International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Review on Workload Characterization Methodology Using Supervised and Unsupervised Deep Learning

Aniket Dattatrey Deshmukh[1], Dr N. S. Bagal[2]

[1,2]Department of Computer Engineering, PVPIT Bavdhan, Pune

Abstract
Effective workload characterization and prediction are crucial for enhancing system performance, scalability, and resource usage in today's dynamic computing environments. In order to accomplish precise workload prediction and intelligent resource management, this project proposes a Workload Characterization Methodology utilizing CNN-LSTM that combines statistical analysis, clustering, deep learning, and optimization techniques. To make sure that only pertinent characteristics are taken into account, the procedure starts with data preparation and feature selection using Pearson Correlation. Fuzzy C-Means Clustering is then used to group the chosen features, successfully identifying comparable workload patterns. To improve prediction accuracy, a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) model is used to incorporate both temporal and spatial correlations in the workload data. In order to provide equitable workload distribution and effective resource use, the Hungarian Model is finally used for optimal job allocation. Applications in cloud computing, data centers, edge systems, and large-scale distributed environments can benefit from the suggested methodology's enhanced predictive capability, adaptability, and decision-making efficiency.

Keywords: CNN-LSTM , Supervised learning , deep learning

I. INTRODUCTION

In today's computer environments, workload characterization is essential because of the fast expansion of data and the dynamic behavior of systems, which necessitates the use of clever methods to recognize, categorize, and forecast workload trends. When dealing with high-dimensional, noisy, and constantly changing datasets, traditional analytical techniques frequently fail. This paper suggests a hybrid workload characterization methodology that combines supervised and unsupervised deep learning techniques in order to overcome these issues.

To make sure that only the most pertinent characteristics are chosen for additional analysis, the methodology starts with methodical data preprocessing and Pearson correlation-based feature filtering. Fuzzy C-Means (FCM) clustering is used in unsupervised learning to group comparable workload behaviors and uncover hidden patterns in the dataset. High-accuracy performance prediction is made possible by simultaneously learning temporal and spatial relationships in the workload data using a CNN-LSTM deep learning model.

The Hungarian Model then combines the clustered outputs and supervised predictions to optimize task assignment and match projected workload categories with real resource needs. Lastly, a strong and trustworthy workload prediction is generated by the integrated decision-making module, which can help with efficient scheduling, resource allocation, and system optimization.

This hybrid framework guarantees enhanced interpretability, accuracy, and adaptability, which makes it appropriate for large-scale distributed systems, cloud environments, and data centers where operational effectiveness depends on an understanding of workload characteristics.

In contemporary computing systems, workload characterization has grown essential, particularly as cloud platforms, massive data centers, and intelligent resource-driven settings become more complicated. Understanding how apps behave is crucial for guaranteeing good performance, effective scheduling, optimal resource use, and overall system stability since they produce enormous and varied workloads. However, problems like nonlinear patterns, high dimensionality, and hidden correlations in the data make it difficult for traditional statistical or rule-based approaches to effectively describe dynamic workload variations. Fuzzy C-Means Clustering (FCM) is used to arrange comparable workload patterns into clusters based on shared characteristics once the most influential features have been chosen. FCM offers greater flexibility and accuracy in workload grouping than hard clustering techniques since it permits each data point to belong to several clusters with differing degrees of membership. The CNN-LSTM hybrid model, which combines the sequential learning capabilities of Long Short-Term Memory (LSTM) networks with the feature extraction capability of Convolutional Neural Networks (CNN), is then fed these clustered features. More accurate workload forecasts result from the system's ability to identify temporal correlations and spatial dependencies in the workload data.

By effectively matching anticipated workloads with existing computational resources, the Hungarian Model is integrated to optimize task allocation. By distributing workloads evenly among nodes or processors, this architecture reduces bottlenecks and boosts system performance. Lastly, the decision-making module makes intelligent workload forecasts using the CNN-LSTM and Hungarian Model outputs, enabling proactive system management and dynamically balancing computing loads.

All things considered, this approach provides a strong and flexible framework for workload characterization that improves prediction accuracy, maximizes resource use, and encourages wise decision-making. It helps create more intelligent, self-aware computing environments that can manage intricate and quickly evolving workloads by utilizing cutting-edge machine learning and optimization approaches.

