ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

Data Analytics-Driven Approaches to Yield Prediction in Semiconductor Manufacturing

Botlagunta Preethish Nandan

Senior Lead Business Intelligence, ORCID ID : 0009-0008-3617-8149

Abstract: In the semiconductor manufacturing industry, the yield prediction of a product is vitally essential. Its accuracy directly affects product cost and customer satisfaction. One commonly known yield prediction and management technique is statistical machine learning models due to their high generalization capacity. With the support of rapidly increasing data volume, driven by advanced wafer-level processes, an increasing number of Deep Learning architectures have recently been adopted for yield prediction in semiconductor manufacturing. With rapid advancements in wafer-level processes, it becomes feasible to collect defect density data through advanced sensors from testing equipment. However, constructing Deep Learning models from scratch requires a lot of expertise in both semiconductor manufacturing and machine learning, which is not easy to obtain in modern semiconductor industries. To automatically assist semiconductor manufacturing engineers in building accurate Deep Learning models in yield prediction, this study first designs and builds a fully-automated Deep Learning yield prediction framework. This framework can assist engineers in developing Deep Learning models in a timely manner without time-consuming data preprocessing, feature engineering, or architecture searching. It consists of yield simulation, data preparation, candidate models building, and ensembling formation. Besides yield prediction, another significant concern is the explanation of Deep Learning model outputs. On the one hand, many model-agnostic explanation algorithms have been successfully adopted in many fields, providing valuable information for improving model quality and transparency. The explanation logic is intuitive; e.g., it calculates the contribution of a feature to the output of the prediction, which helps in finding out pixel areas or categorical reasons to focus. On the other hand, due to the complexity of semiconductor manufacturing processes, existing interpretable models lack effective inductive bias for proper yield prediction in semiconductor manufacturing, and turned out to be unable to simulate disentangled understanding rules. Therefore, current model-agnostic explanation methods for explaining Deep Learning model outputs either fail to reason rules about semiconductors, due to insufficient representation ability, or contain prohibitive time complexity for searching the activation of nodes or features. As a result, existing approaches struggle to simultaneously achieve both high prediction accuracy and high-quality explanations.

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.

I. INTRODUCTION

Semiconductor manufacturing constitutes one of the most intricate, highly automated, and advanced manufacturing sectors. The immense difficulty in designing and manufacturing IC chips has resulted in the development of a highly specialized and segmented supply chain in the semiconductor industry. In a typical semiconductor supply chain, reticles are manufactured in a photomask house and are shipped to a wafer fabrication facility (fab) to create wafers with IC chips. Subsequently, packaging houses process the wafers to produce packaged IC chips before they are shipped to customer assembly houses for testing and then delivered to customers. Yield enhancement significantly influences process reliability, cost efficiency, product robustness, and subsequently market competitiveness. Hence, enhancing yield has become a crucial lever to cut costs and increase net profit. In advanced logic wafer fabrication facilities, a mere 1% increase in yield would lead to an additional estimated net profit of $150 million, while a 1% drop in yield would give rise to a remarkable $250 million loss.

To facilitate yield enhancement, data analytics-driven (DA-driven) approaches, namely machine learning (ML) or deep learning (DL) methods, have been increasingly leveraged. Such DA-driven yield prediction approaches are applicable to a wide range of tasks in different stages of semiconductor manufacturing and enable the comprehensive monitoring of such processes. Specifically, the DA-driven yield prediction approaches are applicable to front-end-of-line (FEOL) yield prediction in wafer fabrication, back-end-of-line (BEOL) yield prediction, defect yield prediction in IC packaging, wafer-

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

test yield prediction in wafer testing, die-level functional yield prediction in package testing, and test time prediction after the circuitry design is frozen. The central role of DA-driven approaches to yield prediction in semiconductor manufacturing can be essentially attributed to the types of data ubiquitous in semiconductor smart manufacturing (SSM).

II. OVERVIEW OF SEMICONDUCTOR MANUFACTURING

The manufacture of semiconductor devices (ICs) constitutes an important and challenging domain of production engineering, which is also referred to as wafer fabrication or wafer processing. The fabrication of ICs starts with an electronic grade silicon ingot, which is sliced to wafers, polished, and oxidized. Then, thin films are deposited, pattered, and etched; details of these steps are recorded in extensive engineering knowledge. The cause for 99.9% of processed wafers is 'defect-free,' and each IC dies or chips is formed from exposure through mask patterns, including interconnections to terminals [1].

Most devices are intended for consumer products compatible with high-volume fabrication. Wafer fabrication facilities (fabs) cost billions and take years to be built. Yield has emerged as a pivotal lever to curb costs and increase profits. For advanced logic fab, a 1% increase in yield would result in an estimated additional net profit of several hundreds of millions. However, enhancing the yield and passing the pre-sail wafer final test are extremely difficult. The IC product is composed of millions of transistors; first, defects must remain below a critical size and volume for process nodes below 65 nm. The process is inherently interrelated; a discrepancy in a critical dimension immediately affects electrostatics operation with delay. Random non-detection process anomalies in aberrations or drift could be a killer that wipes out revenue entirely. Device development time can take years.

Yield enhancement analyses to measure mechanisms and models comprise statistical and data-analytics-based approaches. There is increasing employment of machine learning to augment yield enhancement strategies, such as analyzing critical process steps by widely applied feature selection to select process step-contributing yield (Y) detrimentally, assisting in troubleshooting and process optimization by data mining in search of processing parameters discoverable from massive process logs of similar rectangles, detecting the potential cause referring to the process or test steps with high likelihood malfunction for anomalies owing to instantaneous or periodic causes by clustering algorithms, automatic defect classification by pattern-recognition algorithms and supervised ML using server observable aberrantly measured metrics for test tool-to-test tool malfunctions, spotting lot/module defects building on spatial-temporal information.

III. IMPORTANCE OF YIELD PREDICTION

The semiconductor manufacturing industry is one of the major industries that contribute to the worldwide economy [1]. Return on investment (ROI) improvement has become an urgent priority for this industry, especially the yield improvement. Yield is defined as the ratio of good (i.e., defect-free) products to the total number of produced products. The yield rate in wafer fabrication or semiconductor manufacturing is usually very low due to defect introduction. Hence, waivers need to undergo an extensive set of inspection, test, and/or repair operations to identify and rectify defects or faults. Those chips or dice that pass all inspection and/or test operations are sold. On the other hand, those chips or dice that fail in any operation are classified as non-functional products. Wafers that fail to pass inspection/repair operations are scrapped. If the yield rates in terms of good or functional products can be predicted in advance before the inspection, testing, and repair operations, the operations can be executed in a more cost-effective manner.

In the context of semiconductor manufacturing, such a prediction task, also referred to as yield prediction in this paper, is defined as predicting the yield state of the wafer of interest or target wafer based on the past wafer-level inspection, testing, and repair data. It is required to perform such a prediction task on each newly and/or incoming wafer after all the related yield attributes have been set. Thus, the visualization-based exploratory data analysis needs to be executed in an automatic or semi-automatic manner. The return on investment of production activities in the semiconductor industry relies heavily on high semiconductor yields. Thus, yield improvement becomes an urgent priority in semiconductor manufacturing industry. The semiconductor fabrication process involves several hundreds of different fabrication processes leading to hundreds of defects. Each of these processes is performed by dedicated machines known as equipment. The yield outcome of the semiconductor manufacturing process is a binary outcome determined based on the measurement parameter values obtained from the measurement machines.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

IV. DATA ANALYTICS IN MANUFACTURING

Data analytics plays an important role in manufacturing where it can be applied to forecast/estimate product specification, equipment performance, resource utilization, yield outcome and other key business objectives. Semiconductor manufacturing is a highly sophisticated and capital-intensive business with numerous complex processes for circuit pattern transfers onto silicon wafers. In this manufacturing environment, the often unexpected yield loss due to the introduction of a multitude of process or equipment disturbances can have a significant impact on product delivery and thus the overall business performance. Therefore, automatic systems/methods to predict the yield outcome of a wafer processing job prior to the final inspection are thus highly desired by the manufacturers [2]. The rapid advancement of artificial intelligent technology enables the development of a data-driven yield prediction type of knowledge discovery system, which detects possible non-conformance wafers even before the final visual inspection is completed [3]. This proactive type of predictive analytics approach improves the efficiency of the final inspection, optimizes the inspection planning and reduces the wasted inspection resources.

The work to be reported in this paper covers two major aspects of the data analytics-driven process yield prediction in the semiconductor manufacturing, with a focus on the key data visualization and modeling techniques employed. The first aspect is the data visualization of the process data generated during the wafer processing machine operation and near-infrared chemical imaging inspection and the challenges in visualizing a high-dimensional dataset with a nonlinear relationship. The second aspect is the choice of data-driven classification techniques for the data-predictive yield modeling and compares their performance obtained on a diverse test-bed. In addition to the emphasis on the data technique aspects, the paper concludes with a brief note of future work directions in the topic.

