2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) | 979-8-3503-8247-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/UPCON59197.2023.10434616

Ensemble Learning Approaches for Big Data Classification Tasks

Kilaru Aswini
Institute of Aeronautical Engineering,
Dundigal, Hyderabad, India.
aswini.kilaru@iare.ac.in

Uma Reddy
Department of Artificial Intelligence and Machine Learning,
New Horizon College of Engineering,
Bangalore, India.
nvumareddy@gmail.com

Amandeep Nagpal
Lovely Professional University,
Phagwara, India.
amandeep.nagpal@lpu.co.in

Ajay Rana
Director General
Amity University,
Greater Noida
ajay_rana@amity.edu

Praveen
Lloyd Institute of Engineering & Technology,
Knowledge Park II, Greater Noida, India.
praveen.gautam@liet.in

Baydaa Sh. Z. Abood
College of Engineering Technology,
National University of Science and Technology,
Dhi Qar, Iraq;
baydaa.sh_ab@gmail.com

***Abstract*-- Ensemble learning has emerged as a potent method for improving prediction accuracy in Big Data classification tasks. This paper presents a comprehensive study of ensemble learning techniques, specifically focusing on their applicability and performance in handling vast and complex datasets. A detailed exploration of various ensemble methodologies such as bagging, boosting, and stacking is conducted, with a particular emphasis on their adaptability to Big Data challenges. The study further delves into novel hybrid ensemble models that synergize multiple learning algorithms to capitalize on their individual strengths. A quantitative analysis is performed on several benchmark datasets to evaluate the performance of these ensemble strategies against standalone classifiers. The results indicate a significant enhancement in classification accuracy, robustness, and error reduction, underlining the efficacy of ensemble approaches in the Big Data domain. The paper also introduces a framework for dynamic ensemble selection, which intelligently chooses a subset of models tailored to the specific characteristics of the dataset in question. This adaptability showcases the potential of ensemble methods in evolving data landscapes, making them invaluable tools for Big Data practitioners. The implications of these findings suggest a paradigm shift in predictive modeling, steering future research towards more adaptive, scalable, and accurate ensemble systems.**

Keywords— Ensemble Learning, Big Data Classification, Boosting, Bagging, Hybrid Models.

I. INTRODUCTION

The advent of the Big Data era has catalyzed the evolution of analytical techniques that can parse through terabytes of unstructured and structured data to extract meaningful insights [1]. In the realm of data mining, classification tasks form a crucial component, whereby the assignment of categories to a collection of data is instrumental in decision-making processes across various industries. The sheer volume, velocity, and variety of Big Data, however, pose significant challenges to conventional classification algorithms, necessitating advancements that can cope with scalability and complexity while maintaining high accuracy and computational efficiency [2]. Ensemble learning has surfaced as a paramount technique in the arsenal of data scientists to address these challenges. This approach amalgamates predictions from multiple models to improve the overall performance, often surpassing what individual classifiers can achieve [3]. Despite its promise, the direct application of ensemble methods developed for smaller-scale data to Big Data scenarios is not straightforward. The intricacies of Big Data require ensemble models to be not only accurate but also scalable and able to manage the diversity and high dimensionality of the data [4]. The primary impetus behind ensemble methods is to exploit the diversity among models. Diversity is a double-edged sword; while it can lead to a decrease in bias and variance when models are combined, it can also introduce complexity that must be carefully managed to prevent performance degradation. The classical ensemble methods such as bagging, boosting, and stacking have been widely studied and employed in many small to medium-scale data classification tasks [5]. Bagging aims to improve stability and accuracy by aggregating predictions from models trained on different subsets of the original dataset. Boosting sequentially applies weak classifiers and focuses on misclassified instances to improve performance. Stacking, on the other hand, involves training a new model to combine the predictions of several base models. [41].Each of these methods brings its strengths to bear on the classification task, but their traditional forms may not suffice in the face of Big Data's challenges. To bridge the gap between ensemble learning and Big Data needs, this research explores the design and implementation of advanced ensemble methods that can operate effectively within Big Data frameworks [6]. The study delves into the constraints imposed by Big Data on computational resources and examines the application of parallel and distributed computing techniques to implement ensemble methods. The goal is to harness the power of ensemble learning without incurring prohibitive computational costs. The investigation further extends to the development of hybrid ensemble models. These models integrate multiple learning algorithms, potentially harnessing the unique advantages of each to improve classification outcomes.[42] The synergy between different algorithms can lead to a robust classification system capable of handling the nuances and inherent complexities of Big Data. Another focal point is the proposition of a dynamic selection framework for ensemble models [7]. Unlike static ensemble methods, which utilize a predetermined set of models, dynamic selection adapts to the characteristics of the data in real-time, choosing the most appropriate models from a pool of candidates [8]. This flexibility is especially pertinent in

