fault detection. However, large-scale testing results in excessive consumption of both storage and computation resources. To overcome these challenges, efficient data preprocessing is necessary, including disk removal and down-sampling, which help concentrate all the defects within a pane while discarding the unimportant areas, thus significantly decreasing the model complexity

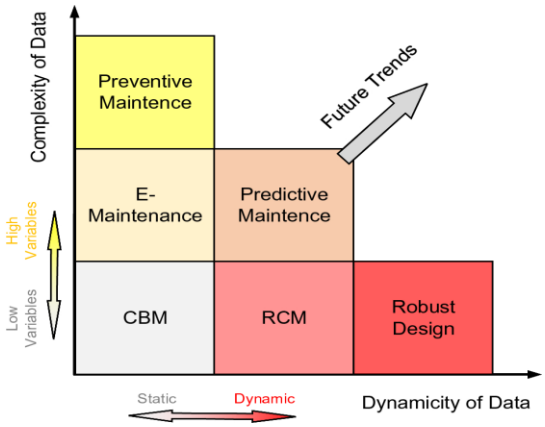

Fig 3 : A systematic review of data-driven approaches to fault diagnosis and early warning

Although numerous works in academia and industry are constantly helping to improve the fidelity and accessibility of tagged data, data poverty remains a hurdle in industrial applications. As a general recommendation, when insufficient amounts of labeled data are provided for training, researchers should consider utilizing labeled datasets from other domains that are similar to the target domain. Some unsupervised learning methods, such as clustering, can also be helpful to estimate the classes even without using any labeled images. However, when even the bearing manufacturers do not provide any labeled data, the option of semi-supervised learning can help researchers use notable amounts of unlabeled images, as well as a few images that are tagged manually, or about 100 frames. When the targeted fragments are relatively small, and attentional regions outside the critical area take up a considerable part of the original frame, using a 0.5-1.0 pixel range can help to improve the detection ability. The small grids can be filled with the corresponding probabilities and added to the loss function to train a superb data-centric model.

As the size of models continues to increase dramatically, the computing burden grows, making it almost impossible for researchers in normal institutes to retrain or even fine-tune the models with public weights. Self-supervised pretraining of imaging tasks based on random cropping, flipping, and rotation helps to improve the performance of the model in the limited datasets. Masking a part of the seed images and then recovering them can also boost performance and robustness. Researchers are also encouraged to design robust and efficient pipelines for industrial scenarios where data is acquired primarily on the fly. As Chinese semiconductor equipment companies become increasingly competitive, data security has become an issue of great concern since the malicious tampering of the ground truth can easily result in severe economic loss and jeopardize national security.

4.1. Data Collection Methods

In semiconductor manufacturing, the data were obtained from equipment processing data measured with sensors through the manufacturing process to detect abnormalities. The data consisted of a collection of values sequentially ordered over time, and each dataset was a collection of time series data for the same process through the same procedure. On purpose, the data were finally collected using the same processing equipment and the same process. Rawdata includes the results from 260 sensors. In the experiment, only tilt 24 data were used, which were used in [1] and the signal ranges were scaled to be in the [0,1] range. The purpose was to detect 14 anomalous signals from normal signals. The noisy signals were also handled through a simple smoothing method.

The semiconductor fabrication has a lot of cleaning steps to remove fluids which adhere to the wafer. Here the tag based on the cleaning type and the tags indicating whether the wafer had dropped after came out from the chamber were considered as the target attributes preventing them from getting into the history. The cleaning steps had various durations which were longer than or comparable to the signal lengths. Although the abnormal processing parameters were given along with the normal processing parameters, it is not possible to handle them in training if the different types were not specified as the target classes. Hence, only normal signals were used to check how well the cleaning steps were detected.

Semiconductor manufacturing is composed of many different processes. The detected fault identifies the process maturity that the equipment underwent, which usually has various levels of severity. With some knowledge about the signal extraction process, fault history logs can help to improve the detection performance.

4.2. Data Preprocessing Techniques

Semiconductor fabrication processes, which are operated in factories, result in semiconductor devices made from silicon wafers. This process can be non-stationary, and the acquired data from sensors mounted on the machines used in the process can vary widely depending on hardware settings, environmental conditions, and others. Also, these raw data for each device are often very large and complex: in manufacturing automotive components, one wafer can have from 70,000 to approaching 1,000,000 individual measurements per device. Thus, as semiconductor fabrication technology advances, a real-time fault detection system that can detect abnormal devices within a few minutes before those semiconductor devices are tested further is required. Time-series classification models have been extensively studied, which could accurately classify snippets of time-series data representing multiple types of signals. However, while sensors continuously record raw data over time for many fabricated devices, these devices cannot be classified into “normal” or “abnormal.” This is because semiconductor fabrication is expensive, and no further testing can be conducted for abnormal devices, as they are entirely audited or scrapped. Most sensitivities of data points are close to each other in normal devices, while abnormalities can induce larger sensitivity shifts in a few data points. Thus, the next goal is to obtain abnormality points from sensed time-series measurements of an unclassifiable dataset based on a one-class approach that classifies only a “normal” device.

Semiconductor manufacturing and the diffusion process are examined for such a problem, and real raw trace data acquired from an equipment maker’s Test/Sort environment are analyzed. The main contributions include a preprocessing approach using dimensionality reduction techniques and a one-class classification-based fault detection method using various machine learning models. The preprocessing technique improvements lie in understanding input raw trace data to defect detection models and overcoming noise or measurement variations. Commonly, mass fault detection has been performed using a designated machine learning model trained with previous readily obtained datasets. However, the input datasets for fault detections in the semiconductor industry differ from those in conventional manufacturing because there are no previously obtained defect datasets using sensors in the factory. Hence, preprocessing techniques are required to either enrich the defect dataset using the weakly-observed defect labels or enable proper training of machine learning models to classify manifold shapes projected in large dimensional measurement spaces.

5. Feature Engineering for Fault Detection

In the semiconductor industry, high-precision performance of tools is essential for production processes. The requirements for thickness, depth, diameter, and angle of holes in the silicon wafer are extremely strict, as minute defects cause chip defects, preventing normal system operation and significantly degrading reliability. Therefore, it is essential to identify failures by analyzing patterns in the data collected from tools that conduct such processes. A pattern recognition framework that detects tool faults in real time, processes and analyzes time-series data, and distributes alarms to users was developed.

To detect faults in a semiconductor fabrication plant, the automated pattern-recognition system presented was applied to error data generated by an etch tool used to cut holes in silicon wafers. As the widths of drill holes become thinner and the number of drill holes in a wafer increases, the steps involved in etching holes on wafers are more time-consuming, resulting in a longer idle time in the fabrication tool park. Each individual and distinct state is regarded as a class. It is critical to detect state misalignment in order to avoid idle times in idle design communication lines. State misalignments in tool faults in large fabrication plants can lead to severe processing delays. Identifying the state of the tool in real time is essential for tool-park scheduling. The implemented system could identify the state of each tool in real-time with greater than 90% accuracy.

Batches of capitals are loaded before and/or after fabrication and remove tooling faults occurring inside the capital equipment in semiconductor fabrication tools. If defects can be effectively recognized, the time required to replace parts can be minimized, allowing coin-counting-removal operations to be performed in a timely manner. Toward this end, an automated fault-detection and recognition system based on hidden-error pattern tree reducts was developed. The system detects defective events (error events) by monitoring the status of system error log files on each tool. The pattern tree reduct features are optimal features built in a hierarchical fashion by refining them with respect to classes and errors during the fault-reduction phase. The performance of the implemented event-detection and error-recognition systems was evaluated and various testing data revealed that the detection of good-to-bad natural errors is highly successful while the classification performance of catastrophic defects is affected by their pattern similarity [6].