Workload characterization has become an essential procedure for enhancing the effectiveness, dependability, and scalability of computational systems in the age of large-scale computing, cloud infrastructures, and data-intensive applications. Understanding the behavior of a workload—which is the collection of activities, procedures, or tasks carried out by a computing environment—allows system designers to make well-informed choices about scheduling, resource allocation, and performance optimization. However, because workloads are complex, time-varying, and interdependent in today's dynamic and diverse contexts, traditional workload characterization methods that rely on static analysis and linear models are insufficient.

This study presents an integrated workload characterization methodology that integrates deep learning models, clustering techniques, statistical analysis, and optimization algorithms to address these issues. By examining multi-dimensional data patterns, the suggested approach guarantees precise workload prediction, ideal task distribution, and astute decision-making.

The dataset, which forms the basis for analysis, is where the process starts. Performance indicators, job parameters, and system usage statistics gathered from actual or simulated computer environments are frequently included in this collection. Cleaning inconsistencies, resolving missing values, standardizing numerical ranges, and converting categorical variables into understandable representations are all crucial steps in the pre-processing stage of data preparation. In addition to improving data quality, appropriate pre-processing keeps noise and bias from influencing analysis later on.

The linear relationships between various workload variables are then measured using Pearson Correlation Analysis. This stage aids in discovering the factors that have a substantial impact on workload behavior by measuring the degree of correlation between features. The next step is feature selection, which lowers dimensionality and computing overhead by removing attributes that are redundant, unnecessary, or weakly associated. By ensuring that only the most discriminative and instructive characteristics are kept for further modeling stages, feature selection enhances interpretability and learning performance.

[1] Baekgyu Kim et al. Narrate author's empirical analysis indicates that understanding the edge server's composite load requires accounting for the intricate interaction between traffic flow (e.g., volume, density, speed) and local load (e.g., periodic, speed-dependent offloading). Even when vehicles generate the same type of local loads, the composite loads on edge servers can vary significantly based on the speed, number of vehicles, and the size of the area they cover. Geographically adjacent edge servers often exhibit similar peak load patterns; however, the magnitudes and timings of these peaks can differ depending on how traffic flow diverges or converges between areas. Building on these findings, author aim to develop a theoretical model of the edge server's composite load. This model will mathematically capture its characteristics, enabling formal analyses such as end-to-end latency analysis, resource allocation, and schedulability tests.

[2] Mohammad Newaj Jamil et al. Proposed Multi-access Edge Computing (MEC) is a standard network architecture of edge computing that has emerged to handle a massive amount of diverse data traffic, enhance computation capabilities, and reduce communication latency. However, there exist several critical challenges during workload orchestration in MEC. Firstly, a task generated by delaysensitive or delay-tolerant applications should appropriately be distinguished when determining the optimal edge/cloud server for processing that task since task requirements for delay-sensitive and delay-tolerant applications are different. Secondly, due to the high mobility of end-users, unknown numbers of offloading requests from different types of UE, and intermittent data traffic, the state of the network condition keeps changing, and computational capacities of edge and cloud servers keep fluctuating, which arises a VOLUME 11, 2023 118021 M. N. Jamil et al.: Workload Orchestration in Multi-Access Edge Computing rapidly changing dynamic environment at the network edge due to uncertainty. Finally, if offloaded tasks from UE are distributed unevenly and randomly within the overall MEC infrastructure, it will lead to imbalance workloads between edge and cloud servers, increase task latency due to improper task offloading, and impair the overall performance and efficiency of MEC. In order to address these workload orchestration challenges in MEC, a BRB workload orchestrator is proposed in this study. The proposed BRB workload orchestrator uses belief rules to specify the required workload orchestration operations with regard to network conditions, computational resources, and properties of offloaded tasks from UE and decide the optimal execution location for each offloaded task within the overall MEC infrastructure. The performance of the proposed BRB workload orchestrator is compared with four workload orchestration approaches from the literature under various workloads in regard to three performance metrics. According to the result of simulation experiments, the proposed workload orchestrator outperforms other approaches in aspects of task failure rate, average service time, and average VM utilization on edge servers. Under various workloads, it reduces overall service time, minimizes task