V. TYPES OF DATA IN SEMICONDUCTOR MANUFACTURING

This section introduces a series of data types regarding semiconductor manufacturing, which aids in the following review of the corresponding data analytics-driven approaches. More specifically, the section reviews batch data with an eye towards advance wafer-level virtual metrology, a typical type of challenging task in semiconductor manufacturing.

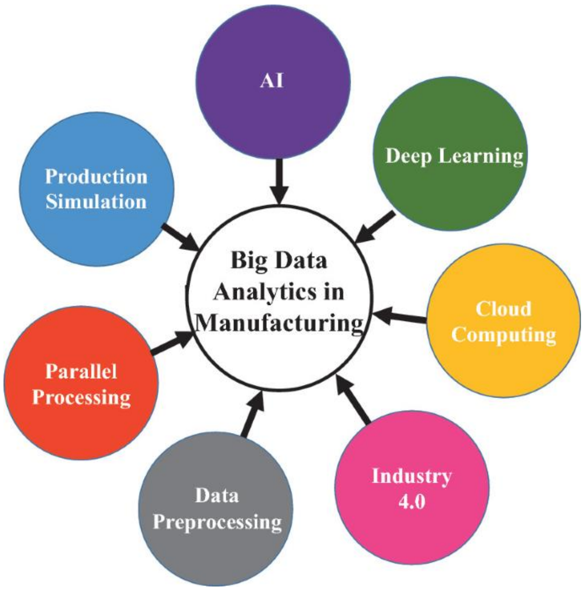

Fig 1: Big Data Analytics for Semiconductor Manufacturing

With the increasing deployment of data analytics technologies, the semiconductor industry is experiencing a data revolution. Genomic sequencing generates more than 1 exabyte of raw data every year. This is comparable to the size of all the worldwide written words in human history. In the semiconductor industry, the fast-growing data sources include fault diagnostic sensors, process scanners at various stages of the manufacturing flow, and probe test and final test at final assembly stages. Added up, these sensors generate terabytes of raw data every week. Such an abundance of sensor data creates an unprecedented opportunity to leverage predictive analytics to automate much of the data monitoring and postmortem process analysis involved in current operations. Data in the semiconductor industry may be classified into time-stamped and non-time-stamped data.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

Time-stamped data are generated by sensors to monitor the ongoing processing of raw wafers. Smart production test in the semiconductor industry monitors the product quality by testing measurements that typically involve complicated test setup and data analysis. A huge amount of test data may be generated and retained for long periods of time. With advanced process control being widely adopted for production quality assurance, there is a pressing need to develop data analytics-driven software systems to monitor the processing of raw wafers for automatic detection of process anomalies and clues for the suspicious cause. Non-time-stamped categorical data represent information in discrete categories. Fixed attribute data include continuous measurements such as external metrology results, categorical measurements such as identification information describing the batch processes on specific tools. Variable attribute data include regression outputs such as spatially formative pulses describing 2-D product structures or summed-up measurement variable sources describing total figure of merit.

Process data analytics-driven virtual metrology is a new data analytics-driven approach to improve manufacturing yield and efficiency. Using this approach, prediction models are built using real-time process monitoring data to replace physical metrology tools that may be slower, more expensive, or less accurate [4]. A standard prediction process is described, which is usually composed of data acquisition, data preprocessing, and model training/test/prediction. The conventional comparison method is implemented to predict final metrology values at final assembly. Long short-term memory (LSTM) and auto-regression commonly used in time series prediction are adapted and benchmarked to predict impending or intermediate metrology values. Illustrative examples using multilayer perceptron and LSTM-prediction models are provided based on a plasma etching data set described in.

5.1. Process Data

This chapter details the application of data analytics-driven yield prediction in semiconductor back-end assembly and test. The data analytics framework focuses on predictive modeling and fault detection. In the framework for predictive modeling, several data-driven yield prediction approaches, either using historical lot-level non-visual data, inherent visual data, or combined models of both non-visual and visual data, are introduced. In the framework for yield fault detection, innovative approaches for novelty detection, using wavelet transform and generative modeling, are introduced. The method can effectively detect new yield faults, even without prior known fault data. This study provides a better understanding of the development of data analytics-driven approaches to yield prediction. And also discusses future opportunities to build on established methods in non-semi-conductor applications, including big data, deep learning, forecasting, autoML, and natural language processing.

To maximize yield, product quality needs to be ensured at every stage of the process—from wafer manufacturing and circuit designing, through fab fabrication, to back-end packaging. Traditional yield prediction relies on prior product knowledge. With scaled-down rules of thumb, statistically significant models can be established to predict yield based on design for manufacturing criteria. As the demand for faster time to market and shorter design cycles increases, prior product knowledge cannot be easily obtained, and rapid yield prediction is needed [5]. As a result, predictive modeling approaches that solely rely on process data instead of prior product knowledge have led to intensive research interests in data analytics-residents in semiconductor manufacturing.

This chapter focuses on the application of data analytics-driven approaches to yield prediction in semiconductor back-end assembly and test. Though yield prediction in post-fab processes, including substrate probing, assembly, final test, and packing, has been studied in the past, these models remain relatively static, of limited generalizability, and computationally intensive. A rapidly updating and flexible data analytic framework that allows simulation-driven modeling is still needed. In this regard, there are opportunities to explore the abundant process data to develop data analytics-driven approaches for predicting and understanding yield-fault relations. Furthermore, technology scaling is paving the way for wafer-scale package-level assembly. Advanced three-dimensional integration (3DI), in which chips are assembled and interconnected in a package, is driving the miniaturization of devices to achieve higher levels of integration. However, it can also lead to yield loss, which may not be comprehensively modeled with prior knowledge.

5.2. Quality Data

Machine learning and data analytics have been widely utilized in yield prediction, diagnosis, and process optimization. Machine learning alone is an unlikely candidate for a suitable production solution in many problems. Therefore, data analytics-driven approaches combining mathematical models and data sciences methods would be the most suitable yield prediction solution. Yield prediction methods with two approaches of data analytics-driven are proposed, and yield models with various structures are built with multivariate soft sensor technologies, Gaussian process regression (GPR) and hybrid Metamodel-based design of experiments, simulation-based modeling.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

When combining the two data analytics techniques, the multivariate soft sensor technology and GPR-based time sequence prediction model, the detailed processing conditions and factors of 70- to 130-mm in diameter wafers predict efficiently. In this research, a hybrid simulation-based modeling with a technique based on the Morris screening for Exponential function, a Polynomial function, and a Gaussian function as surrogate models are combined to meet faster yield prediction with accuracy.

The advancement of artificial intelligence (AI) technology generates a system with rapid prediction timings and approximate predictive accuracies. However, the AI system's prediction cannot check the valid range of the producing numbers and parameters, and the complex design structures are required. To design an accurate yield prediction solution, a methodology including detailed verification processes and sensitivity analyses against controlling parameters will be investigated. There is also a necessity for innovative technology that ensures very stable and high yields amongst the AI-based smart factories. The potential direction includes a hybrid model design with devising ensemble prediction methodologies and metamodels. The semiconductor industries adopt innovative technologies for the highly productive smart factories. Because of the tremendously complex manufacturing process in these industries, metamodeling with data analytics-based soft sensors will be one of the most suitable solutions. To enhance estimation timing and prediction accuracy in metamodeling-based soft sensor technologies, high-dimensional quantity measurement data can represent input features by various vision-processing and pattern-recognition technologies.

Translating highly unstructured pattern data into structured data by image processing has been focused. Most manufacturing industry challenges are yield prediction problems, where the outcome class of interest is usually significantly under-represented. Yield prediction classes are also examined by focusing on unrepresented class visualization to comprehend the underlying pattern better. The challenges are addressed through improved data sampling techniques, carefully revealing the minority to majority class relationship without losing other data distribution characteristics. For the future direction, potential techniques to further enhance visualized pattern yield prediction and other industries through an anomaly detection approach will be proposed [6].

5.3. Equipment Data

In semiconductor manufacturing, equipment data are widely adopted in data analytics-based yield improvement tasks. Equipment-level or product-level measurements that contain the information of equipment performance/behavior are called equipment data. Equipment data usage includes but are not limited to the maintenance scheduling based on operating status changes, yield prediction based on the performance as well as process variable changes, and product quality/time prediction based on condition monitoring. The equipment data collected in the smart factory are generally time-series data or time stamps. Therefore, it may require nontrivial preprocessing steps to be used in a variety of data analytics tasks. Dealing with high-dimensional features is another common challenge, especially for using shallow models that highly rely on the proper feature engineering in many cases.

Directly applying black-box deep learning models for time-series prediction tasks has shown superior performance over the state of the art [2]. Given their strength, natural and easy-to-use models, long short-term memory (LSTM)-based regression models are proposed to predict impending metrology measurements based on sensor data of the equipment in an efficient way. Recent studies have developed various approaches to advance the analytics of time-series data, either for the sake of product yield improvement or for the sake of process control. In those analytics-based data-driven approaches, detailed equipment time-series data that were measured by sensors installed on each piece of equipment are also used.