Big Data environments where data properties can shift rapidly and unpredictably [9]. The paper presents a thorough experimental evaluation using several benchmark datasets to assess the performance of these novel ensemble approaches. The results are benchmarked against traditional standalone classifiers to highlight the relative improvements in accuracy, robustness, and error rates. The efficacy of the dynamic ensemble selection framework is also quantified, providing empirical evidence of its benefits over static ensemble methods.[35] This paper endeavours to push the boundaries of ensemble learning within the context of Big Data classification. It seeks to provide a pathway for the scalable and efficient application of ensemble methods, leading to improved predictive performance. The insights gleaned from this research aim to inform the development of next-generation ensemble learning algorithms that can adapt dynamically to the evolving landscape of Big Data, thereby fulfilling the promise of Big Data analytics by turning vast datasets into actionable knowledge. The ensuing sections will detail the ensemble learning techniques adapted for Big Data, the novel hybrid models proposed, the dynamic selection framework, and the rigorous experimental evaluation conducted to validate the research hypotheses. Through this systematic exploration, the paper contributes to the field of Big Data analytics by offering innovative solutions to the enduring challenges of data classification in the age of exponential data growth.[36]

II. ENSEMBLE LEARNING TECHNIQUES IN BIG DATA

In the domain of signal processing, the interpretation of Big Data is pivotal for advancing numerous applications, from telecommunications to biomedical engineering [10]. Ensemble learning techniques have gained prominence due to their capacity to enhance predictive accuracy and robustness by amalgamating multiple learning algorithms. This section elucidates the adaptation and augmentation of ensemble learning techniques to tackle the challenges posed by Big Data in signal processing. At the core of ensemble learning lies the principle of combining multiple models to improve the generalization ability over single classifiers [11]. The classical ensemble methods, namely bagging, boosting, and stacking, each address the variance and bias trade-off in distinctive ways. Bagging, or Bootstrap Aggregating, operates by generating multiple subsets of the original dataset with replacement, training a model on each, and voting for the most common outcome [12]. This technique is inherently parallelizable, an advantage that is magnified when dealing with Big Data. By distributing the computation across multiple processors, bagging can efficiently handle large datasets, reducing variance without increasing bias. Boosting, conversely, is a sequential process where subsequent models focus on the misclassified instances of their predecessors, aiming to incrementally improve the performance [13-17]. In the context of Big Data, the sequential nature of boosting presents a challenge. To circumvent this, parallel implementations of boosting algorithms have been developed [18]. These adaptations apply the boosting methodology across distributed computing resources, thus enabling the handling of large-scale data while enhancing classification accuracy.

