

IEEE 2025-26 PROJECT LIST	
Machine learning	
CODE	TITLE AND ABSTRACT
26-ANSP-ML-001	A Hybrid Machine Learning Model for Efficient XML Parsing The Extensible Markup Language (XML) files are extensively used for representing structured data on the web for file configuration, exchanging data between distinct applications, web development, and many other applications. Consequently, effective parsing techniques are necessary for XML files to enhance the performance of applications. The existing parsing techniques have their strengths and weaknesses affecting the performance of applications. Researchers point out that the selection of an efficient and appropriate parser is the most challenging issue regarding a particular condition. This paper proposes a framework XML Parsing Optimization using Hybrid Machine Learning (XPOHML) that makes use of Artificial Neural Network (ANN) and Support Vector Machine (SVM) machine learning techniques for efficient XML parsing. The newly developed framework performs analysis and prediction of different XML parsers using profiling, classification, performance evaluation, and finally generates code for efficient parsing. The XML profiling phase of the XPOHML framework generates a dataset by evaluating the performance of PXTG, SAX, StAX, DOM, and JDOM parsing models on separate cores by applying numerous file sizes. The Classification phase produces the classification model by applying ANN and SVM techniques to identify the appropriate parsing model. The performance evaluation phase of XPOHML assesses the performance of both parsing models through classification metrics (accuracy). Additionally, based on evaluation outcomes, the code generation phase produces an efficient parsing model of XML files. The newly designed and developed XPOHML framework has shown a meaningful improvement in the performance of parsing XML files.
26-ANSP-ML-002	A new fault detection method using machine learning in analog radio-on- fiber MIMO transmission system In this letter, we propose a new fault detection method in a wired-wireless multiple-input and multiple-output (MIMO) transmission over an analog radio-on-fiber (A-RoF) network. The A-RoF architecture makes the configuration of the remote node simple, however, lots of radio frequency (RF)-band analog devices are used in it, such as semiconductor fiber laser, optical fiber link, photo-detector and electrical amplifiers, and the methodology for detecting the fault on these devices has not been discussed. Conventional methods require dedicated devices for sending fault messages via out-of-band wireless data transmission, which increases the hardware requirement of the whole system configuration. This letter proposes a machine learning approach in an A-RoF-based MIMO transmission without any additional devices.
26-ANSP-ML-003	A Systematic Literature Review of Supervised Machine Learning Techniques for Predictive Maintenance in Industry 4.0 Industry 4.0 has driven a paradigm shift in manufacturing, pushing industries to adopt innovative technologies for more efficient decision-making. A key component of this revolution is predictive maintenance, which plays a central role in this transformation by leveraging, among others methods, supervised machine learning techniques to anticipate equipment failures, optimise maintenance schedules, and enhance operational efficiency.

	This study presents a Systematic Literature Review (SLR) of 216 peer-reviewed papers published between 2019 and 2024, analysing the adoption of supervised machine learning techniques for predictive maintenance in various industrial domains, including manufacturing, machinery, energy, and smart systems. Unlike previous SLRs that broadly examine ML applications, this review provides a structured taxonomy of predictive maintenance methods, highlighting their domain-specific usage and prevalence in safety-critical industries. Additionally, we analyse the types of datasets used, revealing a strong preference for real-world data but limited public availability, which poses challenges for reproducibility and benchmarking. This study identifies key trends in ML adoption and offers insights into future research directions, thereby reinforcing the need for open datasets, explainable AI, and cross-domain generalization.