These sensor data contain the information on the lot path history of products as well as equipment conditions. To the best of knowledge, however, the state-of-the-art time-series analytics of the equipment data are focused mainly on yield prediction [4]. Even if the yielded product is classified into defective and non-defective by procedure metrology, equipment data analysis is surrounded by chaotic issues caused by difficulties in deal with variant data encodings across tools, diverse features for different engineering problems, and masking many additional data processing and scoping steps.

5.4. Environmental Data

Environmental data in semiconductor manufacturing include temperature, humidity, and pressure, which are vital for various reasons. In some equinoxes or when manufacturing tools are safeguarded, the temperature around the tools may gradually change significantly while the wafer quality remains acceptable. The environmental condition sensors can help determine when the tool is too cold to produce correct work. The pressure of a vacuum chamber is often monitored, which is similar to monitoring critical variables in etch and deposition chambers.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

Additionally, researchers have found that many random failures of tools can be directly associated with fluctuations in these environmental conditions, and excessively high humidity can cause the glass chip to gimmick. These operating conditions can affect chip-to-chip uniformity and need to monitor over time [5]. Some examples are discussed below.
The normalized environmental data were first preprocessed using a time window of 5 minutes and a median if some value exceeds a threshold. Exploit PCA to select a reduced feature set that can best describe the system states [6]. Time-invariant PCA is proposed to provide more relevant and compact features on these time-varying systems. This online method is based on the singular value decomposition of the past data covariance. For each new data record, the new covariance matrix is formulated. Previous singular value vectors can become less relevant; thus, the old singular vectors need to be updated by the previous two SVD vectors. The eigenvalue threshold is used to determine whether to keep replacing the eigenvectors or start refreshing the eigencomponents. The chip-level process states and time-invariant features can instead be rebuilt for equipment monitoring and process/diagnosing tool spread.

VI. STATISTICAL METHODS FOR YIELD PREDICTION

Similar to statistical process control (SPC), industrial engineers are often keen in applying a yield prediction method to proximately assess the fluctuations of death yield within wafer lots/processes so that immediate actions can be taken to avoid further economic loss due to the unachievable yield target. Particularly for semiconductor fabrication processes, most of the yield prediction methods are statistical and rely on the assumption that financial yield losses are stable over time. As a result, predictions are absolute and not based on the current process capability or system state. In fact, as a process flows downstream, the design rules for financial yield losses or critical defects may no longer hold. Several approaches have been proposed to classify defects that cause yield losses into low- and high-impact defects but are not solely based on 1/f process variations.

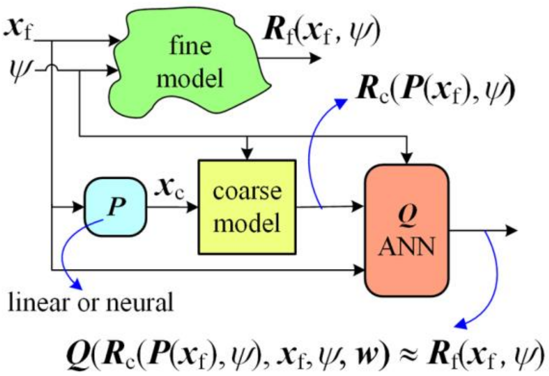

Fig 2: NSM for accurate statistical analysis and yield prediction

This recently-published study proposes a new framework that bridges wide-band defect classification and use of historical yield data to recover the effect of defects on process capability, which is then incorporated into a regression-based zero-defect yield prediction model on two-way decisions (preventive maintenance). This method allows engineers to assess the effect of currently-controlled defects on the capability of future production on early, recommended minimum feature extraction and maximally broad inspection coverage. It is believed to be a new perspective for cost-effective diagnosis and corrective action of yield-ratio defects across different inspection depths [1]. It would also be interesting to combine it with further enhancement of other data analytics methods. Detailed explanation of each method and its applications is also provided. A hierarchy framework to illustrate the flow of data analytics methods in industry is introduced. Understanding and anticipation of data characteristics are the keys to types and objectives of knowledge discovery and utilization leading to various process performance improvements.

6.1. Regression Analysis

Yield prediction is a critical task in the semiconductor industry to parry yield loss. The proactive mitigation measures underpinning this task require an understanding not only of the complex manufacturing processes but also of the implicit derivation of defect distributions based on process data. However, an accurate and holistic comprehension of wafer yield is intricate due to the interaction between multiple big data sources and complex data characteristics. Previous approaches primarily focused on developing effectively accurate prediction models but overlooked the interpretability of the prediction. Knowledge-driven and model-driven methods are either insufficient for accurate prediction or not sufficiently able to interpret cause-effect relationships.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

By integrating high-efficiency knowledge-driven factor extraction, an explainable prediction model for automated high-yield prediction, termed as sERF (explainable regression forest), is proposed.

Unlike the conventional method, which is not automatically tuned or interpretable across different designs, sERF can efficiently and effectively learn and accept semi-supervised predictors without tuning, and deliver transparent prediction with clearly defined cause-effect relationships. Apparently, the proposed schemes present a significant improvement over the conventional retain-and-train method, as they yield a considerable substantial yield improvement. In addition, interpretability and adoption have been mentioned as future areas for work, since the current model is too complicated for engineers to interpret decisions [1]. Rather than being a black box, AI models for yield prediction generally need to deliver rational explanations for users, both for the predictions and for how the features affect decisions.

The rapid advancement in semiconductor electronic products has increased the demand for integrated circuits, driving the rapid growth of the semiconductor manufacturing industry. Because of the complicated texture and intricate geometric structures of ICs, semiconductor manufacture usually consumes substantial fabrication resources, resulting in a lower manufacturing yield. Early yield prediction at the design phase is necessary to obtain manufacturable designs. Numerous works utilize the existing design parameters as inputs to build the yield prediction models, but they do not consider the feature models due to the complexity of building them. A simple good-features-to-good-structures framework is proposed to construct feature models to build quality aware designs under uncertainty, such that early yield prediction is possible. To show the effectiveness of the proposed framework, an industrial case is studied [4].

6.2. Time Series Analysis

To maintain competitiveness and enhance profitability, semiconductor manufacturing companies need to maximize yield and minimize operational and investment costs in a chip manufacturing line [4]. However, with the increasing complexity of a semiconductor wafer fabrication process, it is getting more difficult to increase yield aggressively. Therefore, there exist many interesting yield prediction tasks in a chip manufacturing line. This work focuses on a surface roughness yield prediction task, which uses the measurement values of both manufacturing process and particles in the prior process as features, and aims to predict the wafer level yield of surface roughness. To achieve this goal, a machine learning based approach is developed to rank the importance of features first, and a lightweight feature engineering approach is proposed to reduce the feature dimensionality while preserving most of the original valuable information. Moreover, this approach is generic and easy to implement, and is demonstrated with ordinal categorical target variables and process capabilities quantized as categorical values. An extensive study on the design of regressors is conducted to help production managers to choose suitable regressors to optimize the yield prediction [1]. Semiconductor manufacturing is a multi-stage process involving complex fabrication methods to achieve the desired product properties. In the fabrication process, a wafer is produced through various fabrication processes in different fabrication equipment, during which, the fabrication equipment and processes are adjusted and modified according to the development of semiconductor production technology. Such changes may result in an increase in non-standard devices and lead to variation in the electrical characteristics of wafers in the fabrication process. Consequently, the quality of wafers in time may deviate from the standard or desired value. Thus, a forecasting operation is required to help the fabrication processes maintain the wafer quality outside of manufacturing processes. A time series context-based model is proposed with reasoning mechanisms to address the problem of wafer quality forecasting. Its reasoning mechanism helps to better capture the pattern of localized equivalent quality in the time series. Experimental results on industrial applications demonstrate the effectiveness of the proposed approaches.

VII. MACHINE LEARNING TECHNIQUES

In semiconductor manufacturing, the increasing complexity of integrated circuits leads to more processes operating under tight limits. This leads to an increase in variability, often resulting in defect generation, which decreases manufacturing yield. Metrology data should be analyzed to extract hidden information. A strong industry requirement for data analytics-driven yield prediction is consequently raised, while the increasing accuracy and speed of yield prediction promise business competitiveness.

Yield prediction is a complex regression task with the scattering of class one devices. As an application of data analytics, yield prediction methodology involves capturing mapping relations between the fine-grained data and the corresponding yield before decisions are made on successive operations. In the context of semiconductor manufacturing, fine-grained data encompass parameters and variables quantifying process capability, equipment, environment, circuits, and devices, while yield is defined as the proportion of good devices passed on by manufacturing yield.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

Data analytics-driven approaches to yield prediction in semiconductor manufacturing largely concern data representation learning strategies based on machine learning and deep learning techniques.

