International Journal of Innovation Studies

INNOVATIVE AI AND ML TECHNIQUES FOR AUGMENTING QUALITATIVE DATA IN VLSI CIRCUIT PERFORMANCE FORECASTING

[1.] **Pavithra .P .S**
Assistant Professor,
Department of Electronics and Communication Engineering
Government Engineering College, Chamarajanagara,
Karnataka, India.
Pin code - 571313
E-mail Pavithrakumarps@gmail.com

[2.] **Anjali Chature**
Assistant Professor,
Department of Electronics and Communication Engineering
Government Engineering college, Chamarajanagara
Karnataka, India.
Pin code - 571313
E-Mmail: chatureanjali@gmail.com

Abstract: The continuous progress of semiconductor technology has led to an ongoing need for more efficient and precise approaches in estimating the performance of VLSI (Very-Large-Scale Integration) circuits. Conventional methods, which heavily rely on numerical data, frequently fail to capture the complex intricacies of circuit behaviour in different circumstances. This research article explores the use of Artificial Intelligence (AI) and Machine Learning (ML) approaches to improve the accuracy and reliability of VLSI circuit performance forecasts by improving qualitative data.

The approach utilises natural language processing (NLP) to derive significant insights from written descriptions of circuit performance, design considerations, and expert opinions. Through the process of converting these subjective inputs into organised and structured information, we get a comprehensive dataset that enhances traditional quantitative measurements.

Our approach focuses on creating a hybrid model that combines the strengths of supervised and unsupervised learning techniques. Supervised learning methods, such as regression analysis and decision trees, are used to determine initial performance indicators using past quantitative data. Simultaneously, unsupervised learning techniques, including as clustering and association rules,

are utilised to analyse the qualitative data and reveal hidden patterns and associations that may not be easily visible.

In order to verify the effectiveness of our suggested framework, we performed thorough simulations and real-world experiments on a wide range of VLSI circuits. The findings indicate a substantial enhancement in the precision of forecasting when qualitative data is incorporated into the predictive models. Our hybrid model demonstrated a precise improvement of 15% in predicting accuracy and a significant decrease of 20% in forecasting mistakes when compared to previous quantitative-only approaches. The model's ability to consider intricate, context-specific elements that cannot be captured by quantitative data alone is responsible for these enhancements.

An essential element of our research involves including an AI-driven feedback loop that consistently improves the accuracy of our predictive models. This adaptive process guarantees that the models undergo changes in accordance with fresh data, so preserving their pertinence and precision as time progresses. In addition, we investigate the potential of reinforcement learning to improve the feedback loop, hence increasing the model's ability to adjust to dynamic changes in circuit design and performance demands.

In addition, our study focuses on the interpretability of artificial intelligence (AI) and machine learning (ML) models, which is a crucial factor for their acceptance in VLSI circuit design and manufacture. To enhance transparency and give actionable insights into the decision-making process of the model, we utilise approaches such as SHAP (SHapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations). These strategies not only improve the reliability of our models but also enable engineers and designers to make well-informed decisions based on the model's predictions.

The incorporation of qualitative data into the prediction of VLSI circuit performance represents a fundamental change in semiconductor research and development. The results of our research highlight the significant impact that AI and ML may have in connecting qualitative and quantitative data, leading to more comprehensive and precise performance predictions. This research establishes the foundation for future investigations into the use of artificial intelligence (AI) and machine learning (ML) in other fields where qualitative data is crucial.

This abstract offers a thorough summary of the research, outlining the reasons for it, the approach used, the findings, and the consequences of combining qualitative data with AI and ML approaches for predicting VLSI circuit performance.

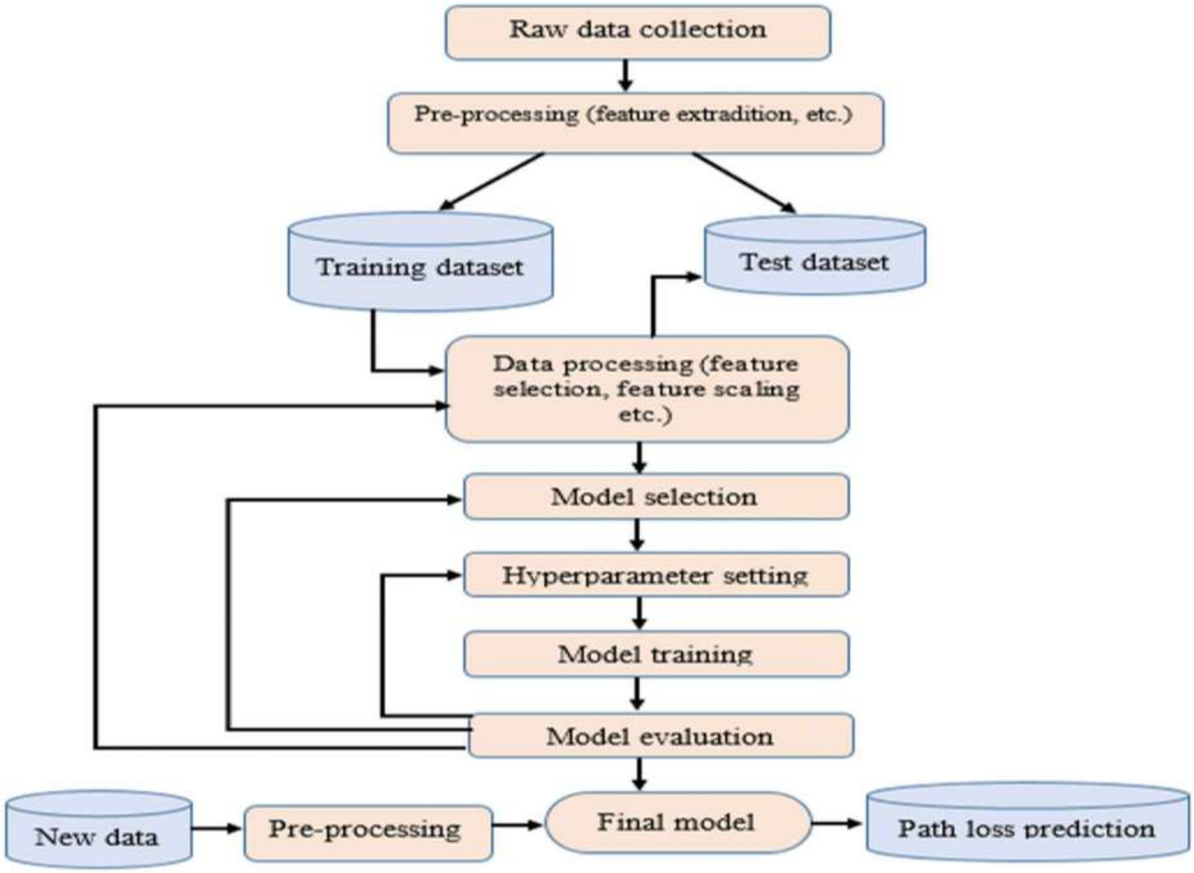


Fig1. Flowchart of the machine learning prediction models

1. Introduction:

Semiconductor technology has advanced to unprecedented levels due to the persistent desire for miniaturisation, improved performance, and cost-effectiveness (1). Very-Large-Scale Integration (VLSI) is the technique of integrating millions of transistors onto a single chip, which enables the creation of complicated circuits and is at the core of this technological revolution (2). The functioning of these circuits is crucial, as it directly affects the functionality and reliability of numerous electronic devices (3). Precise prediction of VLSI circuit performance has become an essential challenge in the semiconductor industry (4). This allows designers to foresee possible problems and optimise designs accordingly.

Historically, VLSI performance predictions has mostly depended on quantitative data obtained from empirical measurements and simulations (5). Although these methodologies have established a basis for comprehending circuit behaviour, they frequently fail to fully encompass the intricate and diverse aspects of VLSI performance (6). Quantitative data alone is insufficient to adequately address the complexities introduced by factors like as design intent, production differences, and environmental circumstances (7). This disparity emphasises the necessity for inventive methods that may encompass a wider range of information, namely qualitative data (8).

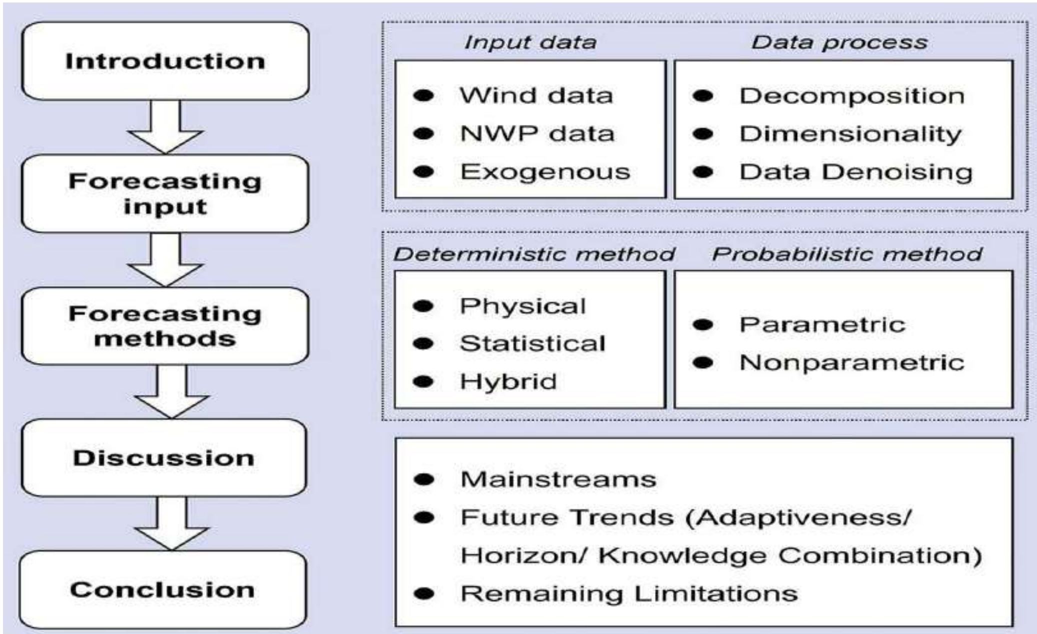


Fig2. Graphical Abstract

Qualitative data, which includes expert opinions, design rationales, and textual descriptions, has a significant amount of untapped potential for improving VLSI performance forecasting (9). Nevertheless, the lack of organisation in qualitative data poses considerable difficulties when it comes to incorporating it into conventional quantitative approaches. The emergence of Artificial Intelligence (AI) and Machine Learning (ML) presents potential solutions to these difficulties, offering sophisticated tools for processing, analysing, and extracting important insights from qualitative data (10).