failures, and provides lower average VM utilization by efficiently utilizing all VMs on edge servers and balancing the workload among edge servers. In this study, tasks generated by the four types of applications are considered independent tasks or stateless jobs. Besides, full computation offloading is considered, which means a whole task from UE is offloaded to an edge or cloud server for processing without any split of the task. In future studies, the dependency between tasks and partial computation offloading can be considered, where after splitting a task into sub-tasks with an optimal task splitting strategy, some sub-tasks can be executed locally by UE if it has enough processing capacity, while the remaining sub-tasks can be offloaded to an edge and/or cloud servers for further processing. The proposed BRB workload orchestrator can be enhanced by considering partial computation offloading, which can further minimize task failures and reduce overall service time for delay-sensitive and resource-intensive applications through a collaboration among UE, Local Edge, Remote Edge, and Cloud servers and improve the overall performance and efficiency of MEC.
[3] Jia Yunfei et al. Present the nonlinear aging phenomenon observed from a web server with varying workloads, has been analyzed. There is an obvious relationship between the aging speed and the system's workload. The aging speed is calculated by a robust algorithm to validate the trend of resource leak. To address the forecasting problems caused by the unstable and nonlinear time series, author have employed the prior information of workload, which has allowed us to accurately estimate the aging slope and to further forecast the system resource exhaustion time. The workload is used as a threshold to tell us which model can be used when forecasting system resource exhaustion. Then, the evolvement of the system resources is described by a TAR model. The accuracy of author's approach is compared with that of the AR model to show the advantage of author's approach. author's approach can be easily implemented and can be applied to any service-oriented applications.

II RELATED WORKS

[1] BAEKGYU KIM et al. Narrate author's empirical analysis indicates that understanding the edge server's composite load requires accounting for the intricate interaction between traffic flow (e.g., volume, density, speed) and local load (e.g., periodic, speed-dependent offloading). Even when vehicles generate the same type of local loads, the composite loads on edge servers can vary significantly based on the speed, number of vehicles, and the size of the area they cover. Geographically adjacent edge servers often exhibit similar peak load patterns; however, the magnitudes and timings of these peaks can differ depending on how traffic flow diverges or converges between areas. Building on these findings, author aim to develop a theoretical model of the edge server's composite load. This model will mathematically capture its characteristics, enabling formal analyses such as end-to-end latency analysis, resource allocation, and schedulability tests.
[2] MOHAMMAD NEWAJ JAMIL et al. Proposed Multi-access Edge Computing (MEC) is a standard network architecture of edge computing that has emerged to handle a massive amount of diverse data traffic, enhance computation capabilities, and reduce communication latency. However, there exist several critical challenges during workload orchestration in MEC. Firstly, a task generated by delay sensitive or delay-tolerant applications should appropriately be distinguished when determining the optimal edge/cloud server for processing that task since task requirements for delay-sensitive and delay-tolerant applications are different. Secondly, due to the high mobility of end-users, unknown numbers of offloading requests from different types of UE, and intermittent data traffic, the state of the network condition keeps changing, and computational capacities of edge and cloud servers keep

fluctuating, which arises a VOLUME 11, 2023 118021 M. N. Jamil et al.: Workload Orchestration in Multi-Access Edge Computing rapidly changing dynamic environment at the network edge due to uncertainty. Finally, if offloaded tasks from UE are distributed unevenly and randomly within the overall MEC infrastructure, it will lead to imbalance workloads between edge and cloud servers, increase task latency due to improper task offloading, and impair the overall performance and efficiency of MEC. In order to address these workload orchestration challenges in MEC, a BRB workload orchestrator is proposed in this study. The proposed BRB workload orchestrator uses belief rules to specify the required workload orchestration operations with regard to network conditions, computational resources, and properties of offloaded tasks from UE and decide the optimal execution location for each offloaded task within the overall MEC infrastructure. The performance of the proposed BRB workload orchestrator is compared with four workload orchestration approaches from the literature under various workloads in regard to three performance metrics. According to the result of simulation experiments, the proposed workload orchestrator outperforms other approaches in aspects of task failure rate, average service time, and average VM utilization on edge servers. Under various workloads, it reduces overall service time, minimizes task failures, and provides lower average VM utilization by efficiently utilizing all VMs on edge servers and balancing the workload among edge servers. In this study, tasks generated by the four types of applications are considered independent tasks or stateless jobs. Besides, full computation offloading is considered, which means a whole task from UE is offloaded to an edge or cloud server for processing without any split of the task. In future studies, the dependency between tasks and partial computation offloading can be considered, where after splitting a task into sub-tasks with an optimal task splitting strategy, some sub-tasks can be executed locally by UE if it has enough processing capacity, while the remaining sub-tasks can be offloaded to an edge and/or cloud servers for further processing. The proposed BRB workload orchestrator can be enhanced by considering partial computation offloading, which can further minimize task failures and reduce overall service time for delay-sensitive and resource-intensive applications through a collaboration among UE, Local Edge, Remote Edge, and Cloud servers and improve the overall performance and efficiency of MEC.