26-ANSP-ML-004	Academic Performance Prediction Using Machine Learning Approaches: A Survey Properly predicting students' academic performance is crucial for elevating educational outcomes in various disciplines. Through precise performance prediction, schools can quickly pin-point students facing challenges and provide customized educational materials suited to their specific learning needs. The reliance on teachers' experience to predict students' academic performance has proven to be less accurate and efficient than desired. Consequently, the past decade has witnessed a marked surge in employing machine learning and data mining techniques to forecast students' performance. However, the academic community has yet to agree on the most effective algorithm for predicting academic outcomes. Nonetheless, conducting an analysis and comparison of the existing algorithms in this field remains meaningful. Furthermore, recommendations for selecting an appropriate algorithm will be provided to interested researchers and educators based on their specific requirements. This article reviews the state-of-the-art literature on academic performance predictions using machine learning approaches in recent years. It details the variables analyzed, the algorithms implemented, the datasets utilized, and the evaluation metrics applied to assess model efficacy. What makes this work different is that relevant surveys in the past 10 years are also analyzed and compared, highlighting their contributions and review methods. In addition, we compared the accuracy of various machine learning models using popular open-access datasets and determined the best-performing algorithms among them. Our dataset and source codes are released for future algorithm comparisons and evaluations in this community.
26-ANSP-ML-005	An Analysis of Semi-Supervised Machine Learning in Electrical Machines This research outlines the significance of semi-supervised machine learning (SSML) in dealing with the intricate characteristics of electrical machines. SSML provides a key benefit in enhancing the effectiveness and precision of predictive models for optimizing electrical machine performance, reliability, and maintenance by leveraging labeled and unlabeled data. The research investigates important SSML algorithms such as self-training, co-training, generative models, and graph-based methods, highlighting their particular uses in fault diagnosis, condition monitoring, and predictive maintenance of electrical machines. Moreover, the document discusses the specific difficulties associated with this merger and offers remedies to improve the utilization of SSML in this significant area. A detailed table summarizes various methods and emphasizes their role in furthering machine learning within the realm of electrical machines.

26-ANSP-ML-006	Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques Anomaly detection in network traffic is a critical aspect of network security, particularly in defending against the increasing sophistication of cyber threats. This study investigates the application of various machine learning models for detecting anomalies in network traffic, specifically focusing on their effectiveness in addressing challenges such as class imbalance and feature complexity. The models assessed include Isolation Forest, Naive Bayes, XGBoost, LightGBM, and SVM classification. Through comprehensive evaluation, this research explores both supervised and unsupervised approaches, comparing their performance across key metrics like accuracy, F1-score, and recall. The results reveal that while models like XGBoost and LightGBM exhibit impressive performance, with LightGBM achieving near-perfect training accuracy (1.0) and solid test accuracy (0.85), others like Isolation Forest show limitations with low accuracy. The study highlights the strengths and weaknesses of each model, providing valuable insights into their practical application for network anomaly detection. By comparing different algorithms, this research contributes to advancing the application of machine learning in network security, offering guidance on model selection and optimization for improved detection of cyber threats.
26-ANSP-ML-007	Category-Based Sentiment Analysis of Sindhi News Headlines Using Machine Learning, Deep Learning, and Transformer Models The rapid growth of digital content has made sentiment analysis (SA) an essential tool for understanding public sentiment and classifying textual data. Despite significant progress in natural language processing (NLP), low-resource languages, particularly Sindhi, remain underexplored due to the lack of computational tools and annotated datasets. This study addresses this gap by introducing the Sindhi News Headlines Dataset (SNHD), a novel corpus annotated for both SA and category classification across eight categories: Crime, Economy, Entertainment, Health, Politics, Science & Technology, Social, and Sports. To evaluate the effectiveness of different machine learning (ML), deep learning (DL), and transformer-based approaches, we conduct a comparative analysis of various models on SA and category classification tasks. Furthermore, we leverage Explainable Artificial Intelligence (XAI) techniques, such as Local Interpretable Model-Agnostic Explanations (LIME), to gain insights into model decision-making. Experimental results show that traditional ML models outperform DL and transformer-based models on the SNHD dataset. Specifically, Support Vector Machines with Radial Basis Function (SVM-RBF) achieves the highest performance for SA (0.74 accuracy and weighted F-score), while the Ridge Classifier (RC) delivers the best results for category classification (0.84 accuracy and weighted F-score). Among transformer models, XLM-RoBERTa demonstrates strong performance in category classification (0.82 accuracy and weighted F-score). These findings establish a benchmark for future research in Sindhi NLP and highlight the potential of hybrid approaches in tackling challenges associated with low-resource languages. This work provides a foundational resource for NLP researchers seeking to advance computational methods for Sindhi and similar underrepresented languages.