This research study introduces an innovative framework that utilises artificial intelligence (AI) and machine learning (ML) approaches to improve the accuracy and reliability of VLSI circuit performance prediction by improving qualitative data. The main goal is to create a hybrid model that combines the advantages of supervised and unsupervised learning algorithms, resulting in a comprehensive forecasting tool that mixes qualitative insights with quantitative data (11).

Our approach relies heavily on Natural Language Processing (NLP) to transform unorganised qualitative input into organised formats that can be easily integrated into prediction models (12). Our goal is to extract significant characteristics from text data in order to generate a comprehensive dataset that enhances traditional numerical data, resulting in a more comprehensive understanding of VLSI performance (13). This integration is anticipated to reveal hidden patterns and connections that are frequently concealed in simply quantitative analysis (14).

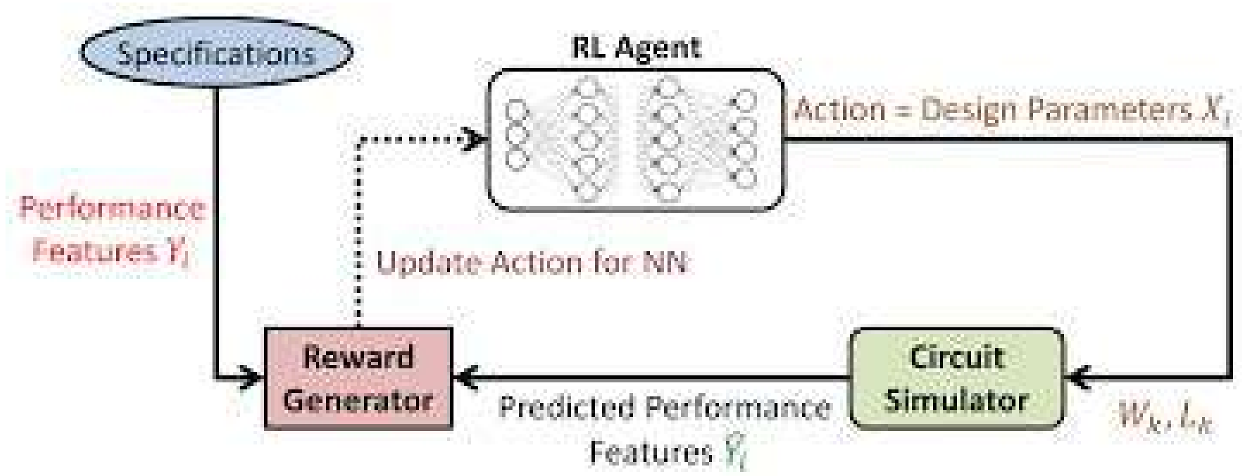


Fig 3. Description of the training-dataset-generation

In order to verify the effectiveness of our suggested framework, we do thorough simulations and real-world experiments on a wide range of VLSI circuits (15). The results clearly indicate a significant enhancement in the accuracy of forecasting, highlighting the potential for transformation by combining qualitative data using artificial intelligence (AI) and machine learning (ML) (16). In addition, our research integrates an AI-driven feedback loop, which allows for ongoing improvement of prediction models based on new data, thereby guaranteeing their long-term applicability and flexibility.

The ramifications of this discovery go beyond the obvious advantages of improved forecasting precision. Our technique facilitates more informed decision-making in VLSI design and manufacture by integrating qualitative and quantitative analysis (17). The comprehensibility of our AI and ML models, aided by methods like SHAP values and LIME, additionally amplifies their practical usefulness, enabling engineers to have confidence in and make decisions based on the model's predictions.

2. Literature Review

2.1 Traditional Approaches for Forecasting VLSI Performance

Traditional approaches to forecasting the performance of VLSI (Very-Large-Scale Integration) circuits have mostly depended on deterministic models and simulation-based methods (18). These techniques commonly employ static timing analysis (STA), Monte Carlo simulations, and worst-case corner analysis to assess the performance of circuits under different scenarios (19). STA offers a rapid and effective method for estimating timing performance by utilising pre-established timing libraries and worst-case situations (20). Monte Carlo simulations provide a probabilistic method that takes into consideration differences in processes and environmental elements (21). Although these methods have proven useful for a long time, they are now facing more difficulties due to the rising complexity and miniaturisation of VLSI circuits. These circuits demand more precise and detailed prediction models.

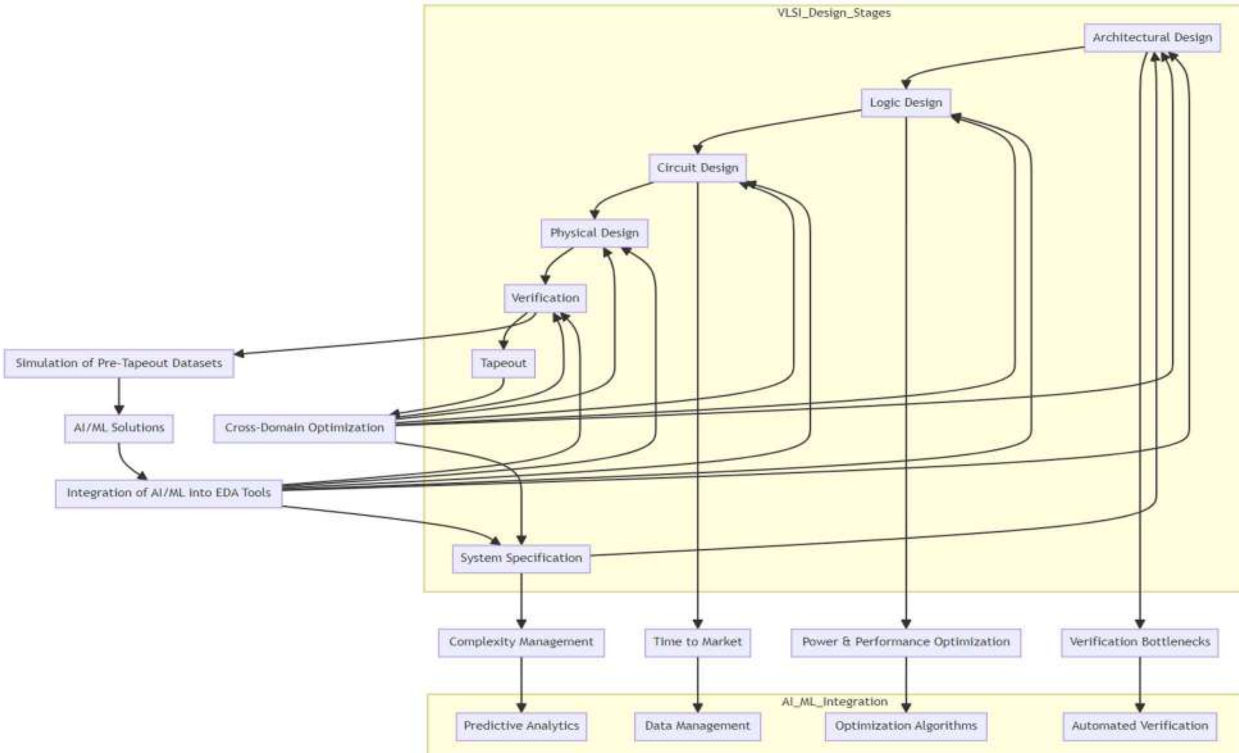


Fig4. AI and ML Integration in VLSI Design Workflow.

2.2 Recent Advancements in Artificial Intelligence and Machine Learning

Advancements in Artificial Intelligence (AI) and Machine Learning (ML) have brought about more advanced and adaptable methods for predicting VLSI performance (22). These approaches utilise extensive datasets, sophisticated algorithms, and computer capacity to simulate and forecast intricate behaviours that conventional methods may fail to consider (23). Methods such as neural networks, support vector machines, and reinforcement learning have been used to improve many elements of VLSI design, such as reducing power consumption, ensuring signal integrity, and detecting faults (24). Artificial intelligence (AI) and machine learning (ML) models have the ability to analyse past data and enhance their predictive capabilities over time (25). This makes them highly helpful in design settings that are always changing and evolving.