[3] JIA Yunfei et al. Present the nonlinear aging phenomenon observed from a web server with varying workloads, has been analysed. There is an obvious relationship between the aging speed and the system's workload. The aging speed is calculated by a robust algorithm to validate the trend of resource leak. To address the forecasting problems caused by the unstable and nonlinear time series, author have employed the prior information of workload, which has allowed us to accurately estimate the aging slope and to further forecast the system resource exhaustion time. The workload is used as a threshold to tell us which model can be used when forecasting system resource exhaustion. Then, the evolvement of the system resources is described by a TAR model. The accuracy of author's approach is compared with that of the AR model to show the advantage of author's approach. author's approach can be easily implemented and can be applied to any service-oriented applications.

[4] ZOHRA AMEKRAZ et al. Narrate author investigated the presence of chaos in cloud workloads. author used chaotic time series analysis to identify the chaotic behavior of the workloads and proposed an improved ANFIS model called the CANFIS model to predict future workloads. The inputs of the proposed model are selected using the phase space reconstruction method and embedding theory. Using real-world web and cluster workloads, author performed experiments to evaluate the performance of the proposed model. Every workload used in these experiments went through several preprocessing operations before being fed to the proposed model. These operations include extraction, aggregation, cleaning, and

normalization. Cleaning was performed using the SG filter, which can effectively eliminate noise and outliers without changing the width and the peak of the original signal. The results of the experiments showed that the proposed CANFIS model achieved higher prediction accuracy than the ANFIS, LSTM, SVM and ARIMA models and some other new neural network-based methods. In addition, the proposed model has a low training time, meaning that it will not have any effect on the provisioning process.

[5] DALAL ALQAHTANI et al. Demonstrate The proposed CVCBM model offers a highly effective solution for improving workload prediction in cloud data centers. By integrating advanced signal processing techniques—such as CEEMDAN and VMD—with deep learning approaches like CNN and Bi-LSTM, the model effectively addresses noise and captures the complex, dynamic patterns in workload data. A hierarchical two-stage decomposition, supported by partitional clustering based on Sample Entropy, further enhances the model's ability to identify trends at multiple scales. This includes a novel architecture of parallel Conv1D layers, designed to extract patterns ranging from long-term to short-term in workload traces, followed by Bi-LSTM layers that capture long-term temporal dependencies. As a result, CVCBM accurately predicts future workloads across various time scales. Extensive experimentation on real-world datasets demonstrates that CVCBM outperforms traditional methods overall, resulting in substantial prediction accuracy gains for cloud data centers. This robust and promising approach is very much suited to resource management in dynamic cloud environments. Future work will focus on extracting additional workload characteristics to enhance prediction accuracy. author also plan to explore reinforcement learning algorithms for resource allocation in CDCs. Moreover, author aim to investigate performing tasks locally on mobile or IoT devices—engaging the cloud only when necessary—to reduce the workload in CDCs.

[6] Genoveva Vargas-Solar et al. Narrate This paper introduced JITA-4DS, a virtualized architecture that provides a disaggregated data center solution ad-hoc for executing DS pipelines requiring elastic access to resources. DS pipelines process big streams and data coordinating operators implemented by services deployed on edge. Given that operators can implement greedy tasks with computing and storage requirements beyond those residing on edge, they interact with VDC services. author have set the first simulation setting to study resources delivery in JITA-4DS. author are currently addressing challenges of VDCs management in simpler environments, on cloud resource management heuristics, big data analysis, and data mining for performance prediction. author give the first experimental results of these aspects. To simulate, evaluate, analyze, and compare different heuristics, author will further build simulators for simpler environments and combine open-source simulators for different levels of the JITA-4DS hierarchy. author are currently defining a "benchmark" with different types of data science workloads that share data collections and functions and study complex resource allocation patterns.

[7] LUYAO LIU et al. Proposed The power consumption and carbon emission problems of data center has attracted worldwide attention with the rapid advance of AI technologies and digital economy, and the integration of data centers with low-carbon renewable wind and solar energy is becoming a promising strategy. Aiming at the volatilities of renewable power, the delay-tolerant workloads within data centers are considered as one of most potential sources for power regulation. In light of the deficiencies of systematic analysis to delay-tolerant workloads in existing studies, the paper clarifies the temporal shift mechanisms of three typical delay-tolerant workloads, i.e. short-running deferrable, long-running continuous, long-running interruptible workloads, and builds the temporal shift models for them to participate in the day-ahead power scheduling of data center micro-grid. By applying the developed models into a data center case across various scenarios, results confirm the effectiveness of the developed

energy management scheme and the efficiency of the time-shiftable workloads in improving the operation economy of data center micro-grid and increasing the utilization rate of renewable energy. The proposed time shift theory and power regulation models can be applied into further planning and operation of other form of data center energy systems, such as data center integrated energy system, which is expected to increase the system benefits and adoption of renewable energy, ultimately accelerating the low-carbon transition of the data center industry.