26-ANSP-ML-008	Comparative Study of Machine Learning and Deep Learning Models for Early Prediction of Ovarian Cancer Ovarian cancer remains one of the most difficult gynecological cancers to detect early, often resulting in poor survival rates. This study presents a comparative analysis of machine learning (ML) and deep learning (DL) models for the early prediction of ovarian cancer using clinical and biomarker data. The dataset undergoes comprehensive preprocessing, including handling missing values, outlier removal, normalization, and dimensionality reduction via PCA. Feature selection methods such as Feature Importance, Recursive Feature Elimination (RFE), and autoencoder-based techniques are employed to enhance model performance. Various classifiers, including KNN, SVM, Logistic Regression, Random Forest, and deep networks like ANN, FNN, CNN, and RNN, are evaluated. Ensemble models such as Bagging, AdaBoost, Stacking, and XGBoost are also implemented. Our results show that the Feedforward Neural Network (FNN), combined with autoencoder-based feature selection, achieved the highest accuracy (85.71%), indicating its potential as a reliable predictive model for ovarian cancer. This comparative study highlights the significance of integrating optimized preprocessing, feature engineering, and model selection for effective early diagnosis.
26-ANSP-ML-009	Enhancing Accessibility Through Machine Learning: A Review on Visual and Hearing Impairment Technologies Assistive technologies powered by machine learning are transforming the way sensory impairments are addressed, offering innovative solutions for individuals with hearing and visual disabilities. This paper provides a comprehensive review of machine learning algorithms designed to enhance accessibility for these groups. For hearing impairments, the analysis focuses on advanced models such as Support Vector Machines (SVM), Random Forests (RF), and Multi-Layer Perceptrons (MLP), examining their effectiveness in auditory assistive applications. In the context of visual impairments, state-of-the-art object detection frameworks like You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD), and RetinaNet are evaluated for their capability to enable real-time object recognition and navigation aids. The study also reviews the Generative Artificial Intelligence based applications for visual and hearing impaired use cases. The study addresses the unique challenges and requirements associated with each type of sensory impairment, with particular emphasis on the customization and fine-tuning of machine learning models for personalized, effective solutions. Additionally, it highlights the transformative potential of deep learning models in advancing assistive technologies, ultimately aiming to enhance the quality of life for individuals with sensory disabilities. By advocating for the development and integration of such technologies, this paper underscores the importance of inclusivity and empowerment in creating a more equitable society.
26-ANSP-ML-010	Enhancing Chronic Disease Prediction in IoMT-Enabled Healthcare 5.0 Using Deep Machine Learning: Alzheimer's Disease as a Case Study Chronic disease significantly affects health on a global scale. Deep machine learning algorithms have found widespread application in the diagnosis of chronic diseases. Early diagnosis and treatment reduce the chance of a disease getting worse and, as a result, raise related mortality. The main objective of this work is to present a deep machine learning-based approach that provides better results in terms of accuracy. These findings have significance for tailored healthcare 5.0, enabling healthcare professionals to predict chronic disease more efficiently. A comparative examination of the most recent methods has been provided in our work reveals that it might be more advantageous to use the proposed model in which segmentation of the MRI is performed using U-net architecture

	and then classification is done using transfer learning for chronic disease prediction. Our proposed model provides 96.06% accuracy, it advances our understanding of deep machine learning's potential for chronic disease prediction and emphasizes the need to tailor model selection to specific disease types using data from IoMT enabled devices. In order to make advanced improvement in the field of healthcare 5.0, future studies should focus on refining these models and investigating how well they work with a wider range of datasets.