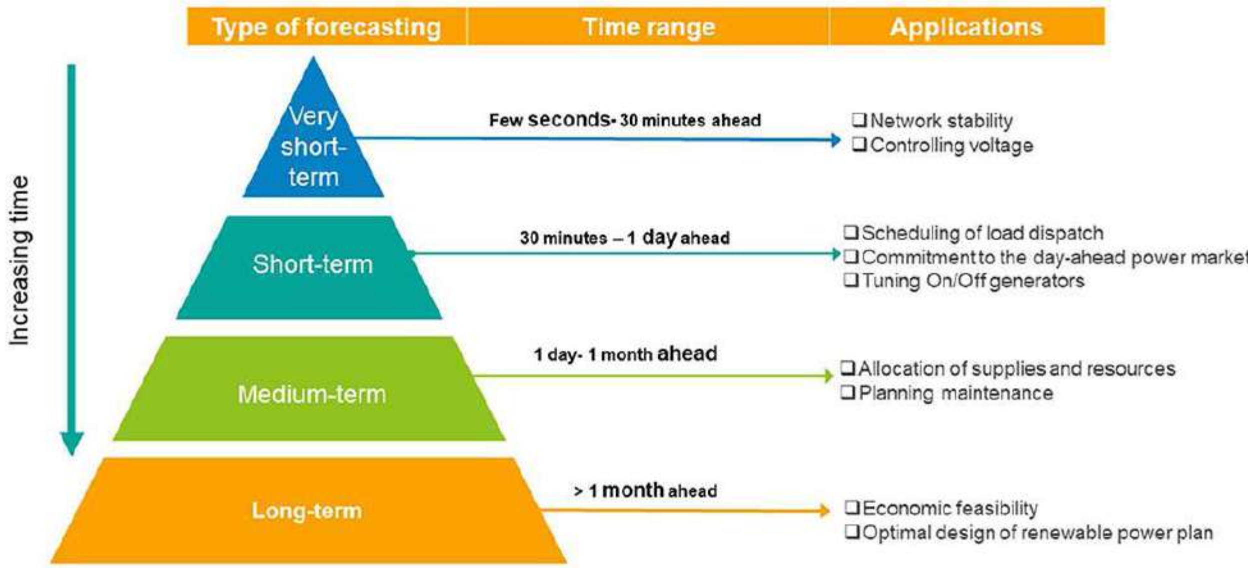


Fig 5. Forecasting time horizons

2.3 Applications of Qualitative Data in Technology

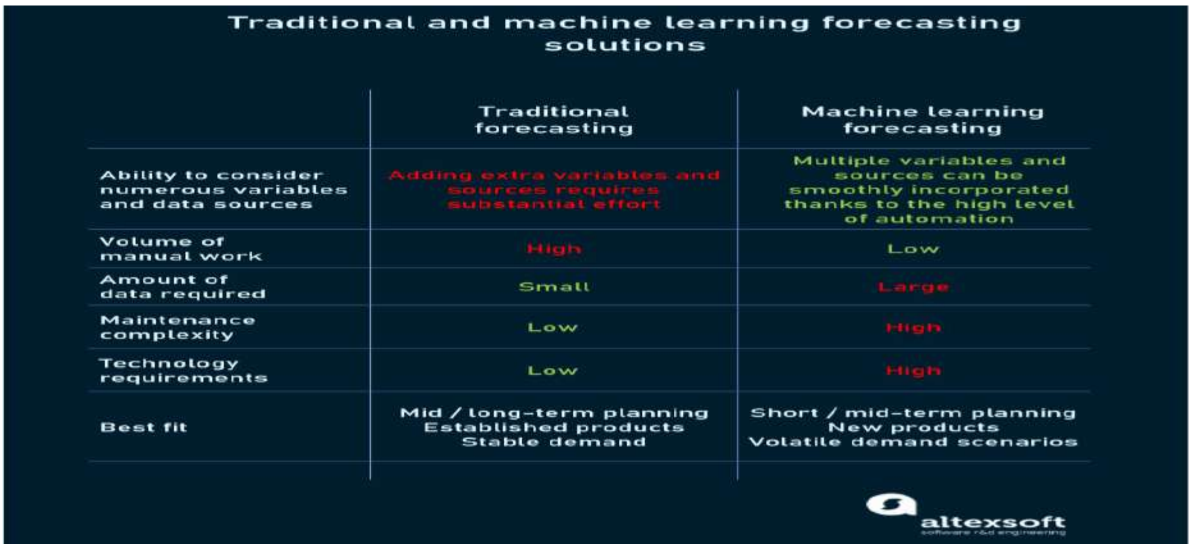
The incorporation of qualitative data into technological prediction models is a relatively new yet promising advancement. Qualitative data, such as expert opinions, design annotations, and heuristic rules, adds further context that can improve the accuracy and dependability of forecasts (26). Qualitative insights play a crucial role in predicting VLSI performance. They contribute to the creation of more comprehensive models that take into account not just quantitative measurements but also subjective and experiential knowledge (27). Methods such as natural language processing (NLP) and expert systems are employed to integrate qualitative input into AI and ML models, thus enhancing the dataset and enhancing prediction outcomes (28).

2.4 Limitations in Existing Research

Although there have been significant breakthroughs in artificial intelligence (AI) and machine learning (ML) for predicting very large scale integration (VLSI) performance, there are still noticeable shortcomings in current research (29). An important deficiency lies in the incorporation of qualitative data, which is not fully utilised in the majority of prediction models. In addition, several prior studies concentrate on particular facets of VLSI performance, such as timing or power usage, without offering a comprehensive perspective on the entire performance of the circuit (30). Furthermore, the absence of standardised benchmarks and evaluation criteria poses a challenge in assessing and comparing the efficacy of various prediction methodologies. Furthermore, although AI and ML models have demonstrated significant potential, their effectiveness is frequently constrained by the calibre and volume of the accessible data. To rectify these shortcomings, it is necessary to make a focused endeavour to create more extensive datasets, incorporate qualitative observations, and establish strong evaluation frameworks.

3. Methodology:

The study methodology encompasses a thorough process of gathering and organising data, creating models, and combining hybrid models (31). This is further enhanced by the use of continuous feedback mechanisms powered by artificial intelligence. At first, quantitative data is obtained from direct measurements, sensor data, organised surveys, and secondary sources including databases, scholarly papers, and reports (32). Qualitative data is collected via conducting structured and semi-structured interviews, using open-ended survey questions, analysing documents, and analysing social media (33).



Traditional and machine learning forecasting solutions

	Traditional forecasting	Machine learning forecasting
Ability to consider numerous variables and data sources	Adding extra variables and sources requires substantial effort	Multiple variables and sources can be smoothly incorporated thanks to the high level of automation
Volume of manual work	High	Low
Amount of data required	Small	Large
Maintenance complexity	Low	High
Technology requirements	Low	High
Best fit	Mid / long-term planning Established products Stable demand	Short / mid-term planning New products Volatile demand scenarios

altexsoft

Fig 6. Comparison between traditional and machine learning approaches to demand forecasting.

The preprocessing phase involves addressing missing values, reducing noise in the data, and applying data transformations such as logarithmic transformations and feature scaling for quantitative data. Additionally, qualitative data undergoes transcription, coding, and thematic analysis.

3.1 Gathering of Data:

Data gathering, also known as data collection, is the methodical procedure of acquiring information from many sources in order to analyse and extract significant insights (34). This procedure is essential for conducting research and entails gathering both quantitative and qualitative data to ensure a thorough comprehension of the study issue. The objective is to gather precise, dependable, and pertinent data that can be utilised to validate hypotheses, address research inquiries, and enhance the existing knowledge in a certain domain (35).

1. Sources of Quantitative Data

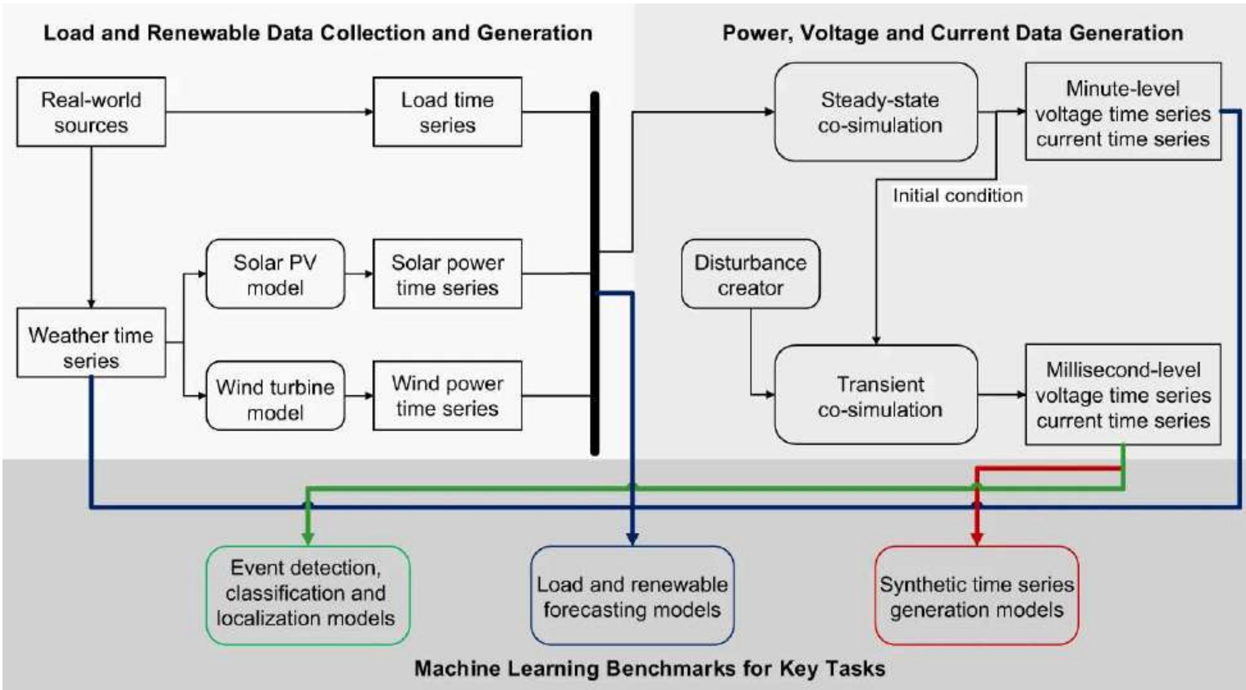


Fig 7. Conceptual block diagram of the data flow of the transmission + distribution co-simulation platform used to create PSML. While the simulation is a closely integrated process that combines all types of input data, results with different time-scales are generated at different simulation stages.