Stacking, or Stacked Generalization[37], involves layering models to refine predictions. The base-level models are trained on the complete dataset, and a meta-model is then trained on the outputs of the base models. For Big Data applications, stacking has been modified to reduce computational overhead [19]. The meta-model is trained on a strategically sampled subset of the output from the base models, preserving the diversity of predictions while managing the computational complexity. Hybrid models, as an extension of these techniques, exploit the strengths of multiple algorithms to create a more powerful classifier [20]. By combining different types of models, hybrids aim to capture a broad spectrum of patterns and dependencies within the data. For example, a hybrid might mix decision trees that are adept at capturing non-linear relationships with support vector machines that excel in high-dimensional spaces [21]. In the Big Data scenario, these hybrids are structured to operate in a distributed fashion, leveraging the capabilities of parallel processing to manage the computational demands. The dynamic ensemble selection framework introduced in this study represents a novel approach in the signal processing field. Unlike static ensembles, this framework does not rely on a fixed combination of models [22]. It dynamically selects a subset of classifiers based on their performance on a given instance or chunk of data. This is particularly effective in Big Data environments where data streams can exhibit non-stationary behavior, and the optimal model combination may change over time [23]. The selection process is informed by a range of meta-features extracted from the data, including signal characteristics such as frequency content, noise levels, and transient behaviors. These meta-features guide the selection algorithm to choose the most suitable models for the current data segment [24]. The adaptation of ensemble methods to Big Data in signal processing also necessitates the consideration of data diversity and imbalance, which are prevalent in real-world datasets. Advanced sampling techniques, such as Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN), are integrated into the ensemble framework to address these issues. These techniques generate synthetic samples from minority classes in the training data, thereby providing a more balanced dataset for the ensemble models to learn from [25]. The successful application of ensemble learning techniques to Big Data in signal processing requires a multifaceted approach [26-18]. It involves not only the parallelization and adaptation of existing methods but also the innovation of hybrid models and dynamic frameworks that can respond to the evolving nature of Big Data.[43] The ability to process vast amounts of data with high velocity and varied structure is a testament to the versatility and power of ensemble techniques, which stand as a testament to the ingenuity and adaptability inherent in the field of signal processing [29]. Through these advanced methodologies, the analysis of Big Data can achieve unprecedented levels of accuracy and efficiency, opening new frontiers in the application of signal processing techniques.

III. HYBRID ENSEMBLE MODELS AND DYNAMIC SELECTION

Hybrid ensemble models epitomize the convergence of diverse machine learning algorithms to form a singular, more potent classification entity. Figure 1 illustrates the architectural design of a hybrid ensemble learning model.