5.1. Identifying Key Features

In semiconductor manufacturing, various factors can cause defects, such as improper process, tool malfunction, and excessive variation of the tooling, as each step of the fabrication involves complex equipment and conditions .A fault detection and classification (FDC) system detects and classifies faults and analyzes causes in manufacturing. With deep learning techniques, automated and accurate FDCs have been developed; although a few studies have applied multi-step state-of-the-art deep learning models in the diffusion process, recent advancements in one-shot or zero-shot learning approaches are yet to be explored. Predicting the proper recipe to deduce non-faulty signal info from process variables can improve the performance of one-shot detection. Application of a supervised segmentation study on opacity data provides the timing of defects and earlier detection possibilities. In semiconductor fabrication, there are hundreds of steps, such as diffusion, etch, deposition, photolithography, and cleaning, and numerous characteristics are generally measured by advanced inline metrology. Non-normal measurements can occur due to various factors, such as poor matching of recipes, incapacity of tools, and change in consumables, which can cause degradation of the product, must be eliminated at an early stage, and of utmost importance in manufacturing. Thus, an anomaly detection and classification (ADAC) scheme that can both detect faults and classify them into diagnoseable types should be developed. However, this as-of-now unexplored topic is quite difficult due to several challenges and demands the development of a novel method. First, an extremely high FDC performance is required because faults directly impact productivity and yield. Some normal and abnormal data show only subtle differences, and general classification algorithms may have limitations in achieving high performance. Thus, unlike conventional models that classify only a single noticed anomaly at a time from the pre-collected batch of totally different anomalies, monitoring signals should be assumed to compose redundant anomalies of possible candidates, which are then filtered by adaptive attention clustering. Second, a critical problem is that the frequency of abnormal data is remarkable low; therefore, the data that engineers can use to develop a model is limited.

5.2. Feature Selection Techniques

As various complex manufacturing processes take place in modern manufacturing systems, advanced sensors generate massive amounts of process data. One of the main objectives of data collection is to detect abnormalities or faults of the processes immediately after they occur. For effective fault detection with great performance, choosing relevant features from a collected dataset is crucial [6]. Semiconductors have become an essential component in diverse industries, for instance, automotive, computer, and telecommunications, and ability to detect faults has become more crucial in high-tech and large-scale manufacturing of semiconductor fabrication processes. These systems consist of hundreds of machines executing different processes at each process step. For these fabrication processes, failure of any machine might lead to expensive fabrication delay and yield loss, hence intelligent automatic faults detection is necessary to detect potential faults early. In deep learning-based classifiers, every process signal must be treated as an independent feature, which generally incurs a high computational cost [1].

A framework is proposed to automatically select relevant features from massive process data for fault detection of semiconductor manufacturing processes. To most optimally capture the temporal relationship and patterns of individual process signals, sliding windows with a fixed size are applied on processing signals of a chosen time interval, and the extracted windows are preprocessed with fast Fourier Transform and Principal Component Analysis before entering a model. With all the signal features converted into a fixed-length representation vector, the window-level features are aggregated into the overall signal-level feature by training a weighted k-means clustering. Then, with the automatically extracted signal-level features representing every signal, an ensemble classifier is trained based on different classifiers' predictions.

This automatic framework essentially comprises two parts: i.e., weighted k-means clustering and aggregation, and ensemble classifiers. The first part is to obtain overall signal-level features of a large number of individual signals which can represent the signals with substantially lower dimensions. The second part converts the individual signal-level features into the final features by combining predictions of multiple base classifiers, which may help to better generalize the predictions. The overall framework is also readily incorporated with any kind of signal representation.

6. Model Development and Training

A semiconductor fabrication model is constructed using ResNet-101, which is trained with a dataset that contains a synthesized set of increasingly challenging images. The synthesized images are preprocessed and classified into positive and negative image chips, and resized to image chips with a fixed dimension of 640 × 640 pixels. The network is altered to output two bounding boxes for the defective class and an object mask that corresponds to the defect.

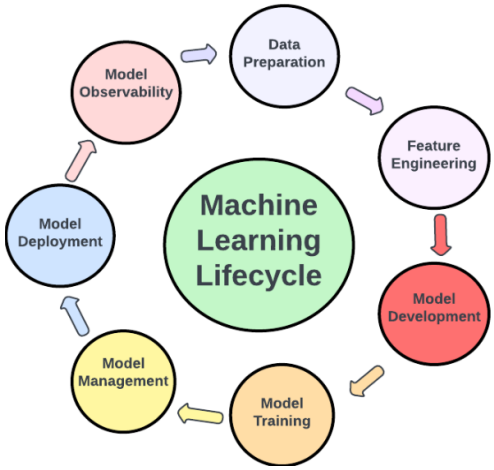

Fig 4 : Demystifying Machine Learning: Comprehensive Guide to Development

The trained model is validated against a validation dataset to explore the area under curve metric. The trained model can significantly reduce false positives by raising the probability threshold for a defective prediction. The image pair with the lowest probability score is used to visualize the predicted areas for both states after passing the probability threshold. The visualizations also revealed that the defect regions predict high probability scores also appear in the innocuous pair, which generates ambiguous predictions.

The trained model is tested against a testing dataset to evaluate its generalizability against defective images generated with a style transfer algorithm. This key challenge is highly significant as, in a real fabrication setting, the model should work autonomously on wafers and detect unseen defects of unspecified widths and styles. Comparison on detection quality shows that detected defects with significantly reduced bounding box areas and mask scores can be filtered and classified as innocuous. It is noted that the inventive generation server can be built at a low cost by training and running a model on a laptop. Good generalizability can be achieved by feeding a naïve model with enough diverse synthetic samples. The success of the proposed end-to-end defect detection framework and underlying technologies demonstrate viable pathways for the community to harness deep learning as a tool to build for semiconductor fabrication chips.

6.1. Choosing the Right Algorithms

Selecting suitable AI algorithms for fault detection in semiconductor fabrication is a complex task due to multiple factors including the type and size of data, device characteristics, the variety of fault detection challenges, and the model type. To maximize the potential of AI algorithms and enhance asset performance, expertise is needed for careful selection of model architecture, hyperparameters, input features, performance evaluation methods, etc. In regards to algorithmic characters and possible combinations, execution can be a multi-stage pipeline. The design choices focus on the following aspects. Most manufacturing processes are high-dimensional dysfunction signal data. Featuring dimensionality reduction like Principal Component Analysis (PCA) is usually required. With PCA, the original high-dimensional data can be represented in lower dimension without significant accumulated variance loss, and the manufacturing data will precisely yield interacted variances. Time series models such as Long Short-Term Memory Neural Networks (LSTM) are sufficient to model data for continuous processes like semiconductor fabrication, especially bidirectional versions that can make real-time prediction to both past and future data as well as detect abnormalities hidden behind extensive time lags [1]. Anomaly detection methods usually need to balance two conflicting requirements, processing time and performance. Therefore, a two-stage fault detection network can be designed as a discriminative one-class classifier followed by a generative score mapping function. Different performance evaluation measures are required for different tasks. As for classification tasks, detecting rare classes is equally important. Therefore, metrics including True Positive Rate (TPR), the area under the receiver operating characteristic (ROC) curve, etc. This category can be regarded as classification models separating disjointed feature spaces. On the contrary, as for regression tasks, quantile intervals must be estimated since the target quantity is a continuous one [2]. AI fault detection has been widely adopted in the semiconductor industry amid rapid progress of either fault detection methodologies or semiconductor technology evolution. AI tools have been successfully applied to synthetic image generation, inline equipment issues detection,

yield prediction, manufacturing equipment prediction modeling and optimization, wafer defect classification and defect density estimation, etc. Manufacturing abnormal events are commonly discovered via visual inspection or pre-defined thresholds. However, as the scale and complexity of semiconductor fabrication evolve, the effectiveness of these human-centric strategies has become increasingly limited.