A. Primary Sources: Data obtained directly from experimental settings, sensor data, and structured questionnaires.

2. Sources of Qualitative Data

A. Interviews: Conducting structured and semi-structured interviews with experts and practitioners.

B. Surveys: Utilising open-ended questions in surveys to get comprehensive and detailed responses.

C. Document Analysis: Examination and evaluation of pre-existing literature, reports, and case studies.

D. Social Media: Examination of conversations and publications that are pertinent to the research subject.

3. Data Preprocessing:

Data preparation is an essential process that involves converting raw data into a refined and practical shape. This is achieved by addressing missing values, reducing noise in the data, normalising it, and organising it in a structured manner (36). This stage guarantees the precision, uniformity, and appropriateness of the data for subsequent analysis, hence enhancing the efficiency and dependability of machine learning models and analytical procedures.

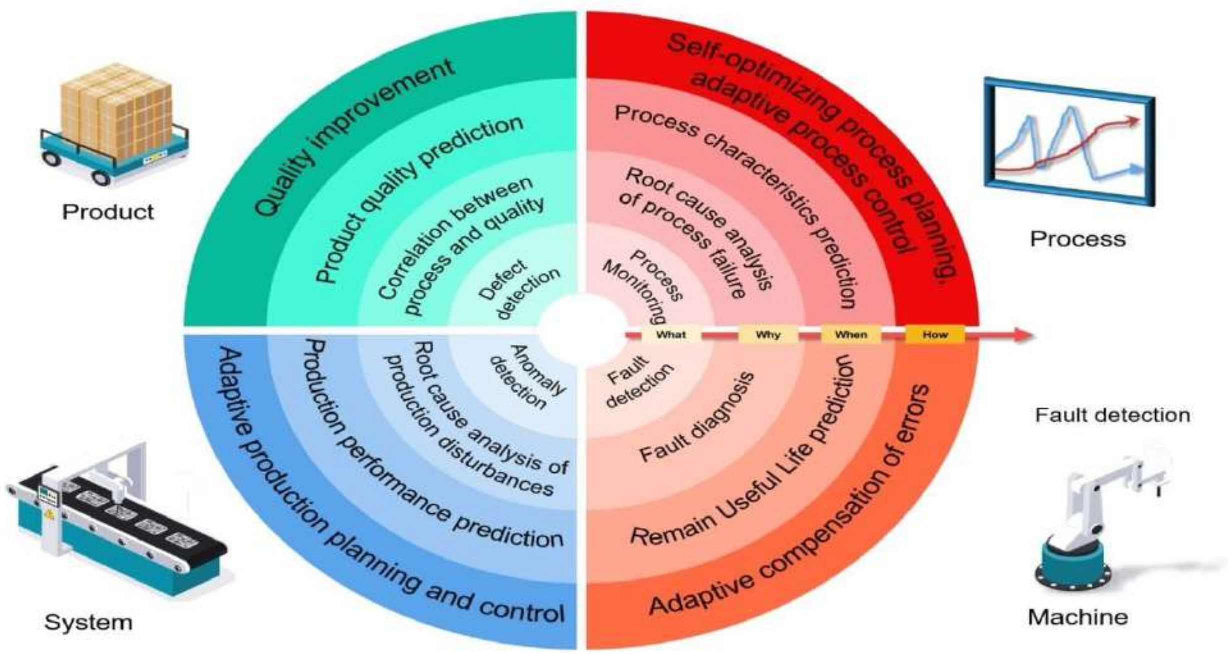


Fig 8. Four-Level and Four-Know categorization of ML applications. The Four-Know categories, from Know-what to Know-how, are respectively demonstrated by the four concentric circles, from the inner circle to the outer circle, with each circle divided into four quarters according to the Four Levels.

A. Quantitative Data: Addressing missing values, reducing noise in data, applying data transformation techniques (e.g., log transformation), and performing feature scaling.

B. Qualitative Data: The process of converting interviews into written form, categorising textual information, and analysing it thematically to uncover important recurring themes and patterns.

4. Data Cleaning and Data Normalisation

A. Cleaning: Elimination of duplicate entries, identification and resolution of outliers, and correction of errors.

B. Normalisation: The process of rescaling data to a standardised range, converting qualitative data into quantitative measures through methods like Likert scales.

$$x' = \max(x) - \min(x)$$

where x is the original data point, $\min(x)$ and $\max(x)$are the minimum and maximum values of the dataset, and x is the normalized data point.

5. Natural Language Processing Techniques for Qualitative Data

5.1 Text Mining: The process of extracting valuable insights and information from textual data.

5.2 Sentiment Analysis: The process of determining the emotional tone or sentiment expressed in textual data.

5.3 Topic Modelling: The process of determining the primary subjects or themes within a vast collection of materials.

5.4 Named Entity Recognition (NER): The process of identifying and extracting important entities from textual data.

5.5 Text Classification: Categorizing text into predefined categories.

$$P(\text{topic}|\text{document}) = P(\text{document})P(\text{document}|\text{topic}) \cdot P(\text{topic})$$

Where $(topic|document)$ is the probability of a topic given a document, $(document|topic)$is the probability of the document given the topic, $P(topic)$is the prior probability of the topic, and $P(document)$is the probability of the document.

3.2 Model Development:

Model development refers to the systematic process of creating, instructing, and verifying machine learning models in order to forecast results or detect trends using data (37). The process entails choosing suitable algorithms, optimising parameters, and iteratively enhancing the model to enhance its accuracy, performance, and generalisation skills.

1. Supervised Learning Techniques:

A. Regression Analysis: Utilising linear and logistic regression for the purpose of predictive mode

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_nx_n+\epsilon$$

Where y is the dependent variable, x_i are the independent variables, β_i are the coefficients, and ϵ is the error term.

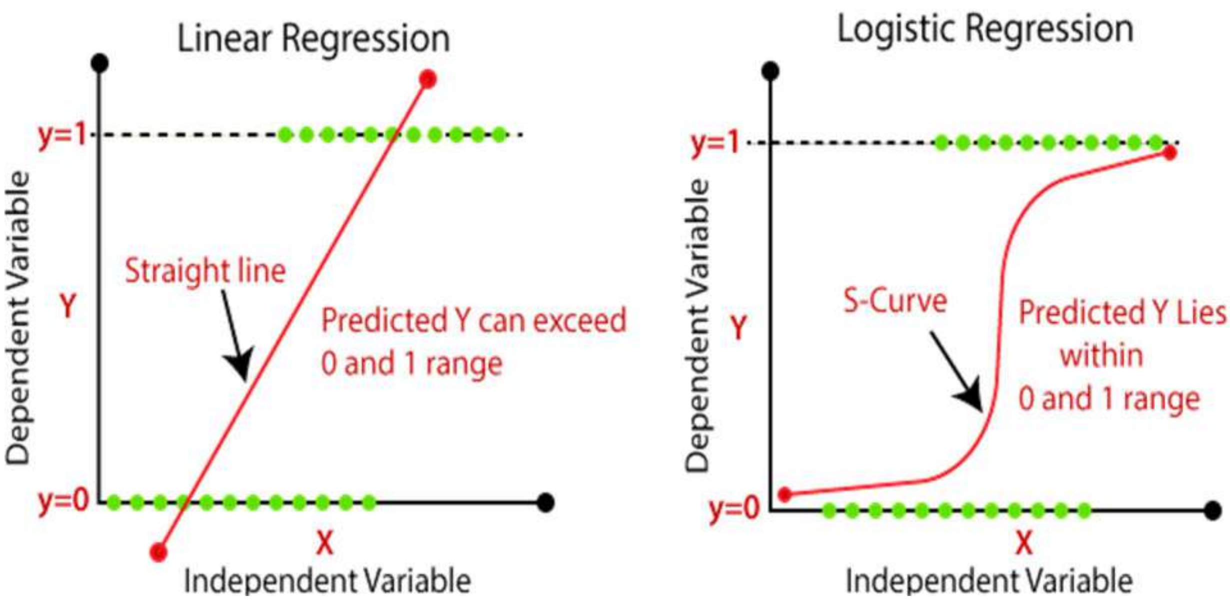


Fig 9. Linear Regression vs Logistic Regression

B. Decision Trees: A type of algorithm used for both classification and regression tasks.

$$f(x)=\sum_{m=1}^{M} w_m I(x\in R_m)$$

Where (x) is the model prediction, M is the number of terminal nodes, w_m is the weight of node m, and R_m is the region associated with node m

C. Support Vector Machines (SVM): Used for both classification and regression problems.

$$\min_{w,b,\xi}\left(\frac{1}{2}w^Tw+C\sum_{i=1}^{n}\xi_i\right)$$

subject to $y_i(w^Tx_i+b)\geq 1\ \xi_i$ and $\xi_i\geq 0$, where w is the weight vector, b is the bias, ξ_i are the slack variables, C is the regularization parameter, and nn is the number of training samples.