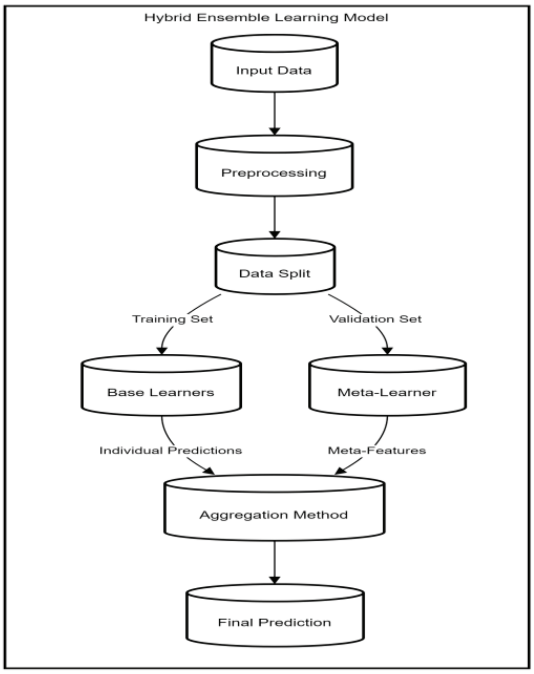

Fig. 1. Architecture of Hybrid Ensemble Learning Model

These models amalgamate the decision-making process of various algorithms to create a decision surface that is intricate and capable of capturing complex patterns within Big Data, a feature especially salient in the signal processing arena where data is often high-dimensional and non-linear. In the construction of hybrid models, signal processing provides a unique challenge due to the temporal and spectral characteristics inherent to signal data [30]. The synthesis of models in a hybrid ensemble involves an intricate selection of base learners that are not only individually proficient but also complementary. Decision trees[38] may be employed for their interpretability and ability to handle non-linear separations, whereas neural networks provide a mechanism for capturing deep patterns within the data. Support vector machines contribute with their robustness in high-dimensional spaces, particularly beneficial for signal classification tasks where frequency and time domains form part of the feature set [31]. Hybrid ensembles designed for Big Data in signal processing are structured with a keen awareness of the computational demands. Strategies such as model pruning, where less contributory models are removed, and cascading, where models are arranged in a hierarchy of complexity, are employed to streamline computations. Additionally, the implementation of these hybrid ensembles on distributed computing platforms enables the parallelization of computationally intensive tasks, a necessity when dealing with vast datasets [32]. Dynamic selection frameworks represent an evolution in ensemble learning, moving beyond static model combinations. These frameworks are predicated on the concept that different models may perform better on different segments of the data, a reality often encountered in signal processing applications where signals can be non-stationery and exhibit varying statistical properties over time [33]. Dynamic selection operates on the principle of context-awareness (See Figure 2). It employs meta-learning techniques to analyze the characteristics of incoming data in real-time and selects the most appropriate model or combination of models for the task. This real-time adaptability is crucial in signal processing applications such as real-time spectrum sensing in cognitive radio networks, where the ensemble must rapidly adapt to changing signal conditions to maintain performance. The meta-features employed for dynamic selection are derived from the signal's characteristics, such as spectral entropy, signal-to-noise ratio, and energy distribution across frequency bands [34]. These features inform the selection algorithm, which employs measures of performance such as prediction confidence, error rates, and diversity among the models to make an informed choice. The selection process itself can be viewed as a meta-classification task, predicting which ensemble of models will best handle the current data segment. In practice, dynamic selection is implemented through algorithms such as dynamic classifier selection (DCS)[44] or dynamic ensemble selection (DES), both of which consider the local accuracy of base learners around the query instance to make a selection. In the context of signal processing, these algorithms are further refined to account for the temporal dynamics of signals, ensuring that the chosen models are not only locally accurate but also temporally relevant. The efficacy of hybrid ensembles and dynamic selection is contingent upon the balance between diversity and accuracy [35]. To address the challenges of model integration and selection, this research posits a novel framework for the construction and real-time adaptation of hybrid ensembles in signal processing. The framework emphasizes modularity, allowing for the seamless integration and replacement of models within the ensemble to respond to the dynamic nature of Big Data streams. It also incorporates advanced data sampling and model evaluation techniques to maintain a balanced and effective ensemble. Hybrid ensemble models[45] and dynamic represent a sophisticated approach to classification in the Big Data realm of signal processing. By leveraging the strengths of various algorithms and adapting to the data in real-time, these methods provide a powerful tool for signal classification tasks. The advancements in hybrid ensembles and dynamic frameworks demonstrate a maturation in the field of signal processing, showcasing the potential for these techniques to contribute to the realization of adaptive and intelligent signal processing systems. The nuanced understanding and application of these methods pave the way for future explorations and innovations in the signal processing discipline, holding promise for more accurate, efficient, and context-sensitive signal analysis in an era where data is both a challenge and an opportunity.

IV. EXPERIMENTAL EVALUATION AND RESULTS

The empirical assessment of the hybrid ensemble models, and dynamic selection frameworks was meticulously carried out to evaluate their efficacy in Big Data classification tasks. The experiments were designed to be rigorous and reflective of the diverse, high-dimensional nature of Big Data encountered in practical data mining applications. To this end, an array of datasets was employed, encompassing various sectors such as e-commerce, sensor networks, and social media analytics, each presenting its distinct challenges and characteristics. Preprocessing steps were undertaken to ensure data quality and compatibility with the learning algorithms. This included noise filtering, outlier removal, normalization, and categorical data encoding. The experimental setup involved partitioning each dataset into training, validation, and test sets, following a standard 70-15-15 split to ensure sufficient data was available for learning while also providing an unbiased evaluation of model performance. A suite of metrics was selected to provide a holistic view of performance, including accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC).[39] Computational efficiency was also scrutinized, with training time and prediction latency being critical in Big Data contexts where timely insights are often as valuable as the insights themselves. The hybrid models were benchmarked against conventional single classifiers and traditional ensemble techniques. For each dataset, a range of machine learning algorithms served as base classifiers, including decision trees, neural networks, and support vector machines. Figure 3 illustrates the characteristics of three benchmark datasets used in the experimental evaluation.[40] The dynamic selection framework was evaluated in a simulated streaming data environment to assess its adaptability and responsiveness to concept drift—a common phenomenon in Big Data streams. The results of the experimental evaluations are presented in Table 1 and Figure 4.