[8] RUIYUN LIU et al. Present a workload based model for optimal planning of geo-distributed data centers in fast developing economies. Instead of modeling the need by the aggregated number of servers, the workload based model enables more flexible resource allocation and can better capture the quickly changing nature of demand composition. It is capable of characterizing the diversified application scenarios and more suitable for the planning of geo-distributed data centers in fast developing economies. author transform the problem into a quadratic programming problem and provide approaches to solve it. By taking into account factors that are eminent in economies like China, the model can achieve considerable cost-saving for long-term planning. For example, 5.8% in cost can be saved in a five-year planning period. With real-life data, author show that the constant evolution of workload composition and their requirements for computing and bandwidth resources have significant influence on the distribution of data centers. author also illustrate that population migration lead to unbalanced load in data centers and may have significant impact on the performance of network services. 5% population migration may lead to 30ms increase in latency. author's study provides a new perspective and useful insights in data center planning in fast developing economies. Note that the factors author consider in this paper are at a macro-level, and should be regarded as ‘‘slow changing’’ over time, at a time span of a few months or even years. This differentiates author's work from existing studies with a micro-level perspective, for instance, scheduling of workflows within a day or even shorter time spans. However, it would be interesting to study how long and short term factors may inter-play in data center planning, and how requirements such as high availability and disaster tolerance may affect the planning results, in the context of fast developing economies.

[9] Junlong Li et al. Narrate This paper develops a 4-stage STR method to reduce electricity costs in EDC clusters via reallocating workload and ESS capacity between EDCs. The proposed method is demonstrated in an EDC cluster and the results testify to the advantages of the proposed method in reducing electricity costs with strong robustness. Through extensive demonstration, the key observations are: • Limited literature studied workload migration between EDCs. This work is the first of its kind, it illustrates that, with a rational optimisation method and coordination with ESS, workload migration between EDCs is a practical way to deliver the benefits of energy cost reduction. • A Bit-Watt transformation, by directly revealing the relationship betweenpower consumption and the amount of computing, reduces decision variables in the workload migration optimization. This could build a mathematical foundation for hybrid modeling of energy systems and communication systems. • A comprehensive power control method, the STR method, is developed to fully utilize existing resources in EDC clusters to reduce electricity costs. This method consists of 4 stages to provide efficient services. Decoupling of spatial and temporal dimensions in stages 1–3 reduces the dimension of decision variables. The rolling adjustment method in stage 4 enhances robustness of the STR method under high uncertaintiesin workload forecasting.

[10] Chao Guo et al. Demonstrate author have considered the workload consolidation problem in CDCs, aiming to minimize the number of active nodes and the number of migrated workload elements. author

have proposed a Q-learning-based algorithm to generate an approximate Pareto front. author have also developed an ILP formulation to validate the performance of the proposed algorithm. The ILP Pareto fronts in various test cases have highlighted the superior consolidation performance of the CDCs over the SDCs in terms of both energy efficiency savings and migration cost reduction. The numerical results have shown that the Q-learning-based algorithm closely approximates the ILP Pareto front while overcoming its key drawback of high complexity. Moreover, by carefully designing modifications to the traditional Q-learning method, especially the dynamic action space, author's proposed method significantly outperforms an SA-based method. Finally, author's Q-learning-based algorithm also overcomes the drawback of the heuristic algorithms, FF and FFD, which do not consider migration costs. Accordingly, author have demonstrated significant performance improvement of author's Q-learning-based algorithm over FF and FFD. It is worth noting that although the proposed Q-learning-based approach is designed for the CDC architecture, it is foreseen that GUO the method can also be applicable to the SDC architecture. In the SDC architecture, the approach can be adapted by considering only computing nodes while ignoring memory nodes and other node types. In addition, the VN demand of each request can be reduced to contain only one element (e.g., VM) instead of multiple ones. However, this work still faces several limitations. One limitation is the assumption of known and constant workload demands. In practice, workload demands can change dynamically, necessitating real-time adaptation. While accurate prediction methods have been studied, achieving high prediction accuracy remains challenging. Inaccurate predictions can lead to decreased efficiency and service degradation. Therefore, it is important to explore strategies for handling inaccurate predictions to mitigate these negative effects. Another possible future work is collaborating with industry partners and deploying author's proposed algorithm in real CDC environments. This will enable a further demonstration and validation of the effectiveness and feasibility of author's proposed method.