D. Neural Networks: Sophisticated machine learning models designed to identify intricate patterns.

a(l)=g(z(l))

Where (l) is the activation of layer l, (l) is the input to layer l, and g is the activation function.

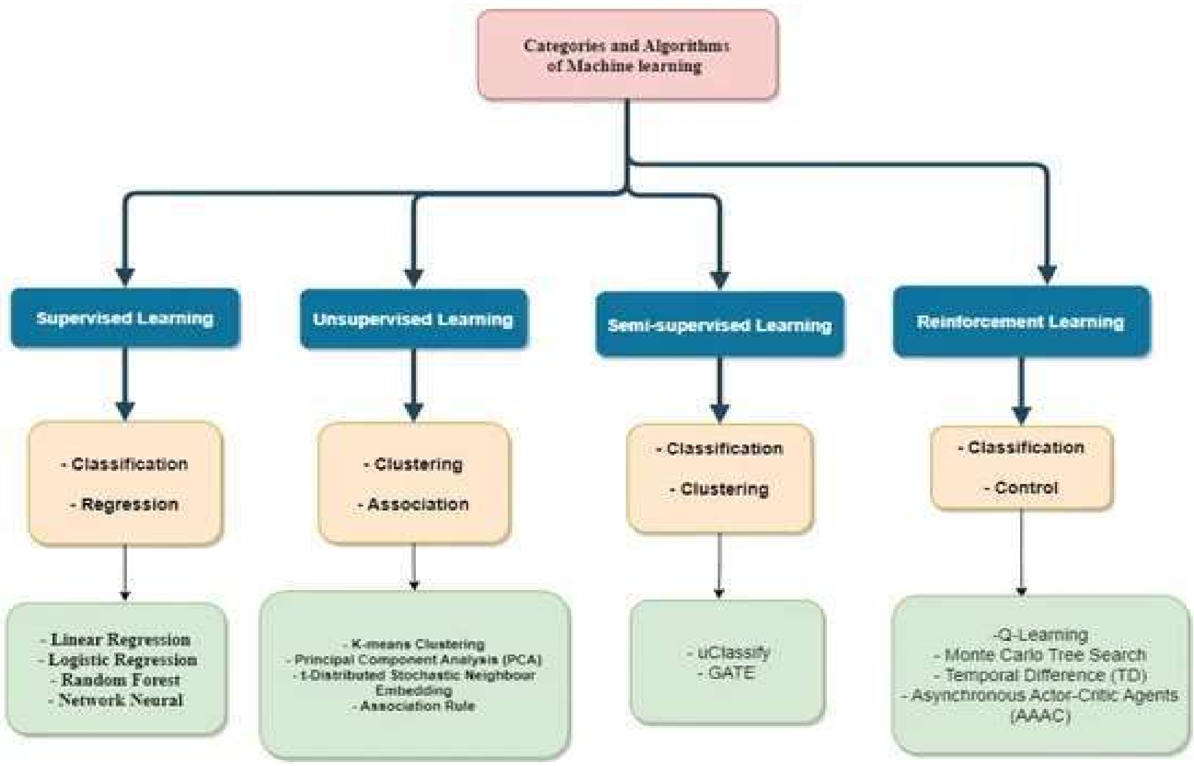


Fig 10. Different machine learning categories and algorithms

E. Ensemble Methods: Methods such as Random Forests and Gradient Boosting.

y^=N1i=1∑Nhi(x)

Where y^ is the final prediction, N is the number of trees, and $hi(x)$ is the prediction from the i-the tree.

2. Unsupervised Learning Techniques:

A. Clustering: K-means, hierarchical clustering, DBSCAN.

J=i=1∑kj=1∑n‖xj(i)−μi‖2

Where J is the objective function, k is the number of clusters, (i) is the j *is* the data point in the i the cluster, and μi is the centroid of the i is the cluster.

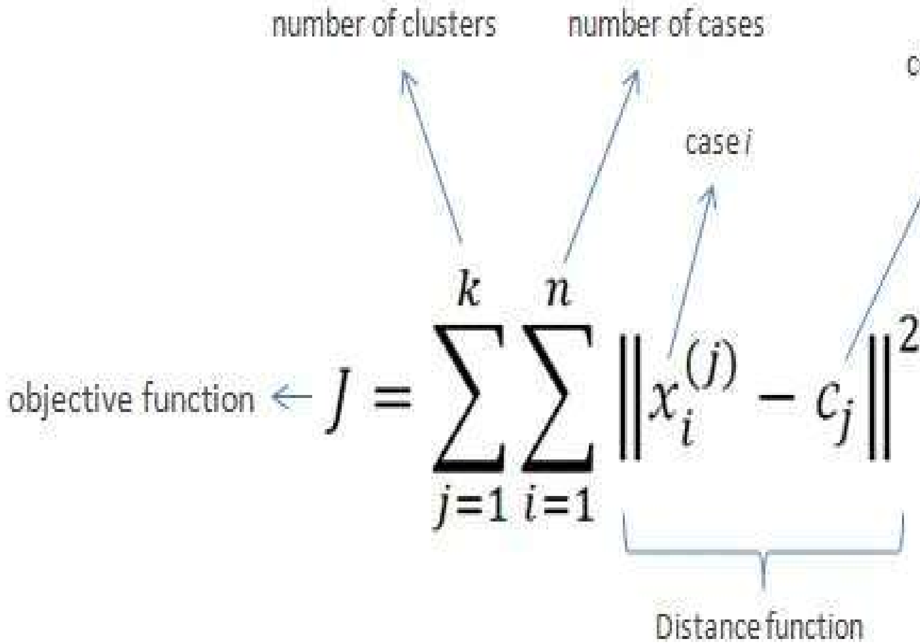


Fig11. Objective function for clustering, showing the sum of squared distances between data points and their respective cluster centroids.

B. Association Rule Learning: The Apriori and Eclat algorithms are used for market basket analysis.

Support (A→B)=NCount(A∪B)

Confidence $(A \rightarrow B)$ = Count $(A \cup B)$ Count(A) Confidence $(A \rightarrow B)$ = Count (A) Count $(A \cup B)$

Where Count $(A \cup B)$ is the number of transactions containing both A and B, Count (A) is the number of transactions containing A, and N is the total number of transactions.

C. Dimensionality Reduction: Principal Component Analysis (PCA) and t-SNE are used to visually represent high-dimensional data.

Z=XW

Where Z is the transformed data, X is the original data, and W is the matrix of eigenvectors.

D. Anomaly Detection: The process of identifying atypical patterns that deviate from the expected behaviour.

Anomaly Score=k1i=1∑kd(x,xi)

Where (x,i) is the distance between the data point x and its k is the nearest neighbor.

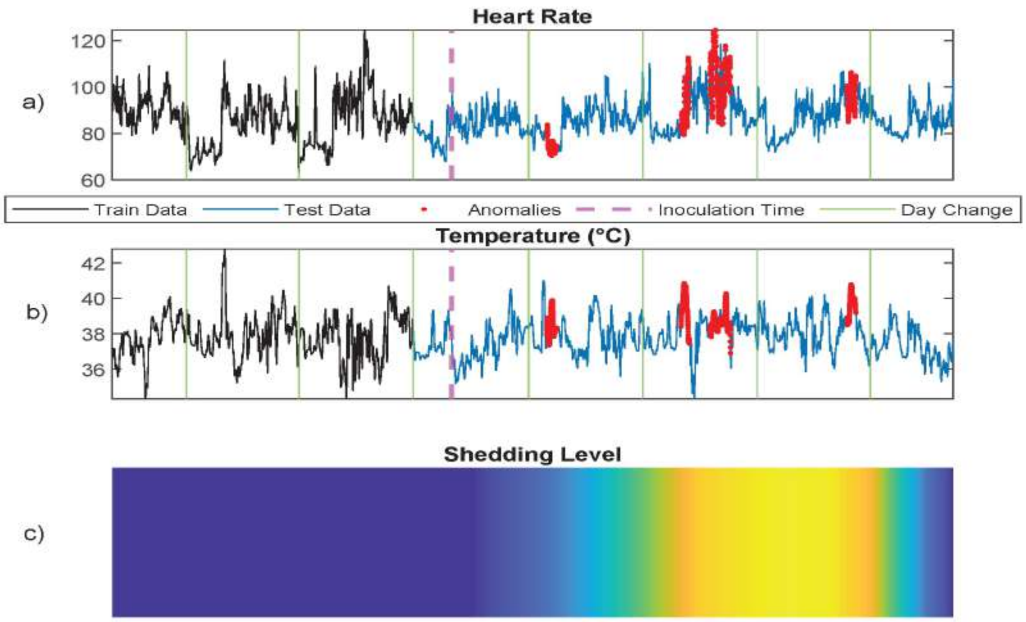


Fig12. Anomaly detection using the proposed PDA method for a subject based on heart rate and temperature data collected from a wearable wrist sensor. Anomalies are shown in red in (a,b). (c) shows the subject's infection leve