6.2. Training Models with Real-World Data

Recent events spur a growth of interest in AI techniques to be deployed in semiconductor fabrication lines. By the convergence of increasing availability of process and equipment data and advances in machine learning techniques, AI models are now being developed and deployed in semiconductor manufacturing companies.

New deep learning models can be trained directly on time series data stream from fabrication equipment. Training these models is often challenging on account of domain shifts, and a trained model often becomes out-of-date after a short period of time. Therefore, AI techniques that enable re-training of models with a minimal amount of annotation and perturbation of normal operations are pursued.

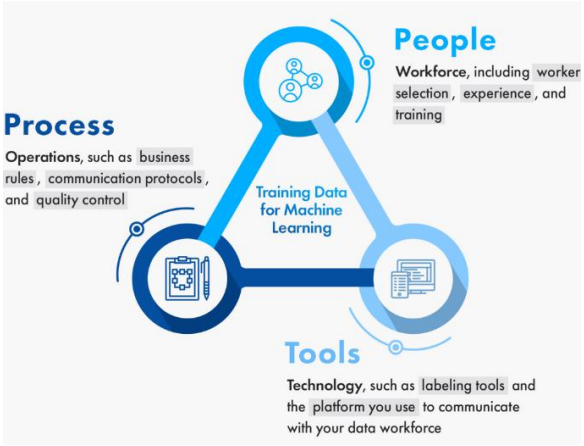

Fig 5 : The Essential Guide to Quality Training Data for Machine Learning

Classification models based on simple additive smoothing have been applied successfully to batch process events detection. Defects on fabricated wafers that are the outputs of the fabrication process are modeled using a model built on engineered features taken from fabrication equipment signals. State-of-the-art deep learning models have been developed to model defect classes directly from process signals and classify the defect classes for each wafer. A novel interpretation framework based on input taxonomies characterizes how the models make predictions. This interpretation tool is used to identify the collected signals of the process that contribute most on defect classification, and suggest extension of the AI models to increase generalization.

7. Evaluation Metrics for Fault Detection Models

This section presents relevant metrics for evaluating the performance of data-driven fault detection models. As a subset of classification tasks, the goal of fault detection in semiconductor fabrication is to determine whether a test sample belongs to the normal class (indicating no faults in the manufacturing equipment used for testing) or anomalous class (indicating a fault in the corresponding manufacturing equipment). In this work, random noise is injected into the dataset so that it is hard to distinguish between "0" and "1" by the fault detection model.

The performance of the model on the test dataset is analyzed using AUROC (the area under the curve of the receiver operating characteristic) and AUPR (the area under the curve of the precision-recall curve) [2]. The AUROC metric is computed based on TPR (the true positive ratio) and FPR (the false-positive ratio) obtained by varying the threshold of segmentation scores that are computed using a mixture of Gaussian distribution model over the clustering labels. A TPR of 1 indicates that none of the faulty samples are misclassified as a normal class, whereas an FPR of 1 indicates that all the normal samples are incorrectly classified as faulty samples. The AUROC varies from 0 to 1, and a higher value indicates a better performance of the model. Similarly, precision is defined as the ratio of true positives to all samples predicted as positives (abnormal), while recall is the ratio of true positives to all actual positives in the prediction [1]. A higher precision value indicates fewer false positives in the prediction; meanwhile, recall indicates how many positive samples are correctly identified.

By varying precision/recall thresholds, PRC (precision-recall curve) is plotted, where the area under the curve is AUPR. As with AUROC, a higher value of AUPR indicates better performance of the model. All metrics score from 0 to 1, and if the predicted results of the models are perfectly similar to the ground truth, high values are obtained.

7.1. Accuracy and Precision

With the fast advancement of the semiconductor fabrication process node, AI is being applied to semiconductor manufacturing for various tasks including fault detection and anomaly detection. Acquiring data is fundamental for AI training and predictions. In the semiconductor industry, cleanroom facilities are highly sensitive to human activity. Notable equipment appears individually under the cleanrooms and they produce vast amounts of signals. Anomaly, time series, or process change resultant response in a 100,000 frame minute video stream is hard to observe. As a consequence, AI trained with a minute

frame series is hard to generalize, mostly due to the dramatic temporal and spatial changes [2]. Time series data analysis is limited in learning toward and identifying high-dimensional cleanroom facility faults.

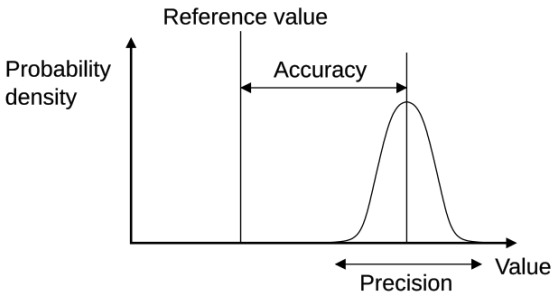

Fig: Accuracy and precision - Wikipedia

The AI-enabled manufacturing ecosystem allows the implementation of data-centric or AI-centric approaches. Instead of considering why, when, how, and who a fault happened, the fault symptoms, time series deviations, or anomalous signals are focused on. However, the distinct performance of failed manufacturing signals creates challenges in transferability, rendering the generalization of AI fault-detection unachievable. On the other hand, minor faults and dense events hard to observe are unavoidable in manufacture processing [3]. For the aforementioned explosion of fabricated signals, it is difficult to process individual signal deviations or observe.

The fabricated surface is smooth and nondifferentiable. A Taylor-like numerical method and a contradiction-based optimization problem are firstly proposed for discrete surface optimization. The condition of global stability is established by assuming that each group of the optimization problems has an optimal solution. The method has second-order convergence. The local minimums are estimated by randomly selected grid points. Interesting results can be obtained if the fabricated surface maps are retained in some shape features. Finite element flexibility modes of the invalid mapping surface model are yielded using the computed input artificial noise force and the available output sensor locations.

7.2. Recall and F1 Score

In this study, recall and F1 score are calculated to evaluate the performance of anomaly detection systems against various failure datasets. The results of the anomaly detection systems indicate that the recall rate is significantly improved when applying the new methodology. Since a high recall rate is critical for FDC, systems achieving a recall rate of above 80% in all datasets can be considered valid. The results show that some systems do not detect any single abnormal data. The major reason for the failure of the one-class classification method is that the density functions of normal and abnormal data overlap too much. Systems SN and SP do not detect any abnormal class in the D1 dataset because they are very similar to normal data and show only subtle differences. These thresholds can be easily tuned down to increase recall at the expense of precision. It shows that the proposed decision criteria are effective in detecting anomalies at various levels of severity, especially when the false-positive rate is set at around 1% [1]. The F1 score is calculated using True Positive, False Positive, and False Negative values as follows: where recall is also calculated with respect to the above values. This study is evaluated based on these metrics in the same time period: (1) the absence of intentional anomalies before the fault, and (2) anomalous signals of various severities. The anomalies may not have sufficient time to result in an apparatus failure. True Positive values are defined as the signals that were conditioned out before fault operations. Variable sets VR3 and VAR2 showed a similar rate of anomalies since they have shown very similar results in all three datasets [2]. Regarding variable sets that led to abrupt deviations, if a single application is chosen, variable set VAR3 is selected as the best set. On moving further towards the other selection criteria, either F1 scores become very skewed, or the data are considered to be just noise.

8. Implementation in Semiconductor Fabrication

Among various AI chip domains, semiconductor fabrication is the main core AI-chip technology that necessitated the need for the AI revolution. Semiconductors account for a major revenue source for technology giants. It is the foundation for the general advancement of technologies including AI chip implementations, wireless communications, automotive electronics, smart homes, portable devices, etc. A semiconductor fabrication includes exceeding hundreds of processes and capital costs million USD, such as ion implantation, chemical vapor deposition, lithography, etc. If a defect occurs during any of the fabrication processes, it might affect the end semiconductor chip performance. In other words, detecting and ensuring the functional yield increase of semiconductor fabrication is an essential and challenging task. Recently, with the growth of AI technology, machine learning algorithms have been introduced to detect faults during the manufacturing of semiconductor chips.

$$R = k_1 \frac{\lambda}{NA}$$

Eqn 1 : Lithography – Resolution Limit

Where:

- R: minimum resolvable feature (resolution)
- λ: wavelength of light
- NA: numerical aperture of the lens system
- k_1: process-dependent constant (typically between 0.25–0.5)

Implementation Use: Guides decisions on equipment (e.g., EUV vs. DUV) and resist materials.