26-ANSP-ML-011	Enhancing Student Management Through Hybrid Machine Learning and Rough Set Models:A Framework for Positive Learning Environments Effective student management is crucial for fostering productive learning environments. This study presents a hybrid framework integrating machine learning (ML) techniques with rough set theory to enhance student management by identifying at-risk students and enabling personalized interventions. The model combines classification algorithms with rough set-based decision rules to analyze complex student data, including academic performance, behavior patterns, and levels of engagement. The ML layered approach detects patterns and outliers, supporting data-driven decisions to improve student well-being and educational outcomes. Evaluation on the Open University Learning Analytics Dataset (OULAD) demonstrated high accuracy (97.85%) in predicting student outcomes and precision (94.62%) in identifying students needing support. The hybrid approach outperformed conventional methods by approximately 15%, showcasing its transformative potential. This framework effectively monitors student performance and enables customized interventions to meet individual learning needs, fostering a more supportive educational environment.
26-ANSP-ML-012	Exploring the Effectiveness of Machine Learning and Deep Learning Techniques for EEG Signal Classification in Neurological Disorders Neurological disorders are among the leading causes of both physical and cognitive disabilities worldwide, affecting approximately 15% of the global population. This study explores the use of machine learning (ML) and deep learning (DL) techniques in processing Electroencephalography (EEG) signals to detect various neurological disorders, including Epilepsy, Autism Spectrum Disorder (ASD), and Alzheimer's disease. We present a detailed workflow that begins with EEG data acquisition using a headset, followed by data preprocessing with Finite Impulse Response (FIR) filters and Independent Component Analysis (ICA) to eliminate noise and artifacts. Furthermore, the data is segmented, allowing the extraction of key features such as Bandpower and Shannon entropy, which improve classification accuracy. These features are stored in an offline database for easy access during analysis, to be then applied for both ML and DL models, systematically testing their performance and comparing the results to prior studies. Hence, our findings show impressive accuracy, with the random forest model achieving 99.85% accuracy in classifying autism vs. healthy subjects and 100% accuracy in distinguishing healthy individuals from those with dementia using Support Vector Machines (SVM). Moreover, deep learning models, including Convolutional Neural Networks (CNN) and ChronoNet, demonstrated accuracy rates ranging from 92.5% to 100%. In conclusion, this research highlights the effectiveness of ML and DL techniques in EEG signal processing, offering valuable contributions to the field of brain-computer interfaces and advancing the potential for more accurate neurological disease classification and diagnosis.

26-ANSP-ML-013	Fuzzy Enhanced Kidney Tumor Detection: Integrating Machine Learning Operations for a Fusion of Twin Transferable Network and Weighted Ensemble Machine Learning Classifier Kidney tumors, often asymptomatic, can lead to serious health problems if left undiagnosed. This study tackles the crucial issue of kidney tumor detection using CT scans. The proposed approach leverages the power of image enhancement using fuzzy systems, deep learning, and machine learning for automated kidney tumor detection in CT images. The study proposes a fuzzy inference system to enhance kidney CT image contrast. This system analyzes image data and uses fuzzy logic to adjust pixel intensities, aiming to improve the distinction between features in the image without creating over-enhancement. Two pre-trained deep convolutional neural networks (PT-DCNNs), DenseNet121 and ResNet101, are used to extract features from the enhanced CT images. These features capture essential characteristics that differentiate between normal and tumor-containing scans. Combining features from twin PT-DCNNs (ensemble approach) creates a richer representation of the image content. The informative features are fed into a combined classifier where Support Vector Machines and Random Forests are combined using a weighted average to achieve the final and potentially more robust classification of kidney tumors. To improve training, we amplified the original dataset by creating variations with added noise and artificial modifications to simulate real-world image imperfections. The integration of Machine Learning Operations practices ensures the scalability, reproducibility, and clinical deployment of the system. The model achieved an impressive accuracy of 99.2% on high-quality images and 98.5% on noisy images, surpassing traditional methods. This automated approach can assist urologists in confirming the presence of kidney tumors, minimizing human error during physical inspection and potentially leading to improved patient outcomes.