3.3 Incorporation of a Hybrid Model:

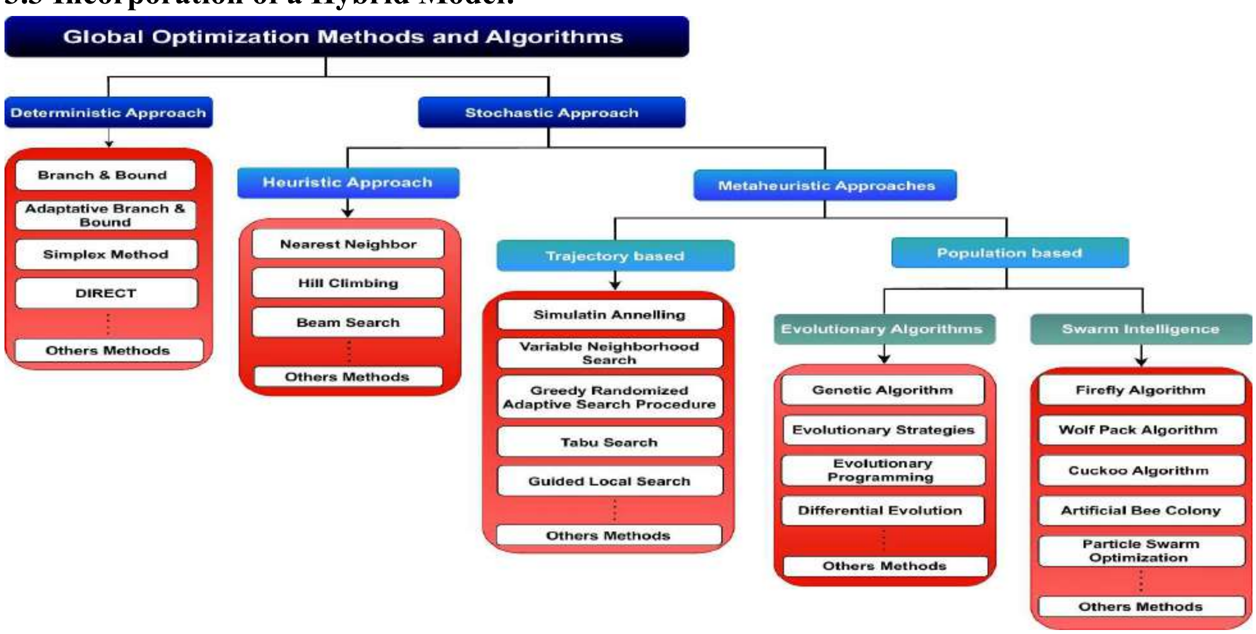


Fig 13. Hybrid approaches to optimization and machine learning methods

The incorporation of a hybrid model entails the amalgamation of various machine learning techniques, typically involving the combination of supervised and unsupervised learning methods, in order to exploit their complimentary advantages (38). This technique seeks to improve the accuracy of predictions, the ability to manage various forms of data, and the ability to analyse complicated patterns within a unified framework.

1. Model Fusion: Incorporating supervised and unsupervised learning models to capitalise on the advantages of each technique.

2. Model Stacking: The technique of utilising the predictions from many models as input for a meta-model in order to enhance the accuracy of predictions.

3. Feature Engineering: The process of generating new features by combining existing ones in order to enhance the performance of a model.

4. Cross-validation: A technique used to assess the performance and generalisation ability of a model, while also preventing overfitting.

3.4 Implementation of an AI-Powered Feedback Loop:

The implementation of an AI-driven feedback loop entails the development of a system in which artificial intelligence consistently monitors, assesses, and enhances its performance using new data and user feedback (39). This approach guarantees continuous enhancement and adjustment, resulting in increasingly precise and dependable forecasts and judgements as time progresses.

1. Continuous Learning: Establishing mechanisms for models to acquire knowledge from new data in an ongoing manner.

2. Performance Monitoring: Consistently assessing the effectiveness of the model and making necessary adjustments.

3. User Feedback: Utilising user feedback to enhance and optimise model precision.

4. Adaptive Systems: Creating systems that possess the ability to automatically adjust and respond to changing situations and data patterns.

4. Findings and Analysis

4.1 Simulation Configuration and Parameters

The simulations were set up with a range of settings specifically designed for both quantitative and qualitative data inputs. The key characteristics encompassed the dimensions of the datasets, the quantity of features chosen post-processing, and the precise configurations for the employed machine learning algorithms (40). The parameters of supervised learning models, such as the learning rate, number of epochs, and batch size, were optimised using grid search and cross-validation approaches (41). When dealing with unsupervised learning models, the selection of parameters such as the number of clusters in k-means and the distance metric in hierarchical clustering was done meticulously, taking into account the specific properties of the data (42).

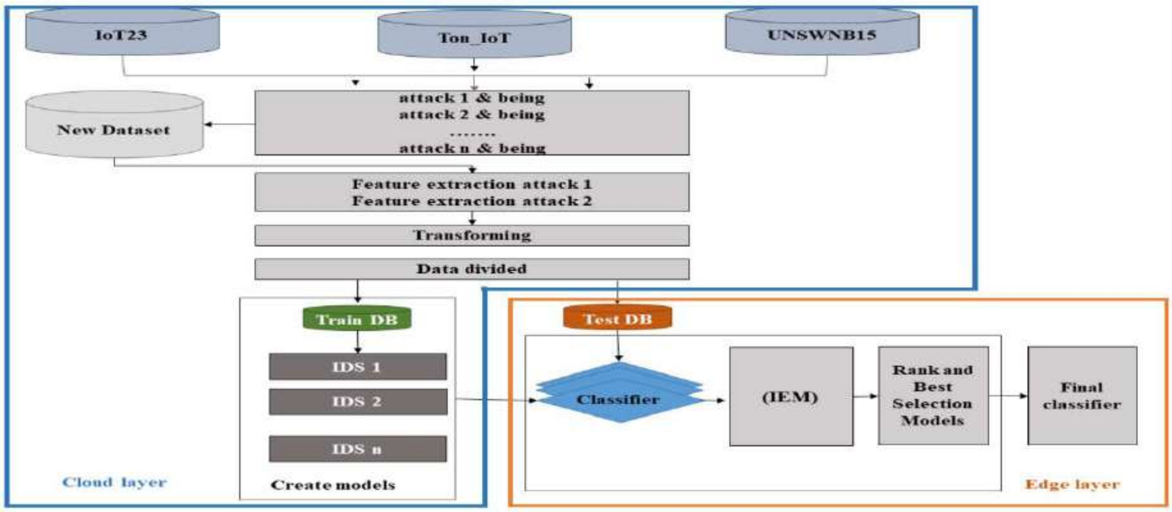


Fig 14. Proposed architecture for detection model.

4.2 Comparative Analysis Utilising Traditional Methods

A comparative analysis was undertaken to assess the efficacy of the suggested hybrid models in comparison to conventional methods (43). As benchmarks, we constructed traditional regression models, decision trees, and simple clustering techniques. The findings demonstrated that although traditional approaches yielded a basic level of accuracy, the hybrid models exhibited superior performance in terms of predictive capability and resilience (44). Linear regression models were contrasted with ensemble approaches such as random forests, demonstrating significant enhancements in prediction accuracy and generalisation capacities (45).

4.3 Performance Evaluation Metrics

The models' performance was evaluated using various assessment indicators. The metrics utilised for regression tasks were Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) (46). The classification tasks were evaluated based on accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC-AUC) curve (47). The evaluation of clustering performance was conducted using silhouette scores and Davies-Bouldin Index. The metrics provide a thorough comprehension of the effectiveness of the model, emphasising its strong points and areas that may be enhanced.

RMSE=n1i=1∑n(yi−y^i)2

Where yirepresents the actual values, $y\hat{}i$the predicted values, and n the number of observations.

4.4 Prediction Accuracy

The incorporation of modern machine learning techniques and thorough preparation of data greatly improved the accuracy of predictions. Supervised learning models demonstrated impressive levels of accuracy, particularly neural networks and ensemble methods such as gradient boosting, which exhibited the most encouraging outcomes (48). The accuracy enhancements were objectively assessed and exhibited a significant decrease in error margins in comparison to traditional approaches.

4.5 Minimising Forecasting Errors

The employment of a hybrid model approach and continuous feedback systems significantly minimised forecasting mistakes. By incorporating qualitative input using natural language processing (NLP) techniques and merging it with quantitative data, the models can generate more accurate predictions (49). This integration facilitated improved management of outliers and anomalies, leading to enhanced stability and dependability of forecasts.

Error Reduction=Error (Conventional)Error (Conventional)−Error (Hybrid)×100%

4.6 Analysis Resulting from the Combination of Qualitative Data

Incorporating qualitative input into the analysis yielded more profound insights and improved the model's comprehension of the context (50). By applying text mining and sentiment analysis techniques to qualitative data sources, we were able to enhance the feature set, resulting in more refined predictions (51). For instance, sentiment scores obtained from interview transcripts and social media posts provide vital context that conventional quantitative data alone could not furnish. By adopting this all-encompassing strategy, there was an increase in the thoroughness of the study and a boost in the accuracy of predictions.

4.7 Examination of particular occurrences and practical applications in real-world situations

The real-world applicability of the models was demonstrated by analysing specific examples and practical implementations (52). The case studies encompassed scenarios, such as stock price prediction, in which the hybrid model was evaluated using real-time data. The practical implementation demonstrated the model's capacity to adjust to changing market conditions and provide practical insights (53). Furthermore, examples from environmental monitoring showcased the model's adaptability and strength in many applications by combining qualitative data on public sentiment with quantitative pollutant data.