Semiconductor chips are in high demand for most of the electronic devices. As they become denser, detecting defects, especially new types of defects not seen before, is becoming very challenging. Traditionally, in semiconductor preferred machine learning for failure analysis and fault detection, input features and classification models are offline batch mode training based on legacy faults. However, this batch mode training approach is deemed ineffective for new types of defects with the challenging critical requirement of no retraining. In one-pass online fashion, new types of defects are detected based on the original legacy feature space. This shift is mainly due to the growth of chip architecture complexity and heterogeneity. Not only is the IC algorithmic design itself more complicated, but the hardware growth also demands heterogeneous and fine-grained implementations. While the 'Rectify After Failure' method works with legacy defects, the adversities raised by new types of defects on such advanced technologies pose a serious challenge.

With the rapid advancement of technology nodes and high integration density, a yield deterioration is observed in semiconductor fabrication. In FPGA, several chip-level tests are usually built in during fabrication for fault detection and yield improvement. Although they are successful for Silicon-On-Insulator technology, many new failure phenomena are emerging with the rapidly decreasing 20 nm node size, including gate-middle contacts shorting, transistor bridge defect, unreliable ground ring formation, etc. The massive increase in the number of design configurations along with the advent of the opportunity for configuration and interconnection errors is making it difficult for safety-oriented verification and test-based redundancy approaches to detect faults. Thus, a need is there for a systematic automatic test generation method that can produce high-quality logic setting tests without considering the original design information. A new test method and an online test system are developed to implement this idea effectively. An importance ratio measure is proposed to characterize the significance of each design configuration and an effective Monte Carlo based test pattern generation method is proposed to select critical configuration settings to be tested.

8.1. Integration with Existing Systems

To successfully deploy the model for future semiconductor chips and process flows, close collaboration with domain experts is required. In this section, a comprehensive approach to ensuring successful integration and adoption of embedded AI is provided, broken down into four stages: prototype testing, R&D tool validation, technology qualification, and ramping-up production. At the R&D lab, an AI-customized defect map will be generated to characterize the initial prototype performance, analyzing cases of interest and false alarms. Collaboration with defect review teams will identify potential improvements encompassing both chip and AI. Altogether, improved conditions describing the initial chip model will be scripted back to the chip design teams for future chips. Once testing of the initial prototype matures to the satisfaction and needs of the teams, a foundry-ready R&D tool will be designed with minimal changes to transmit best proven technology nodes and processes. Before chip deployment, a qualification stress test will be conducted with both random and engineered defects targeting the maximum range of predicted manufacturing needs. This phase will focus on age forward evidence of D2S usability and, importantly, false alarms. Once all defects are executed, a comprehensive analysis will determine model weaknesses and missing features and conditions. Once a target defect productivity is found for the R&D, the production ramp-up can begin with full team-wide training and transition to the production tool vendor. Before full production, safeguards should be implemented for best practices, regular checks, and defect collections to prevent model drift [6].

8.2. Real-Time Monitoring and Alerts

To achieve real-time monitoring, the importance of accurate alarm and warning escalation is presented. It refers to producing an alarm depending on the style-potentials or fault condition severity. It can be processed in the host or a separate cloud device. On chip system, where memory and power are limited is discussed after a chip test is performed. The accurate alarm generation architecture is proposed. Importance of the alarm and warning escalation is generalized as a mapping of detected faults to proper safety escalation levels. An architecture based on this idea is proposed and validation case studies are presented with simulated, fault injected and digitized test data. Further study in alarm process design and safety breach analysis based on fault detection is projected. Alarm systems, widely adopted to detect security breaches, are typically implemented with software, vulnerable to software-based attacks. Fault data acquisition module is adopted for circuit hardening, using controllable attack models, testing error propagation path/scale and alternative actuation. The experiment shows circuit exposed to fault injection, not only keep its safety but also generate fault-tolerable alerts. Since semi-digital systems conduct RF signal generation and mixing in highly-abstract digital domain, it is essential to ensure RF circuits safely handle signals of diverse encoding schemes to achieve skewed performance. Statistical verification using large number of samples with Normal distribution fails to assure incorrect measurements if they occur, due to randomly generated signals covering only a small portion of the entire input space. To prevent this, an analog model-checking technique is developed that dynamically generates edges that easily cause incorrect behaviour in pre-specified timings by cross-simulation of a digital specification and a mixed signal one. With large number of the edges, a model-checking engine finds validation candidates through Binary Decision Diagrams. The method finds a counter-example for the case that the mixing ratio of signals is relatively small and prevents cascaded neighbors from confusion. The number of the counter-examples is orders of magnitudes smaller than tests on the entire input space. In designing a safety-related device architecture, alarm escalation flows dealing with different levels of under-performance conditions are usually required. It describes the assumptions for alarm escalation calculation and its output representation at

each alarm level, together with development of derivations to automatically adapt escalations for accustomed architectures. All these help designers reason on safety requirements more efficiently.

9. Case Studies

AI-powered fault detection techniques have been studied and applied in semiconductor fabrication, including deep learning-based time series classification during the diffusion process and online learning methods for post-fabrication detection to increase speed. Edge computing facilities also have potential applications in semiconductor fabrication, leveraging deep learning methods to detect anomalies. With great efforts in data collection and engineering, a production-ready anomaly detection system has been developed, enabling control of both data collection and data entry for batch-training. Defects in semiconductor chips are elusive and complex, and it is a challenging task to analyze signals emitted from chips to gain insight into defects. Machine learning (ML) has been a powerful technology for failure analysis, and ML-guided failure analysis is conventionally performed by offline batch mode training with new collected signals. However, new types of failure in manufactured devices and changes in data distribution demand retraining the model. Even worse, the unavailability of physically damaged chips further enrains the need for a new model with a different distribution of new incoming signals [2].
The training data are drawn from a stream of observations where each observation arrives in order. The decision maker must choose to accept or reject the current observation. Accepting the observation requires incorporating it in the existing model, while retaining a statistical summary sufficient for future decisions needs to be thought about the accepted observation. One-level quantile learning is first proposed for switching to a new observation distribution. An extended version of quantile learning is similarly proposed for switching across multiple distributions. Finally, the quantile online learning method is tested on various existing benchmark problems under both point-wise and quantile predictions. The proposed method is applied to semiconductor device-level defects: FinFET bridge defect, GAA-FET bridge defect, GAA-FET dislocation defect, and a public database: SECOM. Positive results observed in numerical studies indicate that quantile online learning constitutes a significant advance over existing methods in the semiconductor industry.
Semiconductor fabrication is a complex technological process and involves hundreds of steps. The fabrication process of a single chip can take hours or days, depending on the complexity of the process [1]. In recent years, the equipment is becoming highly automated, and its operational speed is also increasing to improve production yield. However, as equipment is more complicated, it becomes difficult to accurately detect potential faults. Fault detection and classification (FDC) systems are often used in semiconductor manufacturing to detect faults. An FDC identifies the effects of a potential fault on the observed variables during the fabrication process. After the FDC, various traditional methods such as thresholding, control charts, and principal component analysis can flag abnormal processed wafers or indicate warning of potential faults in the process.