26-ANSP-ML-014	GuardianML: Anatomy of Privacy-Preserving Machine Learning Techniques and Frameworks Machine learning has become integral to our lives, finding applications in nearly every aspect of our daily routines. However, using personal information in machine learning applications has raised concerns about user data privacy and security. As concerns about data privacy grow, algorithms and techniques for achieving robust privacy-preserving machine learning (PPML) have become a pressing technical challenge. PPML aims to safeguard the confidentiality of both data and models and ensure that sensitive information remains protected during training and inference processes. Various techniques, protocols, libraries, and frameworks have been proposed for PPML, but choosing the right combination along with the appropriate algorithmic or system parameters for a specific deployment instance can be very difficult. In this work, we introduce GuardianML, an open-source recommendation system for selecting the correct parameters and suitable framework for specific use cases of PPML. GuardianML allows users to search through a wide range of PPML frameworks, techniques, protocols, libraries, and more based on a set of objectives. GuardianML filters potential frameworks based on user-defined criteria, such as the number of parties involved in multi-party computation or the need to minimize communication costs in homomorphic encryption scenarios. The system's recommendations and optimizations are formulated as a maximization problem using linear integer programming to identify the most suitable solution for various use cases. Moreover, this work thoroughly analyzes and presents seventy relevant frameworks in the system's database. Additionally, we offer an open-source repository containing practical examples and documentation for some of the frameworks.

26-ANSP-ML-015	Handwritten Amharic Character Recognition Through Transfer Learning: Integrating CNN Models and Machine Learning Classifiers Handwritten Amharic character recognition presents significant challenges due to the script's syllabic nature and variations in handwriting styles. This study investigates a hybrid approach that integrates convolutional neural networks (CNNs) with machine learning classifiers to enhance recognition accuracy. Transfer learning is applied using four CNN architectures: AlexNet, VGG16, VGG19, and ResNet50 as feature extractors. Initially, their performance is evaluated with softmax classifiers. Subsequently, the softmax layer is replaced with machine learning classifiers, including Random Forest, XGBoost, and Support Vector Machine (SVM), while freezing the pretrained feature extractors. The Hybrid ResNet50 + SVM model achieves the highest accuracy of 91.89%, with a precision of 92.46%, recall of 91.15%, and an F1-score of 91.80%. These results indicate that SVM serves as a potential alternative to softmax, offering robust classification performance for complex handwritten scripts. This research contributes to advancements in handwritten character recognition systems for underrepresented languages.
26-ANSP-ML-016	Imbalanced Data Problem in Machine Learning: A Review One of the prominent challenges encountered in real-world data is an imbalance, characterized by unequal distribution of observations across different target classes, which complicates achieving accurate model classifications. This survey delves into various machine learning techniques developed to address the difficulties posed by imbalanced data. It discusses data-level methods such as oversampling and undersampling, algorithm-level solutions including ensemble learning and specific algorithm adjustments, cost-sensitive algorithms, and hybrid strategies that combine multiple approaches. Moreover, this paper emphasizes the crucial role of evaluation methods like Precision, F1 Score, Recall, G-mean, and AUC in measuring the effectiveness of these strategies under imbalanced conditions. A detailed review of recent research articles helps pinpoint persistent gaps in generalizability, scalability, and robustness across these methods, underscoring the necessity for ongoing improvements. The survey seeks to offer an extensive overview of current approaches that improve the efficiency and effectiveness of machine learning models dealing with imbalanced datasets, thus equipping researchers with the insights needed to develop robust and effective models ready for real-world application.