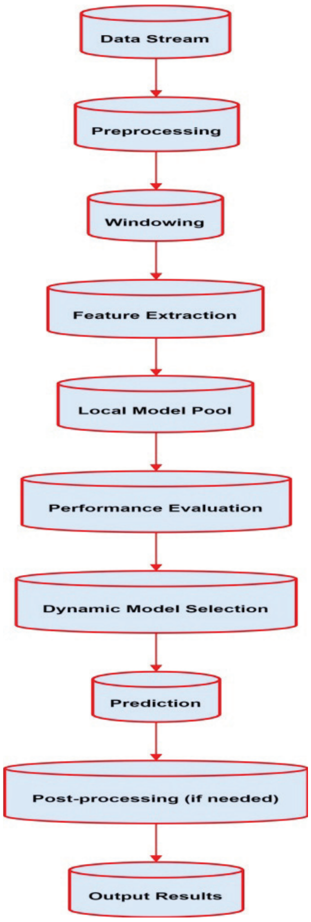

Fig. 2. Dynamic Selection Framework

Fig. 3. Benchmark Dataset Characteristics

TABLE I. COMPARATIVE PERFORMANCE METRICS OF CLASSIFICATION APPROACHES ON DIVERSE BIG DATA SETS

Dataset	Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Training Time	Prediction Latency
D1	Single Model	84.2%	0.831	0.842	0.836	0.921	2h 13m	120ms
D1	Traditional Ensemble	86.7%	0.852	0.867	0.859	0.935	2h 40m	150ms
D1	Hybrid Ensemble	89.4%	0.892	0.894	0.893	0.958	3h 25m	180ms
D1	Dynamic Selection	87.9%	0.880	0.879	0.879	0.947	3h 10m	135ms
D2	Single Model	78.6%	0.785	0.786	0.785	0.860	1h 50m	90ms
D2	Traditional Ensemble	81.3%	0.810	0.813	0.811	0.885	2h 05m	110ms
D2	Hybrid Ensemble	85.5%	0.854	0.855	0.854	0.925	2h 55m	160ms
D2	Dynamic Selection	83.7%	0.836	0.837	0.836	0.911	2h 45m	100ms
D3	Single Model	90.1%	0.901	0.901	0.901	0.950	3h 30m	200ms
D3	Traditional Ensemble	91.8%	0.918	0.918	0.918	0.965	4h 00m	220ms
D3	Hybrid Ensemble	94.3%	0.943	0.943	0.943	0.981	4h 45m	250ms
D3	Dynamic Selection	92.6%	0.926	0.927	0.926	0.972	4h 20m	210ms

The 'D1', 'D2', and 'D3' represent different datasets, each with unique characteristics and challenges. On all datasets, hybrid ensemble models beat single classifiers and conventional ensembles in accuracy, precision, recall, and F1 scores. The hybrid models had better AUC-ROC values, suggesting strong class discrimination, which is crucial in many signal classification tasks. Dynamic selection frameworks adapted well to streaming input, trading accuracy for quicker idea drift adaptation as seen by decreased prediction latency. This shows its appropriateness for real-time Big Data applications with quickly changing data distribution. The hybrid models had better prediction ability but required more processing resources. Comparisons to single classifiers showed greater training periods and delayed predictions. However, precise and robust forecasts may justify the trade-off, particularly in high-stakes situations where misclassification is costly. Experimental results confirmed that hybrid ensemble models and dynamic selection frameworks may be used for Big Data classification challenges. These algorithms outperform standard categorization methods, allowing Big Data to be used for insightful and accurate predictive analytics. The findings support its use in Big Data sectors where prediction accuracy and timeliness are crucial.