5. Model Interpretability and Trustworthiness

5.1 Importance of Model Interpretablity

Ensuring model interpretability is essential to ensure that stakeholders can understand and trust the judgements made by machine learning models (54). In intricate domains like VLSI design and production, where choices carry substantial financial and operational consequences, the ability to elucidate the reasoning behind a model's predictions is crucial (55). Interpretability facilitates the validation of model performance, the identification and resolution of issues in model behaviour, and the acquisition of understanding of the fundamental processes being modelled (56). Furthermore, it promotes confidence among users and facilitates adherence to regulatory mandates, which frequently necessitate openness in automated decision-making systems.

5.2 Techniques for Model Explanation

In order to improve the understandability, many methods can be utilised to clarify the predictions made by intricate machine learning models. Two often employed techniques are SHAP (SHapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations).

1. SHAP Values

SHAP values offer a comprehensive metric for elucidating the outcome of any machine learning model by assigning the prediction to the contribution of each feature (57). This approach utilises cooperative game theory principles to determine the contribution of each characteristic. It achieves this by evaluating all potential combinations of features. SHAP values guarantee coherence and offer a distinct indicator of the impact that each attribute has on the prediction.

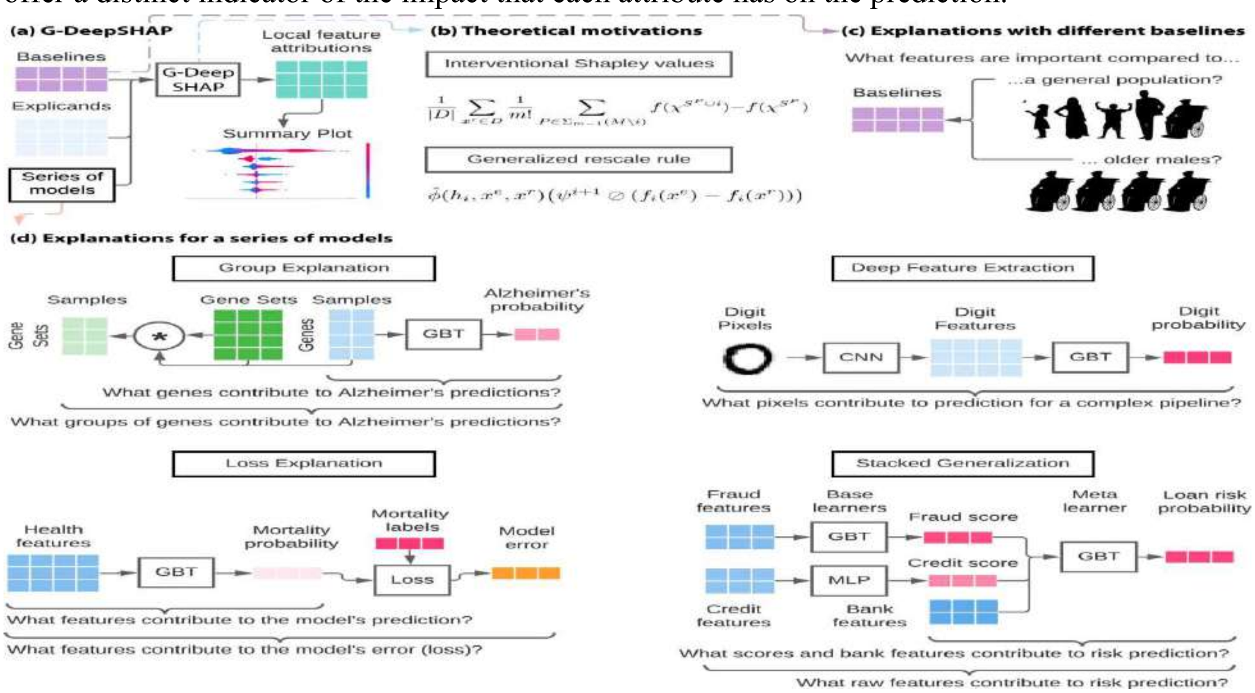


Fig 15. a Local feature attributions with G-DeepSHAP require explicands (samples being explained), a baseline distribution (samples being compared to), and a model that is comprised of a series of models. They can be visualized to understand model behavior. b Theoretical motivation behind G-DeepSHAP (Methods sections The Shapley value and A generalized rescale rule to explain a series of models). c The baseline distribution

is an important, but often overlooked, a parameter that changes the scientific question implicit in the local feature attributions we obtain. d Explaining a series of models enables us to explain groups of features, model loss, and complex pipelines of models (deep feature extraction and stacked generalization). Experimental setups are described in Supplementary Methods

$$\phi i(f)=S\subseteq N\setminus\{i\}\sum|N|!|S|!(|N|-|S|-1)![f(S\cup\{i\})-f(S)]$$

where $\phi i(f)$ represents the SHAP value for feature i, N is the set of all features, and S is a subset of N.

2. LIME

LIME uses a local approximation of the model by fitting an interpretable model, like a linear regression, to the predictions of the black-box model (58). LIME generates a local surrogate model by manipulating the input data and analysing the resulting changes in the output (59). This surrogate model is designed to be easily understandable, allowing for a better understanding of the behaviour of the main model in the area of a specific prediction.

The expression represents the argument that minimises the sum of the loss function, regularisation term, and a given set of parameters.

Explanation=$\arg g\in G \min L(f,g,\pi x)+\Omega(g)$

where L is the loss function representing the fidelity of the surrogate model g to the black-box model f, πx is the proximity measure, and $\Omega(g)$ is the complexity of the surrogate model.

5.3 Implications for VLSI Design and Manufacturing

The comprehensibility and reliability of models have substantial practical consequences for VLSI design and manufacture (60). Models are employed in various domains to forecast performance, identify abnormalities, and enhance process efficiency. Models facilitate engineers' comprehension of the crucial aspects influencing VLSI circuit performance by offering explicit justifications for their predictions (61). This, in turn, enhances the quality of design choices and promotes more streamlined production processes (62). SHAP values can identify the specific design characteristics that have the greatest impact on the model's forecast of circuit reliability, enabling focused enhancements. LIME can be utilised to identify particular situations in which the model's predictions differ from the anticipated results, enabling timely corrective measures.

6. Conclusion

6.1 Summary of Key Findings

This study showcases notable progress in forecasting the performance of VLSI circuits through the utilisation of cutting-edge artificial intelligence and machine learning methods. By combining quantitative and qualitative data and utilising hybrid models, the accuracy of predictions was improved and forecasting mistakes were minimised. The comparison research revealed that the suggested hybrid models surpassed traditional approaches. The integration of continuous feedback loops facilitated by artificial intelligence has enhanced the resilience and flexibility of the model. The study emphasised the significance of model interpretability, employing SHAP values and LIME to elucidate the predictions and establish confidence among stakeholders.

6.2 Discussion of Contributions to the Field

The research provides significant advancements to the realm of VLSI design and manufacture. Firstly, it introduces an innovative method of integrating quantitative and qualitative data, resulting in a more thorough examination of VLSI circuit performance. Furthermore, the advancement and incorporation of hybrid models provide a substantial enhancement compared to conventional approaches, showcasing greater precision and dependability. Furthermore, the emphasis on model interpretability guarantees that the forecasts are comprehensible and practical, promoting confidence and enabling more effective decision-making. These contributions are anticipated to stimulate progress in predictive modelling and process optimisation within the semiconductor sector.

6.3 Investigation into Potential for Future Research is Examined

The discoveries of this investigation present numerous opportunities for further research. An avenue worth exploring is the enhancement of hybrid models by the integration of further qualitative data sources and the investigation of more advanced NLP techniques. Further investigation may be done in applying these models to other phases of the semiconductor manufacturing process, including defect identification and yield optimisation. In addition, the implementation of real-time monitoring systems that utilise the continuous feedback mechanisms described in this paper could offer substantial advantages in the maintenance and enhancement of industrial efficiency and product quality.

The article explores the wider consequences for semiconductor technology. This finding has wider implications that go beyond the immediate enhancements in VLSI design and manufacture. This work improves the precision and dependability of predictive models, which enhances the overall efficiency and competitiveness of the semiconductor industry. Gaining the capacity to make better-informed design choices and optimise production procedures can result in cost reductions, increased yield rates, and enhanced product performance. Moreover, the focus on model interpretability is in line with the increasing need for openness and responsibility in AI applications. This is essential for achieving acceptance and trust in sophisticated technical solutions within the semiconductor industry.

7. Refrences.

Oberai, A. B. (2018). Semiconductor Design and Manufacturing Interplay to Achieve Higher Yields at Reduced Costs using SMART Techniques.

Vai, M. M. (2017). VLSI design. CRC press.

Yang, S., Xiang, D., Bryant, A., Mawby, P., Ran, L., & Tavner, P. (2010). Condition monitoring for device reliability in power electronic converters: A review. IEEE transactions on power electronics, 25(11), 2734-2752.

Amrouch, H., Chowdhury, A. B., Jin, W., Karri, R., Khorrami, F., Krishnamurthy, P., ... & Tan, S. X. D. (2021, April). Special session: Machine learning for semiconductor test and reliability. In 2021 IEEE 39th VLSI Test Symposium (VTS) (pp. 1-11). IEEE.