9.1. Successful Implementations

Estimates suggest that nearly 42 billion semiconductors will be used per month for the 12 months ending in July 2022, a growth of 9% year over year. Despite the increase in the number of foundries and investment in expansion and infrastructure, the yield is still below par due to accidents happening at various stages and leading to faults. There are various inspections at various stations in production to uncover these faults. Following inspection, devices suspected of being faulty are subjected to expensive and laborious FA techniques to discover the possible defect models to replace the mask/photomask. A cost-effective method to identify defective devices is needed to sift through the suspect devices without incurring heavy expenses or resources [2]. To this end, given the data available from numerous inspections and the defects available, machine learning is proposed. But as the technology scaled, the architecture churned, the defects and data patterns became unknown. New machine-learning techniques that could adapt to the ever-changing data better than the former methods were needed. Seeing the past literature and how unsuitable some methods were at this stage prompted a focus on something novel, unseen, and challenging. One such application is semiconductor device-level defect detection, wherein the mask designs, device architectures, and layout patterns change with every new technology node. Thus, any artificially learned setup may degrade with just a slight change, and the data itself takes a toll thanks to the churns [3]. To mitigate this challenge, the focus here is on quantile-based online learning, wherein at any point in time, the technique can be reinitialized with the newest data pattern and commence quickly learning on newly arriving samples. The paper shows how the proposed approach better captures the underlying data distribution remotely similar to the true distribution, reflecting the data insight to detect defective instances along the decision surface. Complementary to this advancement in tool development, the proposal of domain adaptation strategies boosts a base classifier to capture distribution-mismatched inspections better. The proposed methods were applied to semiconductor device-level defects undertaken using real-world data from an anonymous fab.

9.2. Lessons Learned from Failures

The fabricated products not only need to be tested at the functional level, but also need to be analyzed after being failed, especially in advanced technology nodes [3]. In these nodes, with advanced materials and structures, the fault distributions and mechanisms are significantly different from older technology nodes. In addition, the defects often create more diverse failure patterns, leading to more uncertain and intricate instances. As a result, detecting defects from the product images becomes more elusive and complex .

$$L(t) = L_0 + \sum_{i=1}^{t} \Delta L_i$$

Eqn 2 : Cumulative Learning Model (Improvement over Time)

Where:

- $L(t)$: total lessons learned at time t
- L_0: baseline knowledge
- ΔL_i: new lessons from each failure analysis

The machine learning (ML)-guided failure analysis for semiconductor products has been widely studied due to its low-cost and high-throughput advantages. In a conventional way, the models are generally trained in the offline batch mode. When being deployed, the trained models are used to analyze the new samples. However, in the high-volume production process, the occurrence of new-types of failures or the changes in the underlying data distribution may happen. It is expensive and time-consuming to retrain the models with a batch of new samples. As an alternative, online learning (OL) allows classification models to incrementally learn from streaming samples in a sequential manner and update the existing model on-the-fly. Therefore, OL methods can timely adapt the classifiers to the new types of samples without re-training from scratch. Quantile regression is a promising OL method which is particularly effective for the detection of outliers or anomalous samples.
However, there is limited research on the OL methods for semiconductor FA. This paper focuses on quantile online learning for semiconductor FA. The quantile regression, with the corresponding online updating mechanism, is first reformulated from a non-adversarial context to an adversarial context, where the significance of monitoring the absolute value of cumulative loss is discussed. This quantile online learning method is then applied to semiconductor device-level defects and the quantitative evaluation for its performance is presented. It has been shown that the proposed method is able to perform better than the existing benchmarks.

10. Challenges and Limitations

The move towards AI-native semiconductor manufacturing is designed to enable accurate and timely fault detection across the lifecycle of cutting-edge devices, that is notably at the 3 nm node and beyond, with deep learning trained directly on control, measurement and inspection data. However, as the media, techniques and tools continue to transform, certain challenges must be overcome for AI-driven methods to deliver on their potential. As device designs and manufacturing technology evolve to handle increasing transistor count and density, the associated performance, yield and reliability challenges will also change accordingly. A robust, comprehensive, and standardised fabrication knowledge acquisition and data synthesis framework is critical for the data-driven AI method. In addition, the sheer variety, volume and velocity of the fabricated chip operation measurements necessitate a thorough hardware-software co-design modification and optimisation of both AI hardware and software for improved efficiencies and performances [2]. Meaningful structural representation learning from the heterogeneous data relies on the meticulous crafting of the feature engineering and extraction module, which is in turn predicated upon a deep understanding of the choice of the types, forms and levels of the features. To avoid the complexities associated with bespoke feature engineering, decoding by learning from raw data, upon which the feature extraction layer attempts to establish the low-dimensional representation of the data manifold, is gaining traction [3]. Trainable AI models are ripe for efficient incremental and online learning in a on-the-fly fashion, which can be further improved with Active learning and AWS System. Enabling effective decision making relies on the formulation of meaningful evaluation metrics and the construction of a transparent rationale to effectively communicate the prospect of an event and confidence in the recommended actions. In addition, associated performance benchmarks and testing frameworks, tailored explainable-AI measures and evaluation protocols, and active and cooperative learning paradigms must be rigorously crafted. Note that there are additional challenges surrounding interpretability, portability, ethical obstacles, and commercialisation that are beyond the scope of AI methods.

10.1. Data Quality Issues

Fault detection in various fabrication stages of semiconductor devices is a significant engineering issue [1]. Recently, semiconductor fabrication companies started applying AI-based fault detection systems. Transfer learning-based methods have also been developed to allow machinery vendors to transfer their methodology to IDMs with minimal effort. With this trend, the number of candidate methods has been growing rapidly. Most of the methods have been validated with well-designed benchmark datasets. However, in real-world applications, AI models need to deal with data quality issues. First, the number of defective products is much less than that of normal samples. Therefore, it is usually hard to collect sufficiently diverse sample data that cover the full variances between various fault patterns. Additionally, the nominal data would suffer from noisy signals originated from varying conditions under the fabrication processes or malfunctioning of sensors. AI models trained with such data would result in either false positives against a less defective sampling or prototype failures on unseen data due to overfitting.
The aim of this section is to point out the possible data quality issues in semiconductor fabrication involved in AI-based fault detection. It discusses the issues and how they could prevent AI systems from being successfully used in factories. The urgent need to develop data quality assessment methods that are easier and faster to be deployed in factories is emphasized. In-house data set royalty issues should also be addressed as well. Collaborative tests with benchmarks are expected to be encouraging tools for robustly sharing asset models in practical AD systems. Finally, they urge the community to pay more attention to the real world after the model-driven subsystem design.

In 2021, semiconductor fabrication is in the stage of Industry 4.0. Various machines are developed to automate lot transfers, measurement, and disposition decisions. Such machineries inevitably result in data oppressively huge and complex for human-centered inspection and monitoring. Gathering data from various fabrication processes is illustrative of learning such asset knowledge. AI-based approaches to fault detection have been thriving due to the overwhelming availability of event logs and the great advance of deep models.

10.2. Model Interpretability

Despite the blackbox nature of many AI models, various methods aim to ensure the trustworthiness of the model through its interpretability. There are two common forms of model interpretability, namely model-specific and model-agnostic methods. Model-Specific Methods: Saliency maps are a kind of computer vision model-specific method with high popularity and increasing trustworthiness. Saliency maps indicate whether a feature is, for a given input sample, influential for the output of the model and it highlights important regions of interest. The saliency map S of a neural network model f can be computed given an image input x. The S equals to the gradient of model prediction score p ($S = \partial p/\partial x$). Then, the score is back-propagated to the input image x, with regions indicated by more positive gradients being crucial for model decisions. Saliency maps can provide visual insights on learned patterns of the model. Beyond the saliency map generation methods, there are model-specific methods focusing on recovering the learned functioning of a model. For example, pattern-Diffusion is a diagnostic interpretation technique that decouples the action of the classifier into interpretable steps. It analyzes classes based on a given input image and the classifier, says f, in four steps. Initially, the f gets a perturbed copy of the original input where pixels corresponding to active neurons in the first S layer are set to a constant background intensity. This permutes prediction score by amount δ (with δ estimating a μ–σ band of S outputs). Then, secondary image diffusions with varied parameters are generated and scores are computed such that f gets to activate during network diffusion opposing its conservative trade-off. Model-Agnostic Methods: In a case study where the internal mechanics of a fault detection model are decomposed to assess its explainability or the justification of its predictions, model-agnostic methods play a vital role. The model-agnostic framework SparkNose is presented as a fault detection platform by using an interpretable framework called Anchors. Anchors is a model-agnostic explanation approach that generates interpretable explanations based on user-defined examples and can trace individual model predictions back to the original features. The model was trained using a supervised learning algorithm, based on a long short-term memory recurrent neural network architecture, to detect faults and correctly classify their fault types [7].