[11] ANDREW BABAKIAN et al. Narrate This paper highlighted how each generation of network architectures addressed a challenge created by evolved application artifacts. It highlights that cloud-native applications have shifted identifiers from the IP layer and have become an attribute of the application layer used to distinguish workloads and provide location-independent identification schemes. In addition, this paper reviewed the current state-ofthe-art application deployment patterns and the future identifier challenges that may arise. This is an important topic that requires further investigation into three topical areas. Firstly, evaluate existing persistent identifiers in the digital space that could be applied for broader use. Regardless of location, application instantiations should be referred to as the single identifier. Secondly, the verification process, where credentials are analyzed to provide identity attestation and issuance of its imprimatur, should be consistent across any cloud and edge computing platform. Lastly, the delivery method to control and disseminate trust anchors to allow application communication between trust realms will underpin the success of a common identity system. It is hoped this retrospective of identifiers may influence the future design of identifiers and locators used in cloud-native architectures.

[12] Shijie Zhang et al. Present MorphDAG realizes an unexplored workload-aware DAGbased blockchain for elastic storage and transaction processing. Based on the elastic degree of storage concurrency theory, MorphDAG achieves an optimal level of throughput under load variations while guaranteeing system security. Further, a dual-mode mechanism is proposed to support efficient transaction processing under skewed data access. author show that MorphDAG outperforms state-of-the-art DAG-based approaches in throughput and latency. author's sauthor'sce code of the MorphDAG prototype will be available at https://github.com/CGCL-codes/MorphDAG.

[13] NOSIN IBNA MAHBUB et al. Constructed This research addresses the potential risks associated with adversarial attacks targeting cloud workload prediction methods. To the best of author's understanding, there is a lack of research conducted on adversarial attacks against cloud workload forecasting models. Existing research has solely concentrated on forecasting the future workload in cloud data centers. To investigate the robustness and security of predictive models, author therefore focus on the adversarial attack on workload prediction. The experimental results of this study demonstrate that all state-of-the-art workload forecasting models are susceptible to adversarial attacks, which can have disastrous security implications for cloud data centers. In the future, author will focus on the development of enhanced adversarial attack methods specifically tailored for cloud workload data. This is due to the fact that changes in cloud workload data are more perceptible to human observation compared to alterations in picture data. The notion that a change in picture usage is imperceptible to the human eye remains inaccurate in the context of cloud workload data. Furthermore, author's research will encompass an investigation of defense mechanisms aimed at identifying and mitigating hostile risks inside deep-learning forecasting models.

[14] BING HU et al. Narrate The deep learning workload analysis tool (DLWAT) is a data mining tool for characterizing workloads, as demonstrated by the SPEC CPU2006 and CPU2017 benchmark workloads, using supervised and unsupervised deep learning technologies. Comparison studies of supervised study cases showed that CRNN models had the highest accuracy compared with CNN and RNN models. The unsupervised comparison studies have demonstrated that the deep clustering model outperformed standard k-means clustering approach. Both supervised and unsupervised deep learning approaches can capture the essence of these features in latent representations, which produces better classification results. Currently, DLWAT has been used on SPEC CPU benchmark workloads; it has not yet been extended to other workloads such as gaming workloads. Looking forward, author foresee a potential opportunity to study more complex real-life workloads, such as gaming workloads. In the future, the deep learning workload analysis tool could be migrated into a portable stand-alone library to enrich the software community’s capability to analyze complex workloads and provide insights from machine learning perspectives.

[15] ANDREA ALAIMO et al. Proposed This study can be considered a contribution to workload assessment using physiological measures. Specifically, considering time-domain, frequency-domain, and non-linear HRV indexes in a unique experimental design, these findings offer a more comprehensive overview of the different HRV indexes for workload evaluation. These results support the importance of parameters based on HRV measurement for assessing changes in workload, also suggesting the development of non-obtrusive devices that can assess workload in real-time.

III PROPOSED METHODOLGY

The proposed approach for the purpose of achieving A Workload Characterization Methodology Using Supervised and Unsupervised Deep Learning has been depicted in the figure 1 above and the steps taken to achieve this system are elaborated below.