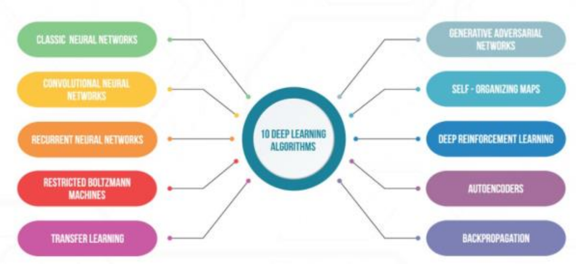

Fig 3: What Is Deep Learning? Definition and Techniques

Numerous machine learning-driven and deep learning-based architectures for yield prediction are constantly emerging, owing to special characteristics like superior performance and robustness. The architecture of yield prediction models sequentially consists of three components: representation, transformation, and prediction. The representation layer collects and pre-processes the input data. The transformation layer encodes fine-grained input data to a yield descriptor via a multi-layers network stack. The prediction layer utilizes a yield descriptor mapped to yield to be strategically predicted. Mathematically, the sequential mapping from data to yield is expressed as follows. The multiple-layer architecture makes it possible to obtain arbitrarily complex functions via a non-linear mapping. Meta-heuristic algorithms are further developed to select an appropriate architecture among its candidates.

7.1. Supervised Learning

The semiconductor manufacturing flow is generally integrated circuits (ICs) wafer preparation and fabrication process followed by packaging and testing process. Wafer preparation includes wafer growth, slicing, lapping, cleaning and surface treatment. While in manufacturing, IC designs are realized and etched onto substrates through a series of fundamental processing steps, which can be classified into four domains: photolithography, chemical vapor deposition (CVD)/physical vapor deposition (PVD), ion implantation/diffusion and etching. Wafer sorting relies on test program equipment (TPE, either functional or parametric testing), whereby chips proving previous fabrication passes qualify as good die and the rest are defined as bad die. With process understanding and an explosion of manufacturing data and computing power, automated yield prediction (Yp) is possible [3]. With more accuracy prediction to identify a possible bad result in wafer processing as early as possible to reduce unnecessary cost incurred in testing and having less human resources intervention in the batch prediction over a period of time.

Generally, machine learning (ML) has two classes. One of them is supervised learning model, which is most common in production yield prediction [4]. Supervised learning models within various machines used in modeling approaches. Determining whether enough results will yield results from the simulation data and whether filtering on data without modeling is necessary for modeling for a soft sensor is likely to be a challenge. Classifying the condition of past manufacturing results as passive and predicting why those conditions yield bad is active. We examined two supervised learning models (Artificial Neural Network, ANN and Extreme Gradient Boosting, XGBoost), machine learning algorithms, to evaluate how they classify the past records of wafer simulation for the condition yielded bad results. With that classification model, the problems needing to be investigated in the examined wafer simulation data can then be of concern.

7.2. Unsupervised Learning

Unsupervised learning (UL) is a class of machine learning methods that uses data with no labels. The aim of unsupervised learning is to discover an underlying structure of the data. Using this underlying structure to aggregate the data into groups makes it possible to perform subsequent supervised learning tasks like classification or anomaly detection on them. It has the potential to push forward both productivity gains in manufacturing and the fundamental understanding of new healthcare data [1]. There exist numerous unsupervised machine learning algorithms, each with its respective assumptions and strength, the most popular which are: k-means, Gaussian Mixture Model (GMM), hierarchical clustering, and non-negative matrix factorization (NMF).

Data preprocessing in unsupervised learning focuses on feature selection and dimensionality reduction leading to better clustering results, with several methods available such as PCA, t-SNE, autoencoders, and so on. Proper partitioned clustering, the default of unsupervised learning, generates a good number of clusters where each cluster contains a considerable amount of points.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

Beyond clustering quality, it's interesting to know how strictly each point belongs to its cluster. The fuzzy c-means (FCM) clustering algorithm is an extension of k-means clustering that allows datasets to have overlapping clusters, i.e. soft membership. FCM achieves this by assigning each point an individualized membership degree determining its strength as a member of each generated cluster. An additional output of the FCM algorithm is the measure of fuzziness per cluster indicating widely each point is supported by its clusters; widely distributed clusters have a high fuzziness degree.

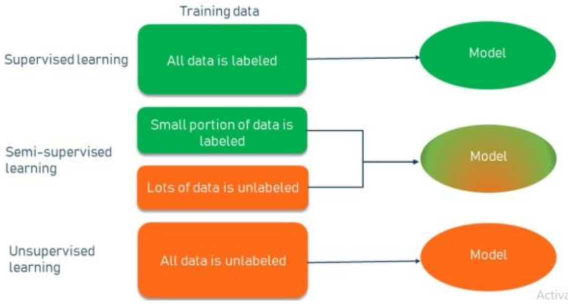

Fig 4: Semi-Supervised Learning

Fuzzy clustering can help resolve the topic of cluster filtering since while the clustering output can clearly identify clusters, these clusters don't necessarily mean any importance unless a subset of the clustering relates to a useful characteristic within the entire dataset. This can help ensure that the number of selected clusters is reasonable and that the fuzzy clustering output can effectively rare meaningless overlapping clusters, encouraging trusts to the clustering output.

7.3. Deep Learning Approaches

Yield prediction, as a subcomponent of yield enhancement, plays a pivotal role in the semiconductor industry. With semiconductor chip designs advancing toward more intricate architectures, complicated manufacturing processes with ever-shrinking feature sizes, and diverse products catering to burgeoning computing capabilities, effectively predicting yield is essential for resource optimization, cost control, and yield improvement strategies. In recent years, many recent deep learning techniques, including comprehensive CNNs, temporal models with attention mechanisms, and ensembles of hierarchical temporal models, have been employed to improve yield prediction reliability by capturing complex features both in a structure-aware and sequence-aware manner. Despite the progress, the rapid evolution of semiconductor manufacturing technologies poses significant challenges for deep learning models to handle potential concept drifts.

To tackle this issue, a multi-scenario in-die yield prediction framework utilizing exception-based explainable ML techniques is proposed. Various data-driven yield prediction models, including shallow ML techniques and neural networks with different complexity, are developed and benchmarked. A novel prediction explanation approach based on counterfactual reasoning is proposed to analyze model prediction behavior under anomalous scenarios. Empirical studies are conducted on production HVM data sets, demonstrating that the proposed framework enriches the current semiconductor smart manufacturing analysis capability with a preliminary attempt to enhance the transparency of black-box models [1]. In semiconductor manufacturing, product quality is assessed using a combination of post-silicon tests including Vmin, IDDQ, and DC test. Test results are aggregated as a product-level yield score which is used to classify the product as either 'good' or 'bad'. Focusing on test quality prediction, the yield score is time series data that contains rich yielding and non-yielding behavior. Modelling time series data is challenging as the model needs to infer the right temporal position and interval between observations. Despite conventional forecasting approaches that heavily rely on engineering heuristics to extract temporal features from the time-series data, the proposed method focuses on deep neural networks to automatically exploit unbiased temporal features hidden in the time-series data.

In each time interval, the Bi-directional Long Short-Term Memory (Bi-LSTM) model and Time Distributed layer exploit temporal features based on state transition and attenuation principles, and the Convolutional Neural Network (CNN) attains near-zero latency and describes cross-sectional behaviors in a single time interval. Further, a focus on explainability motivates use of surf time interval aggregation which allows domain experts to consult and interpret model predictions using a temporal behavior understanding framework. Novel designs of aggregation-level explainable models are proposed as a trade-off between accuracy and explainability [4]. The time- and aggregation-scale- based CNN + Gradient Weighted Class Activation Map model is highly explainable and accurately predicts time- aggregated quality.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

The proposed models can empower domain experts to take actionable insights to correlate and stop malfunctions in the manufacturing process while keeping the product yield high.

VIII. DATA PREPROCESSING TECHNIQUES

Manufacturing semiconductor products is a very complicated and complex process. A semiconductor wafer will pass through many stages during the fabrication processing. Semiconductor yield is the main measure of the effectiveness of a semiconductor process. In these processes, some wafers will end up as 'bad dies' which have to be discarded. Therefore, yield prediction is essential in semiconductor manufacturing.