26-ANSP-ML-017	Implementation and Performance Evaluation of Machine Learning-Based Apriori Algorithm to Detect Non-Technical Losses in Distribution Systems The emergence and augmentation of nontechnical losses (NTLs) in a power distribution system has always been considered a critical issue in the global electricity market. Being considered as uncharged, unlawfully or unfairly charged consumed electricity, NTLs due to fraudulent activities can impose a serious burden on the grid and revenue loss in the state's budget. The detection of NTLs based on the analysis of huge consumer dataset using machine learning is an acknowledged approach due to its expedient performance compared to manual inspection. This paper addresses a comprehensive investigation of monthly electricity consumption data of ~15000 consumers over three years using machine learning for NTL detection. The data is acquired by meters having automatic meter reading (AMR) capability and is processed using support vector machine (SVM) classifier, deep neural network (DNN), gradient boosted reinforcement learning (GBRL) and apriori algorithm. The results of ML algorithms are assessed by various performance metrics based on the confusion matrix and compared among themselves

	as well as with the findings from the published works. The least recall score of 0.8926 is exhibited by SVM classifier and poorest scores of accuracy (0.9376), specificity (0.9451), precision (0.7521), F1 (0.8183) and false positive rate (0.0549) are given by DNN program. With apriori algorithm, the scores of accuracy, recall, specificity, precision, F1 and false positive rate are observed as 0.9964, 0.9927, 0.9971, 0.9843, 0.9885, 0.0029 respectively. In all the performance based six domains, a significant improvement of 6%, 10%, 5%, 23%, 17% and 5% respectively compared to the least scores is demonstrated by the apriori algorithm. Therefore, in all these domains, the apriori algorithm being reported for the first time in this research work outperforms all other methods of this paper as well as the similar published works.
26-ANSP-ML-018	Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements In response to the escalating threat of fake news on social media, this systematic literature review analyzes the recent advancements in machine learning and deep learning approaches for automated detection. Following the PRISMA guidelines, we examined 90 peer-reviewed studies published between 2020 and 2024 to evaluate the model effectiveness, identify limitations, and highlight emerging trends. Our analysis shows that deep learning models, particularly transformer-based architectures such as BERT, consistently outperform traditional machine learning methods, often achieving a high accuracy (Acc), precision (P), recall (R), and F1-score (F1). For instance, a BERT-based model reported up to 99.9% accuracy on the Kaggle fake news dataset and above 98% accuracy on other public datasets, including ISOT, Fake-or-Real, and D3. Similarly, the GANM model demonstrated robust performance on the FakeNewsNet dataset by integrating text and social features. Transfer learning and multimodal models that incorporate user behaviour and network information significantly improve detection in diverse, low-resource environments. However, challenges persist in terms of the dataset quality, model interpretability, domain generalisability, and real-time deployment. This review also underscores the limited adoption of few-shot and zero-shot learning techniques, highlighting a promising direction for future research on handling emerging misinformation using minimal training data. To support practical deployment, we advocate the development of explainable, multilingual, and lightweight models with greater emphasis on human-centred evaluation and ethical considerations. Our findings provide a foundation for researchers and practitioners to build scalable, trustworthy, and context-aware fake news detection systems for global use.

26-ANSP-ML-019	Machine Learning Approaches to Code Similarity Measurement: A Systematic Review Source code similarity measurement, which involves assessing the degree of difference between code segments, plays a crucial role in various aspects of the software development cycle. These include but are not limited to code quality assurance, code review processes, code plagiarism detection, security, and vulnerability analysis. Despite the increasing application of ML technique in this domain, a comprehensive synthesis of existing methodologies remains lacking. This paper presents a systematic review of Machine Learning techniques applied to code similarity measurement, aiming to illuminate current methodologies and contribute valuable insights to the research community. Following a rigorous systematic review protocol, we identified and analyzed 84 primary studies on a broad spectrum of dimensions covering application type, devised Machine Learning algorithms, used code representations, datasets, and performance metrics, as well as performance evaluations. A deep investigation reveals that 15 applications for code similarity measurement have utilized 51 different machine learning algorithms. Additionally, the most prevalent code representation is found to be the abstract syntax tree (AST). Furthermore, the most frequently employed dataset across various code similarity research applications is BigCloneBench. Through this
----------------	--

	comprehensive analysis, the paper not only synthesizes existing research but also identifies prevailing limitations and challenges, shedding light on potential avenues for future work.