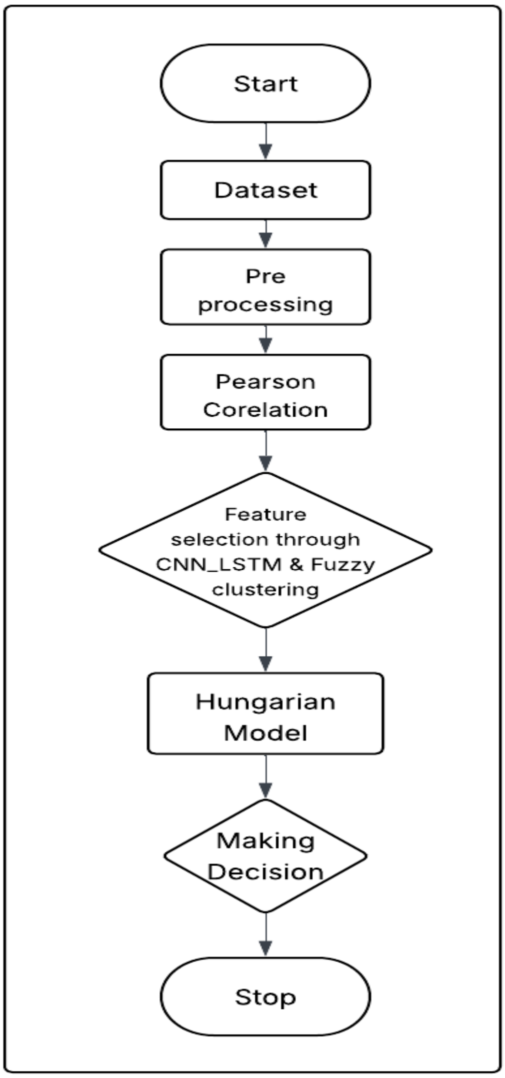

Figure 1: Proposed Methodology

We begin by collecting raw tweet data from publicly available sources (e.g. a labeled Twitter dataset), ensuring each tweet has an identifier, text content, and a ground-truth label (e.g. "cyber bullying" vs "non-cyberbullying"), while preserving any relevant metadata such as timestamps or user info. Next, all tweets undergo preprocessing: the text is cleaned and normalized (special characters, punctuation, URLs/mentions/hash tags are removed or normalized; text is lowercased), then tokenized, with stop-words removed and remaining words stemmed or lemmatized to their base form — reducing noise and standardizing the input for downstream modeling. After preprocessing, the clean tweets are converted into numerical feature representations: this could be a simple Bag-of-Words (BoW) vector, a TF-IDF weighted vector capturing word importance across the corpus, or dense word-embeddings (e.g. pre-trained word vectors) that encode semantic relationships among words. Once feature vectors are ready, you may optionally apply clustering (or soft clustering, e.g. fuzzy clustering) to group similar tweets — for instance clustering by linguistic style, severity, or topical similarity of abuse — which can help reveal latent structure before classification. Finally, a classification model (in your case a hybrid deep-learning architecture such as CNN–LSTM) is trained on the feature vectors (or embeddings) to distinguish abusive

/ cyber bullying tweets from benign ones, leveraging the ability of CNNs to capture local word-pattern features (e.g. n-grams) and of LSTM to model sequential context and long-range dependencies.

V CONCLUSION AND FUTURE SCOPE

This work presents a comprehensive methodology for detecting cyber bullying from social-media data (tweets), integrating data preprocessing, feature extraction/selection, clustering, and deep learning-based classification to build a functional end-to-end pipeline. Starting from raw tweet data, our preprocessing phase cleans and normalizes the text (removing noise, stemming, removing stop-words), which ensures that subsequent analysis operates on high-quality, relevant data. We then convert cleaned text into meaningful features (e.g. using a Bag-of-Words representation, possibly enhanced with embeddings), allowing effective modeling. To account for variability and latent structure in abusive language, we optionally group similar tweets via soft clustering (e.g. Fuzzy C-Means), enabling either cluster-specific modeling or enriched feature context. At the core, a hybrid CNN-LSTM model leverages both local text patterns and sequential context to classify tweets as cyber bullying or not. Finally, a decision module — possibly with a task-assignment or workload-prediction component (e.g. via a Hungarian-style assignment algorithm) — transforms model outputs into actionable moderation or review work.