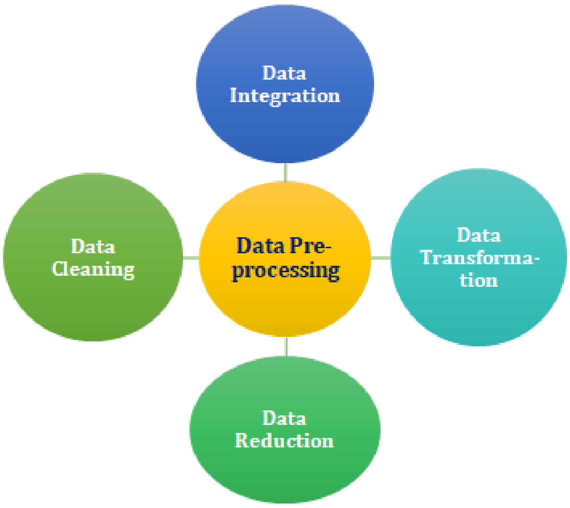

Fig 5: Data preprocessing techniques

Data preprocessing techniques also include handling categorical data, outlier detection, and data normalization. The categorical data handling techniques include one-hot encoding and lab encoding. Numerical data outliers can be identified using interquartile range based filtering. For data normalization, eliminating features can be performed to get rid of constant features. Standardisation and min-max scaling can be used to normalize data feature vector values. Outlier cleaning can significantly decrease predictive yields training time with random forests and gradient boosting model.

The product yield should be classified into two categories: a 'best die' that meets requirements, and a 'bad die' that has been produced with defects. Before proceeding, the data set must be imported to add necessary libraries. Next, the dataset should be split into a training and a test portion. To increase clarity and enhance signal quality, the 'train_df' data set should undergo normalization. Only relevant numerical features are needed in the model training stage, which should undergo NaN yield cleaning, outlier detection and removal, and model feature selection steps prior to receiving training data.

8.1. Data Cleaning

Frequent representations of labeled data patterns indicate either good yield outcome (G) or bad yield outcome (B) due to the faults in the mold/test data or on the steps involved in CHK YLD. The one that reflects bad yield with a high-degree scatter pattern is selected for the analysis. The outlier candidate patterns considered as a new cluster are validated with additional knowledge data or modeled as non-events, if possible [3]. The pre-processed data patterns are clearly visualized by a scatter plot; however, detailed knowledge on the faults causing bad yield is required to preprocess the data before transforming it into a visualized pattern dataset [5]. In a semiconductor IC manufacturing process, many types of structured data, such as parameters setting the flow of fabrication, inspection visualized/physical data, and process metrology signals, are captured and recorded in various data wolckers at every processing step. Nonlivive outward processing test observations are completed for the together returning wafers. However, due to high manufacturing complexities and the nature of the data, predicting the yield outcome is significantly difficult. The predominant class of production is of good-yield (G) wafers, but to ensure that no defective chips get through to packaging, modern ICs often contain many hundreds of MB of test data, with >99% of the test scenarios passed, which often become a performance bottleneck. This situation raises an imbalance classification task to be solved. To propose an accurate yield prediction algorithm, the development philosophy was divided into three blocks.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

The first block was to preprocess visualized pattern dataset too and frequently, separate outlier candidates of B patterns for investigation. In the second block, an information integration process was applied to assist data preprocessing. Hybrid classification and sampling schemes were presented in the third block to improve the modelling precision.

8.2. Data Normalization

To join the analytics pipeline for yield prediction, the data from the manufacturing process, parameters, setup procedure, and yield results must be assembled into a common format. The presentation and structure of manufacturing data can vary significantly across different tools in the fabrication process. Data collected during inspection for testing yield results can be numeric, video, or image data. Undetected noise and defects may also exist. Visualizing and exploring huge datasets together can be difficult and inefficient. Even with advanced data exploration capabilities, data from selected processes must be preprocessed, aggregated, and filtered to maintain high data quality. This data preparation is usually performed differently across different processes and is time-consuming [3].

In practice, most of the yield prediction models either consider data from only the last stage process, or data from several stages without a systematic strategy on what data is useful for prediction. In these cases, the risk of missed out on critical data can be high. It's possible that the selection of data filtered on only one characteristic of the patterns for modeling may lead to lower performance models. With the vast number of possible data combinations, building models at too early a data filtering step may result in lower performance. In addition, eliminating too many data inputs may lead to the loss of valuable data and unexpected yield influences. On the other hand, building a model with all the data collected along the way from all selected processes can be inefficient and result in high model complexity, making it difficult to examine or improve model performance. Therefore, a systematic data exploration methodology to evaluate and select data combinations for yield prediction is developed.

Existing data exploration tools only present the data exploration results but do not provide the integration with modeling steps or an interface for model performance evaluation, making the model building process incomplete. Relying on manual evaluation, a candidate yield prediction model and its input data from tens of millions of possibilities must be selected manually in an iterative, headache process. Automating the candidate model selection process and embedding the data exploration and pre-processing steps in it can greatly facilitate data-driven yield prediction model development.

8.3. Feature Selection

To enhance the understanding of yield from multiple perspectives and boost model performance through rigorous feature selection, a tree data structure is proposed. To shape the yield drivers and extract features from the aspect of process parameters, the features are extracted from Chebyshev polynomial fitting curves from Run to Run process parameter data. At the same time, features are constructed to quantify the relationship among equipment, feature parameters, and wafer attributes based on the procedure of processing a wafer. The constructed features consist of not only orthogonal equipment features but also cross-feature parameters, which contribute more profound insights and quantitative meanings than previously focused features. Based on this tree data structure, massive features are extracted from 1-U equipment parameters to 1-2 interactive equipment parameters, and a dataset of 10320 × 47 features is obtained for further modeling.

Feature selection (FS) is crucial to ensure focused research, enhance model performance, and speed up modeling processes. Recursive feature elimination (RFE), which combines the top two best model performances among ten models to rank and select features as candidates, is performed in a model-agnostic way. The tree structure feature is initially converted to a simpler form, where a coefficient index is generated according to the tree structure and feature characterization. The multi-model integration and RFE are gradually applied in this index list to obtain reduced features while boosting model performances. As a result, FS totally reduces the dimensions of 10320 × 47 features to 84 × 39 features through three iterations of elimination.

IX. MODEL EVALUATION METRICS

Various metrics exist for model evaluation, such as Relative Absolute Error (RAE) and Mean Absolute Error (MAE). However, these metrics assess performance from a single-dimensional viewpoint of measure error with no regard for forecasting quality. In the semiconductor context, models can forecast the yield as an overall percentage of yields and as a distribution. Evaluation metrics for comparison of a pair of yield-predictive models need to treat over-forecasting and under-forecasting of yields on an equal footing. The design of such evaluation metrics is not without merit in other domains of applications. In the Telecom sector, the accuracy of call forecasting models may be assessed as: 'To what extent do forecast calls failed to be made? How many actual calls are not foretold?' Holistic metrics quantify the coverage of calls measured from a network's signaling nodes in terms of CIF, and the delay, error, and non-covered calls measured

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

in terms of PER [7]. These metrics essentially aggregate performance measures across the lenses of non-given call. Outside domain-specific applications, such a framework of designing evaluation metrics needs to treat quality comprehensively, especially when dealing with curvilinear measures.

In the semiconductor industry, inspection of wafers is critical for improving product yield, product quality, and reducing costs [4]. However, it is costly and challenging to implement. To alleviate the cost, the newly proposed onchip soft sensors can sense process and equipment conditions to predict impending normal inspections using runtime sensor data and predictive models. The widely used procedure is to train a regressor from historical data of normal inspections and their corresponding sensor conditions. LSTM can characterize the sequential dynamic pattern of the sensor signals and capture the implication of sensor signals on upcoming inspections. Any tuning policies aimed at maintaining process conditions can apply re-calibrated models for new product lots. However, there are only very few works focused on soft sensing forecasting. Furthermore, LSTMs are novel for the semiconductor MEOL inspection forecasting task since traditional statistical methods like linear regression with auto-regressive moving average terms are dominant.

9.1. Accuracy

Given the complexity of the domain knowledge in yield prediction, the experts in this domain are limited in number and usually cannot be transferred across companies. The evolving and volatile business, sensor, metrology and edge detection, model drift, and concept drift in manufacturing processes can cause prediction accuracy degradation across time. The width of the distribution or population should also be acknowledged, where models having similar accuracy can still make totally different decisions. Finally, an explanation strategy that explains the automated machine learning solutions is decidedly required as end-users need to understand how the solutions came about, and trust it before they act on it [1]. Modern smart manufacturing systems generate extensive multivariate time series data which can be monitored and analyzed for gaining insights into the production-building process. This also applies to semiconductor manufacturing, where time series data can provide critical insights for yield predictions, thus enabling informed decision-making for optimal process control and lower manufacturing costs. Nevertheless, timely predictions of these yields at a required fine granularity remain challenging owing to the additional complexities of massive production volume data, varied process flows, heterogeneous fabrication facilities, and ever-increasing feature space. Data-driven approaches have emerged as a mainstream way to build accurate yield models to address these challenges [6]. Such approaches can leverage modern machine learning (ML) methods to construct accurate models for mapping relevant features to yield values, generally requiring limited prior knowledge and assumptions of the underlying physics. Nevertheless, the data-driven approach relies heavily on the availability of sufficiently high-quality training data.