26-ANSP-ML-020	Neural-XGBoost: A Hybrid Approach for Disaster Prediction and Management Using Machine Learning Effective disaster prediction is essential for disaster management and mitigation. This study addresses a multi-classification problem and proposes the Neural-XGBoost disaster prediction model (N- XGB), a hybrid model that combines neural networks (NN) for feature extraction with XGBoost for classification. The NN component extracts high-level features, while XGBoost uses gradient-boosted decision trees for accurate predictions, combining the strengths of deep learning and boosting techniques for improved accuracy. The N-XGB model achieves an accuracy of 94.8% and an average F1 score of 0.95 on a real-world dataset that includes wildfires, floods and earthquakes, significantly outperforming baseline models such as random forest, Support vector machine and logistic regression 85% accuracy. The balanced F1 scores for wildfires 0.96, floods 0.93, and earthquakes 0.96 demonstrate the model's robustness in multi-class classification. The Synthetic Minority Oversampling Technique (SMOTE) balances datasets and improves model efficiency and capability. The proposed N-XGB model provides a reliable and accurate solution for predicting disasters and contributes to improving preparedness, resource allocation and risk management strategies.
26-ANSP-ML-021	Smell-ML: A Machine Learning Framework for Detecting Rarely Studied Code Smells Code smells are design flaws that reduce the software quality and maintainability. Machine learning classification models have been used to detect different code smells. However, such studies targeted code smells in depth, while leaving other under-explored smells; even so, such smells have a significant impact on source code quality. Recent surveys have highlighted a group of code smells that has rarely been studied by researchers. Furthermore, some machine learning classification models were evaluated on a subset of the source code features while ignoring significant features during classification. This paper proposes a novel approach, called Smell-ML, for detecting five rarely studied code smells: Middle Man (MM), Class Data Should Be Private (CDSBP), Inappropriate Intimacy (II), Refused Bequest (RB), and Speculative Generality (SG). The novelty of this approach stems from the improvement in both the data preparation and classification phases. During data preparation, Smell-ML relies on data balancing and an extended source code feature list to improve accuracy. In the classification phase, different classifiers were assessed, including traditional, ensemble, and multi-level classifiers. We evaluated Smell-ML on a dataset composed of 13 open source Java projects with 125 versions per project. The results show that Smell-ML's detection F1-score values surpass those of previous studies with significant improvements across various code smells. The F1-score measure of the 11 machine learning classifiers improved after using the extended feature list. Data balancing and multi-level classification notably boosted accuracy.

26-ANSP-ML-022	State-of-the-Art Fault Detection and Diagnosis in Power Transformers: A Review of Machine Learning and Hybrid Methods Fault detection and diagnosis (FDD) in power transformers is essential for maintaining reliability and safety in modern power systems. Recent trends in transformer fault diagnosis include the use of advanced monitoring systems and data analysis techniques, and implementing effective FDD strategies can prevent costly repairs, minimize downtime, and enhance the overall reliability of power systems. In this work, a systematic review of FDD in power transformers, focusing on machine learning and hybrid methods applications is presented. The methodology comprised a detailed systematic process undertaken to gather, filter, and analyze relevant research papers using Scopus database, while VOSviewer, and bibliometrix tools were utilized for results analysis. The research findings indicate that there is clear progress in detecting transformer faults, moving from older methods like Dissolved Gas Analysis (DGA) to machine learning models such as Random Forest and Convolutional Neural Networks (CNN). Hybrid models combining machine learning with optimization have made detection more accurate. New tools, including optical sensors, now allow for real-time monitoring. Still, issues like limited data and complex models remain. This study contributes by reviewing how machine learning is applied to transformer fault detection, exploring hybrid methods that combine traditional techniques like DGA with advanced models for better accuracy. It identifies key research patterns, trends, and themes, while also highlighting gaps and offering suggestions for future research to improve diagnostics and monitoring.