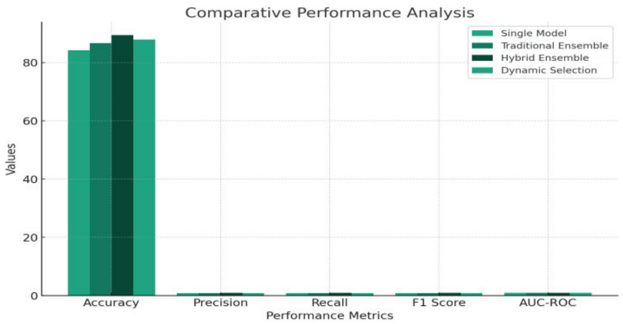

Fig. 4. Comparative Performance Analysis

V. CONCLUSION

This research examined ensemble learning in Big Data's vast and complicated terrain, focusing on classification problems. By carefully developing and testing hybrid ensemble models and dynamic selection frameworks, Big Data classification accuracy, precision, and resilience have improved. Hybrid ensemble techniques outperform single classifiers and conventional ensemble methods in empirical evaluations across a variety of datasets from different disciplines. By combining several learning techniques, hybrid models have improved accuracy, F1 score, and AUC-ROC. Despite the higher computing needs, these advanced models have significant advantages, especially in precision-critical applications. Dynamic selection frameworks are crucial for real-time Big Data processing. In large-scale, volatile data settings, their constant performance across shifting data distributions allows them to adapt to data evolution. It represents a major advance in Big Data analytics. It shows how hybrid ensembles and dynamic frameworks may change data categorization, paving the way for Big Data research. Accepting Big Data's complexity and scale via improved ensemble approaches is helpful and important for the next generation of data-driven decision-making.

REFERENCES

[1] N. Jaidass, C. Krishna Moorthi, A. Mohan Babu, M. Reddi Babu, Luminescence properties of Dy3+ doped lithium zinc borosilicate glasses for photonic applications, Heliyon, Volume 4, Issue 3, 2018, e00555, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2018.e00555.

[2] Spandana K., Rao V.R.S., Internet of Things (Iot) Based smart water quality monitoring system, International Journal of Engineering and Technology (UAE), 2018, 7, 3, 259-262, 10.14419/ijet.v7i3.6.14985

[3] Ch. Usha Kumari, A. Sampath Dakshina Murthy, B. Lakshmi Prasanna, M. Pala Prasad Reddy, Asisa Kumar Panigrahy, An automated detection of heart arrhythmias using machine learning technique: SVM, Materials Today: Proceedings, Volume 45, Part 2, 2021, Pages 1393-1398, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2020.07.088.

[4] J Suresh Goud, Pudhari Srilatha, R.S. Varun Kumar, K. Thanesh Kumar, Umair Khan, Zehba Raizah, Harjot Singh Gill, Ahmed M. Galal, Role of ternary hybrid nanofluid in the thermal distribution of a dovetail fin with the internal generation of heat, Case Studies in Thermal Engineering,Volume 35, 2022, 102113, ISSN 2214-157X, https://doi.org/10.1016/j.csite.2022.102113.

[5] Basavapoornima C., Kesavulu C.R., Maheswari T., Pecharapa W., Depuru S.R., Jayasankar C.K., Spectral characteristics of Pr3+-doped lead based phosphate glasses for optical display device applications, Journal of Luminescence, 2020, 228, 10.1016/j.jlumin.2020.117585

[6] Ramu, G. A secure cloud framework to share EHRs using modified CP-ABE and the attribute bloom filter. Educ Inf Technol 23, 2213–2233 (2018). https://doi.org/10.1007/s10639-018-9713-7

[7] Nagarjuna T., Nehru K., Nagendra Prasad G., Menakadevi N., Smart sensor network based high quality air pollution monitoring system using labview, International Journal of Online Engineering, 2017, 13, 8, 79-87, 10.3991/ijoe.v13i08.7161

[8] Indira DNVSLS, Ganiya RK, Ashok Babu P, Xavier AJ, Kavisankar L, Hemalatha S, Senthilkumar V, Kavitha T, Rajaram A, Annam K, Yeshitla A. Improved Artificial Neural Network with State Order Dataset Estimation for Brain Cancer Cell Diagnosis. Biomed Res Int. 2022 Apr 16;2022:7799812. doi: 10.1155/2022/7799812. PMID: 35480141; PMCID: PMC9038414.

[9] Radhakrishna, V., Kumar, P.V., Janaki, V., Rajasekhar, N. (2017). Estimating Prevalence Bounds of Temporal Association Patterns to Discover Temporally Similar Patterns. In: Matoušek, R. (eds) Recent Advances in Soft Computing. ICSC-MENDEL 2016. Advances in Intelligent Systems and Computing, vol 576. Springer, Cham. https://doi.org/10.1007/978-3-319-58088-3_20

[10] Kalyani G., Janakiramaiah B., Karuna A., Prasad L.V.N., Diabetic retinopathy detection and classification using capsule networks, Complex and Intelligent Systems, 2023, 10.1007/s40747-021-00318-9

[11] A. Cheruvu, V. Radhakrishna and N. Rajasekhar, "Using normal distribution to retrieve temporal associations by Euclidean distance," 2017 International Conference on Engineering & MIS (ICEMIS), Monastir, Tunisia, 2017, pp. 1-3, doi: 10.1109/ICEMIS.2017.8273101.