Corrupt and incomplete data often lead to inaccurate models, thus limiting their deployability in practice. Furthermore, most state-of-the-art data-driven approaches are designed either to accurately predict yield or to select relevant features but selecting relevant features to facilitate accurate yield prediction using the selected features has yet to be explored in the semiconductor manufacturing context. Recent years have seen a growing demand for machine learning based yield prediction models in the semiconductor industry, but there are still unfulfilled needs. Domain knowledge in this area is retained largely in the form of best practice so can be very difficult to share. Some of the knowledge can be extracted from products built in the past, but others cannot, because they are still the object of intense research.

9.2. Precision and Recall

Recently, the focus of attention has been drawn to the issue of precision and recall along with their application in manufacturing yield prediction. It had been argued previously that the best performing machine learning are such algorithms that can handle an extreme skewed data situation. It is very essential to improve yield prediction in a semiconductor manufacturing process. A major source of revenue loss for these manufacturers is the failure of ICs or silicon chips after testing them. These failures can be caused by mismatches during the manufacturing process or deterioration due to environmental factors. On a newly fabricated wafer, inspections are performed after each of the layers. These inspections detect the trouble area to divert the wafer to rework process or to scrap. Noise filtering methods are employed, in order to ensure that the noise data does not corrupt the prediction model i.e. for each table and outcome prediction the observation data should only contain the events of interest that were measured. Predicting the yield outcome in a non-closed-loop manufacturing process can be achieved by visualizing the historical data pattern generated from the inspection machine, transforming the data pattern and mapping it into Machine Learning Algorithm for training, in order to automatically generate a prediction model without the visual interpretation needs to be done by human. This has been successfully implemented with Kaizen repeat inspection, malfunction, and defect population detection in 100 mm Wafer Probe, Tune-in and Blocking foundry event pattern visualization to identify pattern root causes in 12" Wafer Probe-Tune in and other Similar Test, IC device failure interpretation with scoring pattern and residual impact trajectory visualization in IC Sampling Test for Wafer Sort quality analysis and prediction.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

The original per-run depictive pattern generated per inspection machine is then newly employed to directly visualize the trouble area trajectory pattern, thus all subsequent operation stage data can only show the trouble area without noisy non related data. This would become an improved Machine Learning Model visualization pattern input data, yet challenging because it compares to the original visualized pattern, the number of rows and class types of this newly visualized data is highly unbalanced. Data sampling needs to be performed, but the reliable destructed sampling techniques are not yet available. Consequently, it requires a new sampling technique to suit the newly transformed visualized pattern data and highly skewed data situation [4]. The skewness of yield data is manifested by previous studies using the ratio of the number of photographs of in spec and out spec devices. Such data imbalance poses a greater challenge for ML learning processes. Applying a common ML algorithm with standard settings on imbalanced datasets resulted in biased predictions and poor performance on out spec devices. The yield prediction of a semiconductor test is a highly skewed data case. Companies try to minimize costs and maximize profit by producing the highest possible number of in spec devices. Only a very small number of out spec devices are created for further test and failure reports. In previous studies, the yield was depicted with a 1:1 ratio. The outcome for each wafer run prediction is depicted as a blue dot state in the Yield State Vis, representing in spec or good devices, while the pink state represents out spec or fail devices.

9.3. F1 Score

The F1 Score measures the model's accuracy of positive predictions, it takes both precision and recall of the model into account. Overtime the joint detection of both good and bad yield classes will be significant, thus F1 score is chosen as the evaluation measure. The solid curves denote results on the synthetic dataset (with unbalanced classes), while the dash curves correspond to result on the balanced dataset. First, it can be seen the F1 score drop sharply when using the independent limited training dataset (less than 500 records). The proposed method is the best choice among heuristic methods. The best thresholding criterion is found at the very early stage of training (less than 300 records) of the multivariate outlier detector, which indicates the smart manufacturing system is unstable and it is difficult to detect yield relevant fault patterns under such condition. Therefore, it is better to first build robust models for process monitoring and control under uncontrollable variation to stabilize the system. After that, yield prediction models can be built right away since tragic yield faults can be easy detected given a stabilized manufacturing system. The method is less efficient on finding yield predictive features from a limited amount of data comparing with statistical based thresholding criteria, but has great efficiency on training dataset shrinkage, prediction speedup and interpretability which satisfy the requirement for online machine learning.

It can be seen as the amount of training data grows, the gap between the simulation and practical results narrows down first but then they diverged for the simulations settle down to some extreme points once the number of simulated records grows over 20,000. Heavy tail distribution is inherent in semiconductor fabrication process, where a few monitors records a significant portion of total variability. On the contrary, localized covariate shifts exist in real industrial processes, where the majority of observed process conditions vary in a significantly narrow range which may lead to poor performance of monitoring models. Localized data slices are constructed with respect to the temperature of furnaces that maintain a relatively stable working conditions to compare the monitoring performance.

X. CASE STUDIES IN YIELD PREDICTION

This section highlights two case studies of data analytics-driven approaches to yield prediction. Their purposes include explaining how predictive models can be developed by using deep learning and feature engineering, as well as demonstrating the hands-on skill of conducting yield prediction tasks using data analytics tools. All the analysis is carried out in a chosen software platform with no need for deep programming skills.

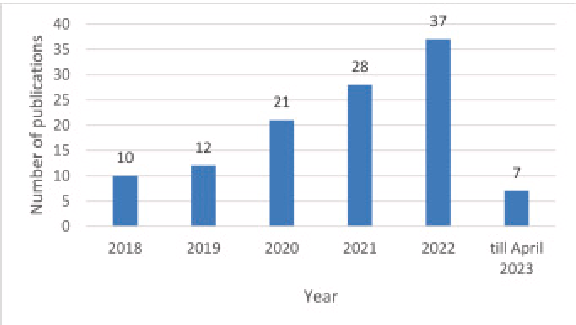

Fig: Crop yield prediction in agriculture

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

The first case study illustrates how predictive models can be developed using historical yield data from a semiconductor fabrication plant (Fab). The task is to predict the yield count of each wafer in the fabrication process. Three types of predictive models are studied, including artificial neural networks (ANN), generalized linear models (GLM), and support vector regression (SVR). It is found that the ANN model substantially outperforms the other two models. The predictive capability of each model is further assessed based on the test dataset [1].

The second case study focuses on deriving the predictive feature set and the predictive model using historical process data from semiconductor stepper equipment. The project involves more than 50 variables from different sources, such as stepper equipment, process files, and inspection results. Among them, features such as exposure field control, focus control, stage control, offset control, on-block wafer number, and defect count after first exposure are found to contain significant predictive capability for yield loss prediction problems. Five data analytics tools are studied at length in terms of general information, user interface, and use cases. Each tool reveals different insights regarding yield loss issues, but all of them show great potential for large-scale applications in semiconductor fabrication which contains billions of records and complexities every day.

10.1. Case Study 1: Process Optimization

As the processes become more advanced, the yield concerns both increase and become more difficult to optimize. The assembly and test yield comprises an array of complex process steps and product characteristics. To effectively improve yield, it is essential to analyze and prioritize these dependent process steps. Many process steps tend to have a tightening shift across different parts of the current data age. Furthermore, the fixing timeline could be limited due to tight coloring or ramp closure due to customer requirements. Therefore, it is important to not only detect but also analyze the problem prior to an attempt for fixing.

Enhance yield has become a pivotal lever to reduce costs and boost financial returns. Yield enhancement is one of the most vital concerns in the semiconductor front-end process. To tackle this issue, machine learning (ML) has been increasingly applied to augment a wide variety of yienhancement strategies, including analyzing critical process steps by feature selection, assisting in troubleshooting and process optimization by data mining, detecting the potential cause of anomalies by clustering algorithms, automatic defect classification, etc.

Despite the inherent contextual differences across those ML techniques, they share a common concept: constructing predictive models based on historical data and model-agnostic analysis tasks to identify feasible actions to either improve yield or prevent yield issues. To enhance the yield of an advanced semiconductor product, it is common to seek for methods to effectively analyze and prioritize the dependent yield concerns. A statistical method is proposed to examine the characteristics as well as the dependency structure of the initial and final product characteristics. Furthermore, a model-agnostic method is also proposed to assess the criticality of the process steps implied by the product characteristics via quantifying their global impact on the yield variability.

10.2. Case Study 2: Defect Detection

Defect detection is a key technology for semiconductor fabrication that provides critical data for both production efficiency analysis and yield enhancement. Semiconductor manufacturing defect detection uses a variety of data streams, such as optical images, infrared images, process variability logs, and particle measurement logs. However, in addition to these conventional data sets, more and more new data types arise, such as wafer-level power maps and images. These characterizations are promising for achieving fault-level defect detection, but there are urgent needs to conduct thorough investigations on how to develop the best model for new data types and how to combine examples of different kinds of data to provide a better detection performance. This section discusses both of these important topics by formulating them under a framework of advances in transfer learning (TL). The development of this TL framework can shed light on exploring interactively anomalies from images to signals and vice versa [1].