11. Future Trends in AI and Semiconductor Fabrication

Due to recent discoveries and successes in AI technologies in various fields, including text summarization, object detection, disease prediction, recommendation systems, and others, there has been an increase in interest in learning and utilizing AI technologies in semiconductor manufacturing. This is due to fewer engineers being available for work after the pandemic compared to before and fewer engineers producing chips. Engineers often spend much of their time manually processing data generated by machines. This time-consuming process could benefit from applying AI technologies. Everyone in the manufacturing field is hunting for a low-HR, no-code, light-weight solution that can help process and analyze massive amounts of data, helping engineers focus and devote themselves to more important tasks. This solution should have an easy learning curve, allowing engineers to easily define probabilistic tag string labels and evaluate results using visualization tools and dashboards

Eqn 3 : Moore's Law (Traditional Scaling Limit)

$$N(t) \propto 2^{t/T}$$

Where:

- $N(t)$: transistor count at time t
- T: time it takes for count to double (historically ~2 years)

Trend: AI helps **optimize beyond Moore's Law** using chiplet design, advanced packaging, and better yield prediction.

Some areas in semiconductor manufacturing include yield prediction, automated processing, PM and GM automation, intra-database correlation, and cross-database correlation. Some low-HR applications that can correlate fail and exists data in W/L and die areas, as well as other fix ideas and criteria, could be created. There are already a few "No-Code" based processes/tools to help data collection, analysis, and debugging, but there is still room for new ideas. There are many MVP problems to be solved and initially demonstrated for the industry. Fake and unknown knowledge remains in massive amounts of data, and a strategic plan/pipeline is needed in semiconductor architecture design scaling generations from NDA/competition to industry. There is a need for a widely used tool/concept of SoC high-performance ML "scalability" for IC design while enabling better data management in EDA tools and SIL/FPGA domains. In addition, there is a generational upgrade for reliability simulation of variability from process variation to RPT, interconnect, and reliability (as key design metrics) including heterogeneous 1D/2D materials.

11.1. Advancements in AI Technology

The rapid advancement of AI technologies has been propelled by the exponential growth of computational power over the last decade. The application of AI is presently and widely expanded across a multitude of industries, creating both opportunities and challenges. In particular, the global semiconductor industry has been tackling a critical roadblock of indeterminate time-to-solution (TTS) simulation of process variability at a rising level of integration. The state-of-the-art Monte Carlo simulation framework with a finite number of particles is incapable of producing robust reliability analysis on time due to its unscalable

TTS. Therefore, the re-acceleration of TTS-imposing simulations has become a top priority in the semiconductor industry for faster design and time-to-market. Acquired through self-learning, AI has emerged as the only existing solution to address this burning problem. Trained on existing TTS-imposing simulation outputs, AI is dubbed to directly predict reliability scores with negligible TTS, revolutionizing design and typically reducing TTS by orders of magnitude [8]. Incorporated into a design, the AI prediction renders "silicon-like" reliability results, making it impossible to identify fault sites and repair it. The huge existing knowledge bases acquired during conventional simulations enormously enhanced process capability but they are inaccessible for advanced physical process understanding due to complexity. The discovery of design and technology physics from large knowledge bases remains a challenge, while it typically requires an exquisite understanding of physical mechanisms with extraordinary talents and experiences accumulated for decades. Nevertheless, thousands of advanced designs and processes have been senselessly and compositionally fabricated in semiconductor industries, and even very recently, it became possible to capture the cycling convergence via transfer matrix analysis. AI is actively mined to recover physical knowledge bases acquired during process simulations. Diverse and broad applications of AI methods are being investigated, including the implementation of generative adversarial networks for knowledge discovery from massive data.

11.2. Potential for Automation

It has been commented that if cars had been developed as fast and with as much expenditure as the integrated circuits, a car would have driven faster than the speed of light, so sophisticated, with built-in features like automatic gear change, road detection, etc. Considering the present science and technology innovations in smart systems, artificial intelligence (AI) algorithms can be integrated for self-inspection and diagnosis purposes in semiconductor manufacturing and fab and can be further developed in the future. Semiconductor manufacturing is truly one of the most important components of high-tech industries. It is required that entire manufacturing fab should not have defects of any kind in the fabricated semiconductor chips, which makes inspection and diagnosis of the entire chip manufacturing process inevitable. Automatic inspection is also needed as it is not feasible to be done manually. Initially, visual inspection of fabricated semiconductor wafers was performed by very trained experts. However, visual inspection is infeasible for the present ultra-large-scale integration (ULSI) chips which are so densely packed with so small features. Moreover, due to increased complexity in design and layout of the fabricated chips, it is probable that a wafer will be defective and it is needed to detect diverse pattern defects out of millions of failures in the short inspection time and dead-scan time [6].

Thus, digital inspection has been developed in the last few decades and has been very effective in test pattern generation procedure in customer selection tests. Still, due to fabrication process variation, chalk (open and short) and positional (incomplete, wrong, flip, and extra) defects are critical, yet these defects with similar appearance cannot be guaranteed to be detected by conventional algorithms just after testing the scan chains [9]. Moreover, false alarms are frequent due to CAD data errors and thus yield degradation. Hence re-test standard interpretation of failures after comprehensive scan-test is very much needed in ensuring detection of all faults, failure localization, and also minimization of false alarms and timely analysis. With increasing diversity of manufacturing process and large number of design rules, especially lay-out rules, passing a new chip design is becoming increasingly difficult. By intelligent CAD-aided design verification, undesirable failures will be avoided when chips are fabricated on the present multi-layered process.

12. Conclusion

In conclusion, challenges in the design-for-high-precision context, model and data challenges, hard technology limitations concerning the knowledge of deep neural networks and AI technology must be addressed before precautions to ensure high yield in training the global SEMI-AI-based architecture. These challenges will apply, generalize, or evolve independently or collectively across diverse technical platforms or other similar supervision architectures. AI technology-to-design, -data, -model, and -factual requirements for addressing high precision and high reliability in semiconductor product design, and addressing high yield in the global SEMI-AI-based training situation are summarized [2]. Also, corresponding short-term, mid-term, and long-term prospect works are proposed to provide more convenience in designing with AI technology, training the SEMI-AI-based product design architecture, and paving the way for encoding the challenges in semiconductor design, data, and model, with well-controlled potential continual paradigm thinking for integrated production [3]. The key challenges in designing AI technology for high precision in chip design context, and the challenges in high reliability in product testing. The knowledge engineering challenges concerning design-for-high-precision chip design with AI technology are summarized. The challenges in building a trustworthy, reliable, and responsible AI-design chip design supervision architecture are discussed. Technology limitations mainly concerning hard-technology limitations and soft-technology limitations are reviewed. Further challenges concerning data and model technicalities are discussed. Data challenges encountered in the chip design domain and relevant nation's security issues are summarized. The large-scale pretrained AutoDL models' rough and complicated inner structures and lack knowledge behind the model structure, parameters, and AI ability that can be effectively accessed, interpretable, and compacted are discussed. A rationale-based generative dual-control strategy to ensure the design precision of any deep learning model is proposed for knowledge-to-design care, and a fuzzy knowledge-based, multi-granularity, and temporal-degree approach to identifying the un-interpretable situation of any AI models is proposed for model explainable care.