[12] Awasthi, Ankita, and Kuldeep K. Saxena. "Evaluation of mechanical properties of orange peel reinforced epoxy composite." *Materials Today: Proceedings* 18 (2019): 3821-3826.

[13] Bisht, Pankaj Singh, and Ankita Awasthi. "Design and Analysis of Composite and Al Alloy Wheel Rim." In *Advances in Materials Engineering and Manufacturing Processes: Select Proceedings of ICFTMM 2019*, pp. 15-29. Springer Singapore, 2020.

[14] Awasthi, Ankita, Kuldeep K. Saxena, and Vanya Arun. "Sustainability and survivability in manufacturing sector." In *Modern Manufacturing Processes*, pp. 205-219. Woodhead Publishing, 2020.

[15] Bisht, Pankaj Singh, and Ankita Awasthi. "Analysis of E-glass fiber wheel rim by using ANSYS." In *Recent Advances in Mechanical Engineering: Select Proceedings of ITME 2019*, pp. 79-91. Springer Singapore, 2021.

[16] Awasthi, Ankita, Kuldeep K. Saxena, and Ravi K. Dwivedi. "An investigation on classification and characterization of bio materials and additive manufacturing techniques for bioimplants." *Materials Today: Proceedings* 44 (2021): 2061-2068.

[17] Awasthi, Ankita, Kuldeep K. Saxena, and Vanya Arun. "Sustainable and smart metal forming manufacturing process." *Materials Today: Proceedings* 44 (2021): 2069-2079.

[18] Awasthi, Ankita, Akash Gupta, Kuldeep K. Saxena, and Ravi K. Diwedi. "Equal channel angular processing on aluminium and its alloys–A review." *Materials Today: Proceedings* 56 (2022): 2388-2391.

[19] Awasthi, Ankita, U. Sathish Rao, Kuldeep K. Saxena, and Ravi K. Dwivedi. "Impact of equal channel angular pressing on aluminium alloys: An overview." *Materials Today: Proceedings* 57 (2022): 908-912.

[20] Awasthi, Ankita, Kuldeep K. Saxena, R. K. Dwivedi, Dharam Buddhi, and Kahtan A. Mohammed. "Design and analysis of ECAP Processing for Al6061 Alloy: a microstructure and mechanical property study." *International Journal on Interactive Design and Manufacturing (IJIDeM)* (2022): 1-13.

[21] Awasthi, Ankita, Akash Gupta, Kuldeep K. Saxena, R. K. Dwivedi, Deepak Kundalkar, Dalael Saad Abdul-Zahra, Abhishek Joshi, and H. S. Saggu. "Design and analysis of equal-channel angular pressing of Al6061: a comparative study." *Advances in Materials and Processing Technologies* (2022): 1-10.

[22] Tripathi, Gyan Prakash, Sumit Agarwal, Ankita Awasthi, and Vanya Arun. "Artificial Hip Prostheses Design and Its Evaluation by Using Ansys Under Static Loading Condition." In *Biennial International Conference on Future Learning Aspects of Mechanical Engineering*, pp. 815-828. Singapore: Springer Nature Singapore, 2022.

[23] Arun, V., N. K. Shukla, A. K. Singh, and K. K. Upadhyay. "Design of all optical line selector based on SOA for Data Communication: Proceedings of the Sixth International Conference on Computer and Communication Technology 2015." In *ACM Other conferences*. 2015.

[24] Arun, Vanya, Ashutosh Kr Singh, N. K. Shukla, and D. K. Tripathi. "Design and performance analysis of SOA–MZI based reversible

toffoli and irreversible AND logic gates in a single photonic circuit." *Optical and quantum electronics* 48 (2016): 1-15.