A semiconductor defects detection system is proposed that consists of a new conditional GAN-based model combined separately trained semantic segmentation and domain adversarial networks. With the embedding construction that contains foreground, background, and non-pattern areas for different application phases, illustration using TL for domain-aware defect detection considering the uniqueness of different data types is described. The architecture of current defects detection system consists of stages from data collection to analysis results. Semiconductor defects presence can be detected by embodied transformation from images to signals through an end-to-end framework. This current development explores how snippets in need can be intelligently suggested by searching in time-frequency and domains at various levels. The proposed methods advance analytics for exploring manufacturers' product stories in wafer-level maps leveraging the uniqueness of Wafer-Level Maps.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

Within a wafer, yields are generally higher for devices within a die group than in the surroundings which may be attributed to cluster effects. The research investigates the combination of assimilation mechanisms for addressing challenges from large-dimensional and scattered feature values. The outputs from different mechanisms can be plugged into XGBoost facilitating performing analyses of production yield and die assembly

XI. CHALLENGES IN YIELD PREDICTION

For semiconductor manufacturing, yield has been widely adopted as a key performance index and criterion to determine the success of the fabrication of integrated circuit (IC) chips. Yield is defined as the ratio of the number of good wafers and dies over the total number of wafers and dies processed by the front-end-of-line and thereafter back-end-of-line manufacturing processes, respectively. Die yield is determined by wafer yield and die manufacturing yield. Wafer yield is usually defined as the percentage of good wafers after various processing steps, and die yield refers to the percentage of good dice after the wafer is diced. Wafer yield and die yield are the most important metrics to determine the efficiency of wafer fabrication plants (WFPs) and assembly and test factories (ATFs), respectively [1]. In semiconductor manufacturing, it is widely accepted that data-driven approaches are a powerful and effective means in the smart manufacturing revolution for the enhancement of various key performance indicators (KPIs), such as yield.

Nonetheless, yield prediction is a challenging task in the semiconductor manufacturing industry. First, the data are usually large in scale. Such large-scale data contain huge amounts of information while making it challenging to effectively analyze them by various means of data-driven approaches. Importantly, large-scale data make it costly and time-consuming to search a hyperparameter combination that can lead to an optimal or near-optimal forecast performance. In addition, data drift is a common occurrence due to environmental changes. Data drift phenomena can significantly deteriorate the model forecast performance. In yield-driven strategy, the expected profits associated with the yield prediction uncertainties must also be considered. The yield prediction model (YPM) may hence be expected to output not just a single yield forecast but also a quantification of the prediction uncertainty in terms of yield percentile bounds, which is an even more challenging task.

11.1. Data Quality Issues

The huge amount of data collected from the semiconductor manufacturing sentry over equipment, data acquisition, control, process monitoring, and defect inspection can provide a large variety of insights into the manufacturing process and product quality. However, there are always some potential issues in data quality which could lead to misleading or wrong inferences. In the semiconductor process where several equipment vendors and different locations are involved, the full understanding of collected data may require additional engineering and business expertise. In this chapter, some common data issues are introduced and discussed. These issues include unmeasured variables, missing data, data preprocessing and calibration, and embedding human expertise. Novel and accurate methods and algorithms that alleviate the effects of these issues in order to enhance the data quality of wafer-level inferences are developed.

Figure 11.1 provides a schematic of potential challenges to the accuracy and reliability of algorithmic data mining. Before deploying any data-driven model that is assumed to be understandable and verifiable, key aspects regarding the quality and characteristics of the available data must first be studied. Although the data used to train and apply the model can be examined, modeling assumptions along with software implementation can be erroneous or incomplete. Each of these categories is further discussed and illustrated with examples.

Most of the data analysis works for yield prediction in semiconductor manufacturing report history data used and mining approaches, but the original data collection process is absent. This is because the semiconductor process involves various equipment from different vendors and implementations, and the acquired data are in an unstructured way. In order to mine historical data, segmentation and mapping of the data need to be performed. Otherwise, mining tools will not be able to process such raw data. Similar to financial data time series analysis, shape matching and pattern definition for discrete real and intersampled data need to be defined first before matching at any time series instance. To the best understanding, there is no work in yield prediction modeling for capturing data in structure, decomposing data into segments, tracking, grouping, and pattern comparison.

Data quality is a crucial aspect in the application of algorithmic data mining to real-world manufacturing. Specific data issues that can influence robustness and performance are outlined. There are many questions and challenges about data quality perseverance in the semiconductor manufacturing process. Essential topics and representative work addressing these issues are presented.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

11.2. Model Overfitting

While applying any machine learning algorithm or predicting model in real life, the feedback time of machine learning model or predicting model to produce the predictive result is an important aspect. If the feedback time required to produce the predictive result is too high, then the predictive model may not be useful for the end users. For the designed yield prediction model using machine learning, the feedback time to produce the prediction result is highly dependent on the number of data points in the dataset, number of features in the dataset and the type of machine learning algorithm used for modeling. In order to shorten the feedback time, several experiments were proposed and run in chapter 5, after that an acceptably short feedback time was achieved while retaining the prediction model's predictive performance. In yield prediction, the classification error not only impacts the overall yield but also affects the net profit. For yield prediction, while designing the predictive model using machine learning, care must be taken not to overfit the model on the training dataset as there will be penalty costs if too many misclassifications take place. Therefore, in the yield prediction model design, the model's overfitting detection and avoidance has been conducted, and acceptable model robustness and generalization capability performance have been achieved for the applied yield prediction model.

Eqn 1: Basic Overfitting Equation Form

$$\hat{y} = f(x;\theta) = y_{\text{true}} + \epsilon_{\text{noise}} + \epsilon_{\text{overfit}}$$

Where:

- $\hat{y}$: Model prediction
- x: Input data
- θ: Model parameters
- y_{true}: The true signal or pattern
- ϵ_{noise}: Random noise in the data
- $\epsilon_{\text{overfit}}$: Error introduced by the model trying to fit the noise

In the above section, through a simulation dataset, an explanation of how a predictive model could be overfit successfully and how to avoid it by using the cross validation method has been provided. Several statistical measurements of the cross validation overfitting detection method have been discussed. In the second experiment, several measures of the overfitting detection method were applied to the actual yield prediction model. It has been determined that out of all the statistical measures, the training sum of squared errors versus number of features graph could clearly show the overfitting occurrence point and has been used to detect yield prediction model overfitting. However, this measure only detects the overfitting occurrence point, and additional measurements should be used to avoid the detected overfitting. Therefore, in the future work, variety of techniques to avoid overfitting could be tried or developed and integrated into the yield prediction model building [1].

11.3. Integration of Data Sources

The semiconductor manufacturing process and environment is a highly technology-driven and complex system consisting of large amount of machines, back-end and front-end manufacturing processes, data sources, etc. A manufacturing process could involve different steps or units with different equipment and materials in the form of a complex manufacturing system [1]. Given the complex manufacturing processes, numerous potential input parameters or factors (N) that might impact the performance are typically involved in manufacturing. However, in reality, only a small proportion of the input parameters ($K \ll N$) are found to be critical or significant in determining the product performance or display a reasonably high correlation with the product performance. Therefore, it is fundamentally important, yet practically very challenging, to explore and identify the important input parameters or features affecting the semiconductor manufacturing performance in real-time.

Data or data source is a fundamental and key element in data-centric intelligent vs. smart manufacturing. In semiconductor manufacturing, various data types of different frequency, value schema, etc., are acquired at different stages of the manufacturing process from various data sources including static local databases or storage, or via streaming or communication wirelessly from multiple sensors installed on machines or equipment in the form of IoT or in an online fashion. Identifying, integrating, preprocessing, and enriching data from heterogeneous data sources (including production, test, and inspection data), would provide a broader and deeper understanding of the manufacturing process and equipment status in a more comprehensive and bigger picture for intelligent vs. smart manufacturing such as yield, on-line performance prediction, optimization, decision-making, etc.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

XII. FUTURE TRENDS IN YIELD PREDICTION

For address yield enhancement in semiconductor smart manufacturing, a data analytics-driven approach involving Explainable AutoML with adaptive modeling is proposed to deploy machine learning techniques. The proposed approach seeks to mitigate the complexities of developing machine learning models in smart manufacturing by automating the implementation of data analytics while maintaining a high level of transparency and interpretability. An Explainable AutoML approach with adaptive modeling is conceived to automatically deploy machine learning techniques for yield enhancement purposes and provide transparent insights. The Explainable AutoML system integrates the AutoML and Explainable AI components with a complexity-adaptive modeling capability to match the problem complexity of different fabrication facilities and yield improvement scenarios.