References:

[1] Kommaragiri, V. B., Preethish Nanan, B., Annapareddy, V. N., Gadi, A. L., & Kalisetty, S. (2022). Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing. Venkata Narasareddy and Gadi, Anil Lokesh and Kalisetty, Srinivas.

[2] Pamisetty, V., Dodda, A., Singireddy, J., & Challa, K. (2022). Optimizing Digital Finance and Regulatory Systems Through Intelligent Automation, Secure Data Architectures, and Advanced Analytical Technologies. Jeevani and Challa, Kishore, Optimizing Digital Finance and Regulatory Systems Through Intelligent Automation, Secure Data Architectures, and Advanced Analytical Technologies (December 10, 2022).

[3] Paleti, S. (2022). The Role of Artificial Intelligence in Strengthening Risk Compliance and Driving Financial Innovation in Banking. International Journal of Science and Research (IJSR), 11(12), 1424–1440. https://doi.org/10.21275/sr22123165037

[4] Komaragiri, V. B. (2022). Expanding Telecom Network Range using Intelligent Routing and Cloud-Enabled Infrastructure. International Journal of Scientific Research and Modern Technology, 120–137. https://doi.org/10.38124/ijsrmt.v1i12.490

[5] Pamisetty, A., Sriram, H. K., Malempati, M., Challa, S. R., & Mashetty, S. (2022). AI-Driven Optimization of Intelligent Supply Chains and Payment Systems: Enhancing Security, Tax Compliance, and Audit Efficiency in Financial Operations. Tax Compliance, and Audit Efficiency in Financial Operations (December 15, 2022).

[6] Mashetty, S. (2022). Innovations In Mortgage-Backed Security Analytics: A Patent-Based Technology Review. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3826

[7] Kurdish Studies. (n.d.). Green Publication. https://doi.org/10.53555/ks.v10i2.3785

[8] Motamary, S. (2022). Enabling Zero-Touch Operations in Telecom: The Convergence of Agentic AI and Advanced DevOps for OSS/BSS Ecosystems. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3833

[9] Kannan, S. (2022). AI-Powered Agricultural Equipment: Enhancing Precision Farming Through Big Data and Cloud Computing. Available at SSRN 5244931.

[10] Suura, S. R. (2022). Advancing Reproductive and Organ Health Management through cell-free DNA Testing and Machine Learning. International Journal of Scientific Research and Modern Technology, 43–58. https://doi.org/10.38124/ijsrmt.v1i12.454

[11] Nuka, S. T., Annapareddy, V. N., Koppolu, H. K. R., & Kannan, S. (2021). Advancements in Smart Medical and Industrial Devices: Enhancing Efficiency and Connectivity with High-Speed Telecom Networks. Open Journal of Medical Sciences, 1(1), 55-72.

[12] Meda, R. (2022). Integrating IoT and Big Data Analytics for Smart Paint Manufacturing Facilities. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3842

[13] Annapareddy, V. N., Preethish Nanan, B., Kommaragiri, V. B., Gadi, A. L., & Kalisetty, S. (2022). Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing. Venkata Bhardwaj and Gadi, Anil Lokesh and Kalisetty, Srinivas, Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing (December 15, 2022).

[14] Phanish Lakkarasu. (2022). AI-Driven Data Engineering: Automating Data Quality, Lineage, And Transformation In Cloud-Scale Platforms. Migration Letters, 19(S8), 2046–2068. Retrieved from https://migrationletters.com/index.php/ml/article/view/11875

[15] Kaulwar, P. K. (2022). Securing The Neural Ledger: Deep Learning Approaches For Fraud Detection And Data Integrity In Tax Advisory Systems. Migration Letters, 19, 1987-2008.

[16] Malempati, M. (2022). Transforming Payment Ecosystems Through The Synergy Of Artificial Intelligence, Big Data Technologies, And Predictive Financial Modeling. Big Data Technologies, And Predictive Financial Modeling (November 07, 2022).

[17] Recharla, M., & Chitta, S. (2022). Cloud-Based Data Integration and Machine Learning Applications in Biopharmaceutical Supply Chain Optimization.

[18] Lahari Pandiri. (2022). Advanced Umbrella Insurance Risk Aggregation Using Machine Learning. Migration Letters, 19(S8), 2069–2083. Retrieved from https://migrationletters.com/index.php/ml/article/view/11881

[19] Paleti, S., Burugulla, J. K. R., Pandiri, L., Pamisetty, V., & Challa, K. (2022). Optimizing Digital Payment Ecosystems: Ai-Enabled Risk Management, Regulatory Compliance, And Innovation In Financial Services. Regulatory Compliance, And Innovation In Financial Services (June 15, 2022).

[20] Singireddy, J. (2022). Leveraging Artificial Intelligence and Machine Learning for Enhancing Automated Financial Advisory Systems: A Study on AIDriven Personalized Financial Planning and Credit Monitoring. Mathematical Statistician and Engineering Applications, 71 (4), 16711–16728.

[21] Paleti, S., Singireddy, J., Dodda, A., Burugulla, J. K. R., & Challa, K. (2021). Innovative Financial Technologies: Strengthening Compliance, Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures. Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures (December 27, 2021).

[22] Sriram, H. K. (2022). Integrating generative AI into financial reporting systems for automated insights and decision support. Available at SSRN 5232395.

[23] Koppolu, H. K. R. (2021). Leveraging 5G Services for Next-Generation Telecom and Media Innovation. International Journal of Scientific Research and Modern Technology, 89–106. https://doi.org/10.38124/ijsrmt.v1i12.472

[24] End-to-End Traceability and Defect Prediction in Automotive Production Using Blockchain and Machine Learning. (2022). International Journal of Engineering and Computer Science, 11(12), 25711-25732. https://doi.org/10.18535/ijecs.v11i12.4746