26-ANSP-ML-023	The Implementation of a Prediction System for Sugarcane's Destruction Rate From Sugarcane Stem Borer via Hybrid Machine Learning Sugarcane stem borer poses a significant threat to sugarcane production, causing extensive damage and economic losses in the agricultural sector. Accurately predicting the destruction rate caused by this pest is crucial for implementing timely and effective management strategies. This paper presents the development and implementation of a time series prediction system designed to estimate the short-time destruction rate of sugarcane in the near future caused by stem borer infestations. The proposed Hybrid system combines Adaptive and Batch machine learning techniques, allowing for continuous model updates and adaptation to new data while maintaining initial accuracy, thereby improving predictive accuracy over time. A dataset comprising historical pest infestation data, manually recorded from the fields, and environmental factors, retrieved online, was used to train the model to capture complex temporal dependencies, with features such as temperature, humidity, seasonal patterns, and pest density being key predictors. The performance of the model was evaluated using the MAE, AIC, BIC, and HQIC, which demonstrated its robustness and reliability in predicting the destruction rate. This proposed Hybrid prediction system offers a valuable tool for sugarcane farmers and agronomists, enabling proactive decision-making to mitigate the impact of stem borer infestations. The findings of this study contribute to the broader application of machine learning in agricultural pest management, highlighting the potential for Hybrid learning approaches to enhance prediction accuracy in dynamic environments.

26-ANSP-ML-024	Toward Drought Modeling in South Asia: Machine Learning Approaches, Challenges, and Opportunities Drought is an environmental and economic problem. Sustainable ecosystems, water resources, food security, and ecosystem sustainability. Machine all are severely affected by drought. Due to the increasing frequency and severity of droughts caused by climate change. Effective drought modeling is crucial for early warning systems and risk mitigation. Recent advances in machine learning (ML) and deep learning (DL) techniques have been developed as potential drought modeling tools, which offer accurate and reliable drought detection. This review paper summarizes the drought modeling(Drought Prediction, Drought Detection and Drought Forecasting) approaches. This paper focuses on three main aspect. 1) The selection of the region for this study, for this study South Asia(SA) is selected as region of interest (ROI) that offer accurate drought modeling, providing policymakers and decision-makers with insightful information. The geographical scope of this study is the region of South Asia. This region is selected because of its heavy reliance on agriculture. 2) This paper focuses on the current and future trends, challenges, and advances of and vulnerability to droughts. The review offers a thorough grasp of how drought conditions are evaluated by gathering and analyzing the most important drought indicators and metrics specific to South Asia. The paper explores the current state-of-the-art in ML and DL for drought modeling. 3) This review encapsulates the indicator and metrics (Complex Machine learning and deep learning models) for drought modeling which are most relevant to the SA region. This study sum up as most common challenges in drought modeling are, highlighting current challenges such as incomplete and inconsistent datasets, lack of explainable and interpretable models, and unavailability of data for model uncertainty analysis. This study proposes that these problems can be solved with modern machine learning techniques such as explainable machine learning and federated or Lack of explainability and interpretability in complex ML/DL models, unavailability of benchmarks. Based on these challenges, this review suggests the following techniques to address these challenges: Data integration (Data fusion), distributed machine learning. (Federated Learning) and explainable AI (XAI, SHAP, LIME, etc.).
26-ANSP-ML-025	VGG-16, VGG-16 With Random Forest, Resnet50 With SVM, and EfficientNetB0 With XGBoost-Enhancing Bone Fracture Classification in X-Ray Using Deep Learning Models Millions of cases of bone fractures are reported every year, and accuracy in classification is crucial to help with proper management and treatment. The recently developed techniques of Machine Learning, particularly Deep Learning, have been effective in increasing diagnosis precision and efficiency. We utilized a diverse dataset comprising 10 different classes of fracture types captured in X-Ray images. This paper makes a comparison of different machine learning models on classifying bone fractures: VGG-16, VGG-16 with Random Forest, ResNet-50 with Support Vector Machine, and EfficientNetB0 with XGBoost. Model performances were evaluated with respect to parameters of precision, recalls, and F1-scores. According to results, VGG-16 and its variant ensemble with Random Forest outperformed with an accuracy of 0.95 when compared to others on every parameter for different classes of fractures. Results indicate that models based on VGG16 are quite effective for bone fracture classification.