Eqn 2 : AI-Driven Hybrid Models

$$Y = \alpha \cdot Y_{\mathrm{DSSAT}} + \beta \cdot Y_{\mathrm{AI}} + \gamma \cdot \Delta Y_{\mathrm{error\_correction}} + \epsilon$$

Where:

- Y_{DSSAT}: Simulated yield from DSSAT crop model (or APSIM, AquaCrop, etc.)
- Y_{AI}: Yield from ML model trained on satellite, sensor, and historical data
- $\Delta Y_{\mathrm{error\_correction}}$: Correction term learned by AI from residual errors
- α, β, γ: Weighted coefficients learned through model calibration

A proof-of-concept Explainable AutoML prototype is developed with its architecture and workflow. The completion of an industrial use case study demonstrates the great potential of the Explainable AutoML approach to facilitate the adoption of data-driven analytics in smart manufacturing. The Explainable AutoML approach holds significant promise in enhancing the efficiency and effectiveness of product yield improvement and profitability margin increase in smart manufacturing, with substantial social impact. Much of the great benefit can be achieved by incorporating the proposed approach in the smart manufacturing framework of other complex manufacturing industries, including aircraft, transformers, integrated circuits, containers, ships, etc. Mathematical modeling, simulation, and other machine learning techniques, such as reinforcement learning, anomaly detection, and outlier classification, can also be adopted in the proposed approach to adapt it to various domains.

Semiconductor smart manufacturing opens new avenues for integrating advanced technologies, driving research worldwide. A technological research gap is identified in closing the promising data-driven analysis techniques and algorithms with the impediment of excessive domain and technical knowledge requirement for implementation. To enable wider adoption of advanced data analytics, the complexity-adaptive Explainable AutoML approach is proposed to take over the development and deployment of data-driven models by seeking automatic solutions. Its great impact in the smart manufacturing domain is highlighted by focusing on a specific implementation to enhance product yield with low defectivity, which is one of the most vital priorities across industries, especially in semiconductor manufacturing.

12.1. AI and Automation

The semiconductor industry is continuously evolving, and with that evolution comes both new challenges and new opportunities for improvement. While there have been many advancement opportunities identified in integrated circuits, there has yet to be an opportunity for integrated circuit improvement identification in the manufacturing realm. However, sourcing current manufacturing processes such as critical features, lithography settings, target films, and slit width has generated insight into improvement opportunities [1]. Additionally, current manufacturing defects in concurrence to these opportunities can be identified and sourced by analyzing the results of Statistical Process Control charts. Using this data, deep learning models can be applied to detect which current defects are most frequent or how defects manifest over time. Once these defects have been identified, data mining techniques can be used to analyze the results. If there are moderate results after analyzing current defects, process variance analysis is recommended.

The background and definition of defects were provided. However, there needs to be a description of how the data is formulated before analyzing it. Parameters such as defect shape, defect type, defect area, defect magnification, and manufacturing processes require a deep learning model that can detect the defect type of interest. The VLSI fabrication process development consists of multiple processing steps that fabricate semiconductor chips. It uses photomasks to transfer the desired design to wafers, however there are limits in terms of the printing fidelity of such features [9]. The yield learning framework estimates the 50% non-defective yield of a design as sample wafer maps are obtained over time. For defect propagation prediction, Monte-Carlo sampling is needed to sample potential defect candidates not declared by the layout rules.

ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91217

Given a set of test failures, the quality analysis engine generates a quality report to help the experts in determining the most efficient logic diagnosis solution. Among numerous solutions, efficiency based on defect coverage and engineering resources is considered. Constrained integer programming is applied to find the solution. In production data analytics, one of the challenges faced today is to leverage entire data from legacy patterns or new data such that quality issue detection schemes can be derived in real-time and without expert involvement. To address the challenge, a Bayesian framework accommodating both pattern level and signal level similarities with hidden factors such as background noise and structural variability is proposed.

12.2. Real-Time Analytics

The semiconductor industry is a rapidly evolving and fast-growing market. Yield, the ratio of the number of good dies to the number of total dies in a wafer manufacturing process, is an important measure of finished products containing functional chips. It is vital for semiconductor manufacturers because of its direct consequence on cost efficiency and market competitiveness. Due to the high fixed costs of building a fabrication unit (fab), to maintain profit margins and long-term growth, enhancing yield has become a crucial lever to cut down costs and increase profits by tightening design and process margins, removing sources of yield loss, and eliminating the impacts of variations on yield. In an advanced logic wafer fabrication facility (fab), a 1% increase in yield corresponds to about-dollar 150 million net profit increase in a fab with a monthly revenue of about dollar 120 million [1]. This is primarily due to the exponential nature of yield loss versus the process defect density according to the "yield learning curve" style of models.

Eqn 3 : Real-Time Analytics Equation: General Form

$$O_t = f(D_{1:t}, \theta_t) + \epsilon_t$$

Where:

- O_t: Output or prediction at time t (e.g., yield estimate, disease risk, irrigation need)
- $D_{1:t}$: All data received from time 1 to time t (streaming inputs)
- θ_t: Model parameters updated dynamically as new data arrives
- ϵ_t: Residual or uncertainty at time t

To gain insights into the manufacturing yield and explore opportunities to improve yield, data analytics-driven approaches have been widely applied in semiconductor manufacturing. Over the past two decades, vast amounts of manufacturing data from production equipment, material handling equipment, and metrology have been accumulated. The vast amount of the data has motivated the application of statistical and machine-learning (ML) models on wafer, lot, process, equipment, sensor, and inspection levels. Applying statistical methods, including logistic regression, linear regression, multivariate analysis, partitioned based regression, and historical BIN prediction methods to yield-related data have been widely studied [2]. These statistical methods have many advantages, including interpretability, robustness, and easy implementation. However, traditional modeling approaches require extensive expertise in terms of methods, implementation, and data management, which becomes a barrier to the rapid integration of data analytics and fine-tuning of models to planter changes in semiconductor smart manufacturing.

As ML technologies advance in performance and efficiency, consideration has been given to applications of various ML techniques in yield analysis. To name a few similar attempts to boost yield, feature selection methods, including embedding, wrapper, and filtering approaches, to find critical process steps driving the wafer yield are proposed. Further data analytics approaches using popular ML techniques such as K-means clustering, clustering algorithms, and random forests to troubleshoot root causes for the outlier wafer yield have been studied. Throughout the path from "data-to-insight-to-impact," statistical and ML techniques have considerable potential to address and mitigate yield issues in the semiconductor manufacturing domain.

XIII. CONCLUSION

Yield prediction is a challenging research domain in semiconductor manufacturing, with the rapid advancement of technology nodes and heterogeneous device integration. As advanced technology nodes shrink, feature design complexity and process step quantity increase, making manufacturing more challenging. These factors, combined with the requirements to improve yield and shorten time to market, result in increased interest in advanced statistical modeling, machine learning (ML), and deep learning (DL) approaches for yield prediction [1]. This study provides a comprehensive summary of current research on yield prediction approaches in semiconductor manufacturing and aims to assist researchers in orienting themselves in this dynamic research area. Key insights of existing yield prediction approaches are analyzed from a data-driven perspective, followed by theoretical insights and future work directions.

Several future directions for potential works are offered, including scrutinizing the combination of powerful feature extraction algorithms with new ML and DL models, mini/maxi modeling considering feature interactions, more spatial yield prediction and related works leveraging the recent achievements in graphs, developing an efficient approach to accommodate nonuniform process variations, and exploring robustness enhancement and explainability investigation on ML/DL models. The attempts outlined above are expected to advance approaches for yield prediction in semiconductor manufacturing, promote collaborative research between the semiconductor industry and academia, and further benefit related communities where research is still in its infancy. In the wake of the new AI paradigm, advancing yield prediction approaches with data analytics-based techniques to facilitate yield gap closure in semiconductor manufacturing has become an emerging demand.

REFERENCES

[1] 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).
[2] 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.
[3] 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.
[4] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.
[5] 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.
[6] 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
[7] 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
[8] 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.
[9] Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", International Journal of Science and Research (IJSR), Volume 9 Issue 12, December 2020, pp. 1899-1910, https://www.ijsr.net/getabstract.php?paperid=SR201212164722, DOI: https://www.doi.org/10.21275/SR201212164722
[10] 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
[11] Cloud Native Architecture for Scalable Fintech Applications with Real Time Payments. (2021). International Journal of Engineering and Computer Science, 10(12), 25501-25515. https://doi.org/10.18535/ijecs.v10i12.4654
[12] 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
[13] 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
[14] 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
[15] 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.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
IJIREEICE
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91217

[16] 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.
[17] Kannan, S., Gadi, A. L., Preethish Nanan, B., & Kommaragiri, V. B. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization.
[18] 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
[19] 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
[20] 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).
[21] Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 111–127. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11582
[22] 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
[23] 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
[24] Kommaragiri, V. B. (2021). Enhancing Telecom Security Through Big Data Analytics and Cloud-Based Threat Intelligence. Available at SSRN 5240140.
[25] 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
[26] 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
[27] Mandala, V. (2018). From Reactive to Proactive: Employing AI and ML in Automotive Brakes and Parking Systems to Enhance Road Safety. International Journal of Science and Research (IJSR), 7(11), 1992-1996.