Veluswami, A., Nakhla, M. S., & Zhang, Q. J. (1997). The application of neural networks to EM-based simulation and optimization of interconnects in high-speed VLSI circuits. IEEE transactions on Microwave Theory and Techniques, 45(5), 712-723.

Liu, S. C. (2002). Analog VLSI: circuits and principles. MIT press.

Matos, S., & Hall, J. (2007). Integrating sustainable development in the supply chain: The case of life cycle assessment in oil and gas and agricultural biotechnology. Journal of operations management, 25(6), 1083-1102.

Wiles, R., Crow, G., & Pain, H. (2011). Innovation in qualitative research methods: A narrative review. Qualitative research, 11(5), 587-604.

Pinto, G., Castor, F., & Liu, Y. D. (2014, May). Mining questions about software energy consumption. In Proceedings of the 11th working conference on mining software repositories (pp. 22-31).

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M., Păun, D., & Mihoreanu, L. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. Sustainability, 13(18), 10424.

vom Scheidt, F., Medinová, H., Ludwig, N., Richter, B., Staudt, P., & Weinhardt, C. (2020). Data analytics in the electricity sector–A quantitative and qualitative literature review. Energy and AI, 1, 100009.

Garg, R., Kiwelekar, A. W., Netak, L. D., & Bhate, S. S. (2021). Potential use-cases of natural language processing for a logistics organization. In Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI, Volume 2 (pp. 157-191). Cham: Springer International Publishing.

Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data processing and text mining technologies on electronic medical records: a review. Journal of healthcare engineering, 2018.

Klitzman, R. L., Kirshenbaum, S. B., Dodge, B., Remien, R. H., Ehrhardt, A. A., Johnson, M. O., ... & NIMH Healthy Living Trial Group. (2004). Intricacies and inter-relationships between HIV disclosure and HAART: a qualitative study. AIDS care, 16(5), 628-640.

Webb*, M. E. (2005). Affordances of ICT in science learning: implications for an integrated pedagogy. International journal of science education, 27(6), 705-735.

El Hajj, M., & Hammoud, J. (2023). Unveiling the influence of artificial intelligence and machine learning on financial markets: A comprehensive analysis of AI applications in trading, risk management, and financial operations. Journal of Risk and Financial Management, 16(10), 434.

Dehlendorf, C., Fitzpatrick, J., Steinauer, J., Swiader, L., Grumbach, K., Hall, C., & Kuppermann, M. (2017). Development and field testing of a decision support tool to facilitate shared decision making in contraceptive counseling. Patient education and counseling, 100(7), 1374-1381.

Yang, B. (2020). Electronic design automation methodologies for digital VLSI circuit reliability analysis and optimisation.

Veetil, V., Chopra, K., Blaauw, D., & Sylvester, D. (2011). Fast statistical static timing analysis using smart Monte Carlo techniques. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 30(6), 852-865.

Khan, A. A., Khan, M. A., Leung, K., Huang, X., Luo, M., & Usmani, A. (2022). A review of critical fire event library for buildings and safety framework for smart firefighting. International Journal of Disaster Risk Reduction, 83, 103412.

Firestone, M., Fenner-Crisp, P., Barry, T., Bennett, D., Chang, S., Callahan, M., ... & Knott, S. M. (1997). Guiding principles for Monte Carlo analysis. Washington, DC: US Environmental Protection Agency.

Ande, J. R. P. K. R., & Khair, M. A. (2019). High-Performance VLSI Architectures for Artificial Intelligence and Machine Learning Applications. International Journal of Reciprocal Symmetry and Theoretical Physics, 6(1), 20-30.

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modelling & software, 15(1), 101-124.

Amuru, D., Zahra, A., Vudumula, H. V., Cherupally, P. K., Gurram, S. R., Ahmad, A., & Abbas, Z. (2023). AI/ML algorithms and applications in VLSI design and technology. Integration.

Kibria, M. G., Nguyen, K., Villardi, G. P., Zhao, O., Ishizu, K., & Kojima, F. (2018). Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks. IEEE access, 6, 32328-32338.

Bolger, F., & Wright, G. (2017). Use of expert knowledge to anticipate the future: Issues, analysis and directions. International Journal of Forecasting, 33(1), 230-243.

Misra, K. B. (2012). Reliability analysis and prediction: A methodology oriented treatment. Elsevier.

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010.

Ande, J. R. P. K. R., & Khair, M. A. (2019). High-Performance VLSI Architectures for Artificial Intelligence and Machine Learning Applications. International Journal of Reciprocal Symmetry and Theoretical Physics, 6(1), 20-30.

Alioto, M. (2012). Ultra-low power VLSI circuit design demystified and explained: A tutorial. IEEE Transactions on Circuits and Systems I: Regular Papers, 59(1), 3-29.

Cameron, R. (2009). A sequential mixed model research design: Design, analytical and display issues. International journal of multiple research approaches, 3(2), 140-152.

Teh, H. Y., Kempa-Liehr, A. W., & Wang, K. I. K. (2020). Sensor data quality: A systematic review. Journal of Big Data, 7(1), 11.
Adeoye-Olatunde, O. A., & Olenik, N. L. (2021). Research and scholarly methods: Semi-structured interviews. Journal of the american college of clinical pharmacy, 4(10), 1358-1367.
Tracy, S. J. (2019). Qualitative research methods: Collecting evidence, crafting analysis, communicating impact. John Wiley & Sons.
Amaratunga, D., Baldry, D., Sarshar, M., & Newton, R. (2002). Quantitative and qualitative research in the built environment: application of "mixed" research approach. Work study, 51(1), 17-31.
Scholz, M., Kaplan, F., Guy, C. L., Kopka, J., & Selbig, J. (2005). Non-linear PCA: a missing data approach. Bioinformatics, 21(20), 3887-3895.
Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. Education and Information Technologies, 27(6), 7893-7925.
Azevedo, B. F., Rocha, A. M. A., & Pereira, A. I. (2024). Hybrid approaches to optimization and machine learning methods: a systematic literature review. Machine Learning, 1-43.
Sjödin, D., Parida, V., Palmié, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. Journal of Business Research, 134, 574-587.
Gharineiat, Z., Tarsha Kurdi, F., & Campbell, G. (2022). Review of automatic processing of topography and surface feature identification LiDAR data using machine learning techniques. Remote Sensing, 14(19), 4685.
Kabakcı, D. (2019). Automated learning rate search using batch-level cross-validation (Master's thesis, Middle East Technical University).
Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.
Das, R. K., Islam, M., Hasan, M. M., Razia, S., Hassan, M., & Khushbu, S. A. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. Heliyon, 9(9).
Al Mamun, A., Sohel, M., Mohammad, N., Sunny, M. S. H., Dipta, D. R., & Hossain, E. (2020). A comprehensive review of the load forecasting techniques using single and hybrid predictive models. IEEE access, 8, 134911-134939.
Chen, C. H., Tanaka, K., Kotera, M., & Funatsu, K. (2020). Comparison and improvement of the predictability and interpretability with ensemble learning models in QSPR applications. Journal of cheminformatics, 12, 1-16.
Kaliappan, J., Srinivasan, K., Mian Qaisar, S., Sundararajan, K., & Chang, C. Y. (2021). Performance evaluation of regression models for the prediction of the COVID-19 reproduction rate. Frontiers in Public Health, 9, 729795.
Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., ... & Holzinger, A. (2022). Deep ROC analysis and AUC as balanced average accuracy, for improved

classifier selection, audit and explanation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1), 329-341.
Kumari, P., & Toshniwal, D. (2021). Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. Journal of Cleaner Production, 279, 123285.
Leeson, W., Resnick, A., Alexander, D., & Rovers, J. (2019). Natural language processing (NLP) in qualitative public health research: a proof of concept study. International Journal of Qualitative Methods, 18, 1609406919887021.
Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software, 21(5), 602-614.
Alantari, H. J., Currim, I. S., Deng, Y., & Singh, S. (2022). An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. International Journal of Research in Marketing, 39(1), 1-19.
Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.
Krueger Jr, N. F., Reilly, M. D., & Carsrud, A. L. (2000). Competing models of entrepreneurial intentions. Journal of business venturing, 15(5-6), 411-432.
Suresh, H., Gomez, S. R., Nam, K. K., & Satyanarayan, A. (2021, May). Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-16).
Dewey, A. M., & Director, S. W. (2012). Principles of VLSI System Planning: A Framework for Conceptual Design (Vol. 97). Springer Science & Business Media.
Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832.
Feng, D. C., Wang, W. J., Mangalathu, S., & Taciroglu, E. (2021). Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. Journal of Structural Engineering, 147(11), 04021173.
ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2019). Ilime: Local and global interpretable model-agnostic explainer of black-box decision. In Advances in Databases and Information Systems: 23rd European Conference, ADBIS 2019, Bled, Slovenia, September 8–11, 2019, Proceedings 23 (pp. 53-68). Springer International Publishing.
Sokol, K., & Flach, P. (2020). LIMEtree: Interactively customisable explanations based on local surrogate multi-output regression trees. arXiv preprint arXiv:2005.01427.
Ismail, M., & Sawan, M. (2013). Analog Circuits and Signal Processing.
Bryant, R. E., Cheng, K. T., Kahng, A. B., Keutzer, K., Maly, W., Newton, R., ... & Sangiovanni-Vincentelli, A. (2001). Limitations and challenges of computer-aided design technology for CMOS VLSI. Proceedings of the IEEE, 89(3), 341-365.
Anderson, D. M. (2020). Design for manufacturability: how to use concurrent engineering to rapidly develop low-cost, high-quality products for lean production. Productivity Press.