[25] Chaitran Chakilam. (2022). AI-Driven Insights In Disease Prediction And Prevention: The Role Of Cloud Computing In Scalable Healthcare Delivery. Migration Letters, 19(S8), 2105–2123. Retrieved from https://migrationletters.com/index.php/ml/article/view/11883
[26] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.
[27] Avinash Pamisetty. (2021). A comparative study of cloud platforms for scalable infrastructure in food distribution supply chains. Journal of International Crisis and Risk Communication Research , 68–86. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2980
[28] Gadi, A. L., Kannan, S., Nanan, B. P., Komaragiri, V. B., & Singireddy, S. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization. Universal Journal of Finance and Economics, 1(1), 87-100.
[29] Dodda, A. (2022). The Role of Generative AI in Enhancing Customer Experience and Risk Management in Credit Card Services. International Journal of Scientific Research and Modern Technology, 138–154. https://doi.org/10.38124/ijsrmt.v1i12.491
[30] Gadi, A. L. (2022). Connected Financial Services in the Automotive Industry: AI-Powered Risk Assessment and Fraud Prevention. Journal of International Crisis and Risk Communication Research, 11-28.
[31] Pamisetty, A. (2022). A Comparative Study of AWS, Azure, and GCP for Scalable Big Data Solutions in Wholesale Product Distribution. International Journal of Scientific Research and Modern Technology, 71–88. https://doi.org/10.38124/ijsrmt.v1i12.466
[32] Adusupalli, B. (2021). Multi-Agent Advisory Networks: Redefining Insurance Consulting with Collaborative Agentic AI Systems. Journal of International Crisis and Risk Communication Research, 45-67.
[33] Dwaraka Nath Kummari. (2022). Iot-Enabled Additive Manufacturing: Improving Prototyping Speed And Customization In The Automotive Sector . Migration Letters, 19(S8), 2084–2104. Retrieved from https://migrationletters.com/index.php/ml/article/view/11882
[34] Data-Driven Strategies for Optimizing Customer Journeys Across Telecom and Healthcare Industries. (2021). International Journal of Engineering and Computer Science, 10(12), 25552-25571. https://doi.org/10.18535/ijecs.v10i12.4662
[35] Adusupalli, B., Singireddy, S., Sriram, H. K., Kaulwar, P. K., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks. Universal Journal of Finance and Economics, 1(1), 101-122.
[36] AI-Based Financial Advisory Systems: Revolutionizing Personalized Investment Strategies. (2021). International Journal of Engineering and Computer Science, 10(12). https://doi.org/10.18535/ijecs.v10i12.4655
[37] Karthik Chava. (2022). Harnessing Artificial Intelligence and Big Data for Transformative Healthcare Delivery. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 502–520. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11583
[38] Challa, K. (2022). The Future of Cashless Economies Through Big Data Analytics in Payment Systems. International Journal of Scientific Research and Modern Technology, 60–70. https://doi.org/10.38124/ijsrmt.v1i12.467
[39] Pamisetty, V., Pandiri, L., Annapareddy, V. N., & Sriram, H. K. (2022). Leveraging AI, Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management. Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management (June 15, 2022).
[40] Innovations in Spinal Muscular Atrophy: From Gene Therapy to Disease-Modifying Treatments. (2021). International Journal of Engineering and Computer Science, 10(12), 25531-25551. https://doi.org/10.18535/ijecs.v10i12.4659
[41] Kaulwar, P. K. (2022). Data-Engineered Intelligence: An AI-Driven Framework for Scalable and Compliant Tax Consulting Ecosystems. Kurdish Studies, 10 (2), 774–788.
[42] Operationalizing Intelligence: A Unified Approach to MLOps and Scalable AI Workflows in Hybrid Cloud Environments. (2022). International Journal of Engineering and Computer Science, 11(12), 25691-25710. https://doi.org/10.18535/ijecs.v11i12.4743
[43] Nandan, B. P., & Chitta, S. (2022). Advanced Optical Proximity Correction (OPC) Techniques in Computational Lithography: Addressing the Challenges of Pattern Fidelity and Edge Placement Error. Global Journal of Medical Case Reports, 2(1), 58-75.
[44] Raviteja Meda. (2021). Machine Learning-Based Color Recommendation Engines for Enhanced Customer Personalization. Journal of International Crisis and Risk Communication Research , 124–140. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3018
[45] Rao Suura, S. (2021). Personalized Health Care Decisions Powered By Big Data And Generative Artificial Intelligence In Genomic Diagnostics. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v7i3.3558
[46] Implementing Infrastructure-as-Code for Telecom Networks: Challenges and Best Practices for Scalable Service Orchestration. (2021). International Journal of Engineering and Computer Science, 10(12), 25631-25650. https://doi.org/10.18535/ijecs.v10i12.4671
[47] Vamsee Pamisetty, Lahari Pandiri, Sneha Singireddy, Venkata Narasareddy Annapareddy, Harish Kumar Sriram. (2022). Leveraging AI, Machine Learning, And Big Data For Enhancing Tax Compliance, Fraud Detection, And Predictive Analytics In Government Financial Management. Migration Letters, 19(S5), 1770–1784. Retrieved from https://migrationletters.com/index.php/ml/article/view/11808

[48] Someshwar Mashetty. (2020). Affordable Housing Through Smart Mortgage Financing: Technology, Analytics, And Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 99–110. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11581

[49] Srinivasa Rao Challa,. (2022). Cloud-Powered Financial Intelligence: Integrating AI and Big Data for Smarter Wealth Management Solutions. Mathematical Statistician and Engineering Applications, 71(4), 16842–16862. Retrieved from https://philstat.org/index.php/MSEA/article/view/2977

[50] Paleti, S. (2022). Fusion Bank: Integrating AI-Driven Financial Innovations with Risk-Aware Data Engineering in Modern Banking. Mathematical Statistician and Engineering Applications, 71(4), 16785-16800.

[51] Pamisetty, V. (2022). Transforming Fiscal Impact Analysis with AI, Big Data, and Cloud Computing: A Framework for Modern Public Sector Finance. Big Data, and Cloud Computing: A Framework for Modern Public Sector Finance (November 30, 2022).

[52] Kommaragiri, V. B., Gadi, A. L., Kannan, S., & Preethish Nanan, B. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization.

[53] Annapareddy, V. N. (2022). Integrating AI, Machine Learning, and Cloud Computing to Drive Innovation in Renewable Energy Systems and Education Technology Solutions. Available at SSRN 5240116.

[54] Transforming Renewable Energy and Educational Technologies Through AI, Machine Learning, Big Data Analytics, and Cloud-Based IT Integrations. (2021). International Journal of Engineering and Computer Science, 10(12), 25572-25585. https://doi.org/10.18535/ijecs.v10i12.4665

[55] Venkata Bhardwaj Komaragiri. (2021). Machine Learning Models for Predictive Maintenance and Performance Optimization in Telecom Infrastructure. Journal of International Crisis and Risk Communication Research , 141–167. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3019

[56] Paleti, S. (2021). Cognitive Core Banking: A Data-Engineered, AI-Infused Architecture for Proactive Risk Compliance Management. AI-Infused Architecture for Proactive Risk Compliance Management (December 21, 2021).

[57] Harish Kumar Sriram. (2022). AI-Driven Optimization of Intelligent Supply Chains and Payment Systems: Enhancing Security, Tax Compliance, and Audit Efficiency in Financial Operations. Mathematical Statistician and Engineering Applications, 71(4), 16729–16748. Retrieved from https://philstat.org/index.php/MSEA/article/view/2966

[58] Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29-41.

[59] Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). International Journal of Engineering and Computer Science, 9(12), 25289-25303. https://doi.org/10.18535/ijecs.v9i12.4587

[60] Pallav Kumar Kaulwar. (2021). From Code to Counsel: Deep Learning and Data Engineering Synergy for Intelligent Tax Strategy Generation. Journal of International Crisis and Risk Communication Research , 1–20. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2967

[61] Pandiri, L., & Chitta, S. (2022). Leveraging AI and Big Data for Real-Time Risk Profiling and Claims Processing: A Case Study on Usage-Based Auto Insurance. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3760

[62] Kummari, D. N. (2022). AI-Driven Predictive Maintenance for Industrial Robots in Automotive Manufacturing: A Case Study. International Journal of Scientific Research and Modern Technology, 107–119. https://doi.org/10.38124/ijsrmt.v1i12.489

[63] Gadi, A. L. (2022). Cloud-Native Data Governance for Next-Generation Automotive Manufacturing: Securing, Managing, and Optimizing Big Data in AI-Driven Production Systems. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3758

[64] Dodda, A. (2022). Secure and Ethical Deployment of AI in Digital Payments: A Framework for the Future of Fintech. Kurdish Studies. https://doi.org/10.53555/ks.v10i2.3834

[65] Gadi, A. L. (2021). The Future of Automotive Mobility: Integrating Cloud-Based Connected Services for Sustainable and Autonomous Transportation. International Journal on Recent and Innovation Trends in Computing and Communication, 9(12), 179-187.

[66] Dodda, A. (2022). Strategic Financial Intelligence: Using Machine Learning to Inform Partnership Driven Growth in Global Payment Networks. International Journal of Scientific Research and Modern Technology, 1(12), 10-25.

[67] Just-in-Time Inventory Management Using Reinforcement Learning in Automotive Supply Chains. (2021). International Journal of Engineering and Computer Science, 10(12), 25586-25605. https://doi.org/10.18535/ijecs.v10i12.4666

[68] Srinivasa Rao Challa. (2021). From Data to Decisions: Leveraging Machine Learning and Cloud Computing in Modern Wealth Management. Journal of International Crisis and Risk Communication Research , 102–123. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3017

[69] Kommaragiri, V. B. (2021). Enhancing Telecom Security Through Big Data Analytics and Cloud-Based Threat Intelligence. Available at SSRN 5240140.