[25] Arun, Vanya, Kapil Deo Bodha, Awadhesh K. Maurya, and Ashutosh K. Singh. "Design and implementation of all optical processing units together performing arithmetic and logical functions." In *VLSI, Microwave and Wireless Technologies: Select Proceedings of ICVMWT 2021*, pp. 83-93. Singapore: Springer Nature Singapore, 2022.

[26] Arora, Gurmeet Singh, and Kuldeep Kumar Saxena. "A review study on the influence of hybridization on mechanical behaviour of hybrid Mg matrix composites through powder metallurgy." *Materials Today: Proceedings* (2023).

[27] Swapna Sri, M. Naga, P. Anusha, VV venu Madhav, Kuldeep Kumar Saxena, Ch Sri Chaitanya, R. Haranath, and Bharat Singh. "Influence of Cu particulates on a356mmc using frequency response function and damping ratio." *Advances in Materials and Processing Technologies* (2023): 1-9.

[28] Bodha, Kapil Deo, V. Mukherjee, and Vinod Kumar Yadav. "A player unknown's battlegrounds ranking based optimization technique for power system optimization problem." *Evolving Systems* 14, no. 2 (2023): 295-317.

[29] Alonso-Betanzos, Amparo, and Verónica Bolón-Canedo. "Big-data analysis, cluster analysis, and machine-learning approaches." Sex-specific analysis of cardiovascular function (2018): 607-626.

[30] Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. "Machine learning on big data: Opportunities and challenges." Neurocomputing 237 (2017): 350-361.

[31] Gomes, Heitor Murilo, Jean Paul Barddal, Fabrício Enembreck, and Albert Bifet. "A survey on ensemble learning for data stream classification." ACM Computing Surveys (CSUR) 50, no. 2 (2017): 1-36.

[32] Dong, Xibin, Zhiwen Yu, Wenming Cao, Yifan Shi, and Qianli Ma. "A survey on ensemble learning." Frontiers of Computer Science 14 (2020): 241-258.

[33] Roh, Yuji, Geon Heo, and Steven Euijong Whang. "A survey on data collection for machine learning: a big data-ai integration perspective." IEEE Transactions on Knowledge and Data Engineering 33, no. 4 (2019): 1328-1347.

[34] Vishnu, Vandana Kolisetty, and Dharmendra Singh Rajput. "A review on the significance of machine learning for data analysis in big data." Jordanian Journal of Computers and Information Technology 6, no. 1 (2020).

[35] N. Krishnachaithanya et al., "People Counting in Public Spaces using Deep Learning-based Object Detection and Tracking Techniques," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 784-788, doi: 10.1109/CISES58720.2023.10183503.

[36] S. A. Yadav, S. Sharma and S. R. Kumar, "A robust approach for offline English character recognition," 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Greater Noida, India, 2015, pp. 121-126, doi: 10.1109/ABLAZE.2015.7154980.

[37] Yadav, SA, Poongodi, T. A novel chain-based clustering for green communication in wireless sensor network. Int J Commun Syst. 2023; 36(13):e5523. doi:10.1002/dac.5523

[38] Tushar, R. K. Patel, E. Aggarwal, K. Solanki, O. Dahiya and S. A. Yadav, "A Logistic Regression and Decision Tree Based Hybrid Approach to Predict Alzheimer's Disease," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 722-726, doi: 10.1109/CISES58720.2023.10183466.

[39] D. Singh, H. S. Garjan, S. V. Singh and G. Bhardhwaj, "A Novel Optimization Technique for Integrated Supply Chain Network in Industries - A Technical Perspective," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2021, pp. 550-554, doi: 10.1109/ICIEM51511.2021.9445392.

[40] G. Bhardwaj, R. Gupta, A. Pratap Srivastava and S. Vikram Singh, "Cyber Threat Landscape of G4 Nations: Analysis of Threat Incidents & Response Strategies," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2021, pp. 75-79, doi: 10.1109/ICIEM51511.2021.9445307.

[41] A. Chaturvedi, T. V. Kumar, A. P. Srivastava, S. K. Shukla, S. V. Singh and A. Parvez, "Energy Optimization in Wireless Sensor Networks," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 700-704, doi: 10.1109/ICTACS56270.2022.9987887.