Overall, this methodology achieves a balance between robust text-processing, powerful classification, and practical applicability (task assignment / workflow support). It demonstrates how modern NLP and machine-learning tools can be combined to address the challenge of cyber bullying detection on social media, offering a scalable and automated solution. At the same time, by clearly structuring each stage — from preprocessing to decision-making — the system remains modular and extensible, making it possible to refine or replace individual components (e.g. swap Bag-of-Words for embeddings, change the clustering algorithm, or update the classification architecture) without reworking the entire pipeline.

REFERENCES

1. B. Kim and D. Gangadharan, "Edge-Server Workload Characterization in Vehicular Computation Offloading: Semantics and Empirical Analysis," in IEEE Access, vol. 12, pp. 89082-89097, 2024, doi: 10.1109/ACCESS.2024.3419156.
2. M. N. Jamil, M. S. Hossain, R. U. Islam and K. Andersson, "Workload Orchestration in Multi-Access Edge Computing Using Belief Rule-Based Approach," in IEEE Access, vol. 11, pp. 118002-118023, 2023, doi: 10.1109/ACCESS.2023.3326244.
3. J. Yunfei, Z. Zhiquan and W. Renbiao, "A workload-based nonlinear approach for predicting available computing resources," in Journal of Systems Engineering and Electronics, vol. 31, no. 1, pp. 224-230, Feb. 2020, doi: 10.21629/JSEE.2020.01.21.
4. Z. Amekraz and M. Y. Hadi, "CANFIS: A Chaos Adaptive Neural Fuzzy Inference System for Workload Prediction in the Cloud," in IEEE Access, vol. 10, pp. 49808-49828, 2022, doi: 10.1109/ACCESS.2022.3174061.
5. D. Alqahtani, H. Imani and T. El-Ghazawi, "Enhanced Workload Prediction in Data Centers Using Two-Stage Decomposition and Hybrid Parallel Deep Learning," in IEEE Access, vol. 13, pp. 64115-64132, 2025, doi: 10.1109/ACCESS.2025.3558743.
6. G. Vargas-Solar, M. S. Hassan and A. Akoglu, "JITA4DS: Disaggregated Execution of Data Science Pipelines Between the Edge and the Data Centre," in Journal of web Engineering, vol. 21, no. 1, pp. 1-26, January 2022, doi: 10.13052/jour1540-9589.2111.

7. L. Liu, X. Shen, Z. Chen, Q. Sun and R. wennersten, "Optimal Energy Management of Data Center Micro-Grid Considering Computing Workloads Shift," in IEEE Access, vol. 12, pp. 102061-102075, 2024, doi: 10.1109/ACCESS.2024.3432120.
8. R. Liu, W. Sun and W. Hu, "Workload Based Geo-Distributed Data Center Planning in Fast Developing Economies," in IEEE Access, vol. 8, pp. 224269-224282, 2020, doi: 10.1109/ACCESS.2020.3043949.
9. J. Li, C. Gu, L. Fang, X. wei, Y. Zhu and Y. Xiang, "Spatio-Temporal Reallocation Method for Energy Management in Edge Data Centers," in CSEE Journal of Power and Energy Systems, vol. 10, no. 5, pp. 1901-1910, September 2024, doi: 10.17775/CSEEJPES.2022.02000.
10. C. Guo, L. Li and M. Zukerman, "Q-Learning-Based Workload Consolidation for Data Centers With Composable Architecture," in IEEE Transactions on Industrial Informatics, vol. 21, no. 3, pp. 2324-2333, March 2025, doi: 10.1109/TII.2024.3503776.
11. A. Babakian, P. Monclus, R. Braun and J. Lipman, "A Retrospective on Workload Identifiers: From Data Center to Cloud-Native Networks," in IEEE Access, vol. 10, pp. 105518-105527, 2022, doi: 10.1109/ACCESS.2022.3211293.
12. S. Zhang et al., "MorphDAG: A Workload-Aware Elastic DAG-Based Blockchain," in IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 10, pp. 5249-5264, Oct. 2024, doi: 10.1109/TKDE.2024.3382743.
13. N. I. Mahbub, M. D. Hossain, S. Akhter, M. I. Hossain, K. Jeong and E. -N. Huh, "Robustness of Workload Forecasting Models in Cloud Data Centers: A White-Box Adversarial Attack Perspective," in IEEE Access, vol. 12, pp. 55248-55263, 2024, doi: 10.1109/ACCESS.2024.3385863.
14. B. Hu, K. Kempf and N. Mason, "A Workload Characterization Methodology Using Supervised and Unsupervised Deep Learning," in IEEE Access, vol. 12, pp. 181907-181913, 2024, doi: 10.1109/ACCESS.2024.3509857
15. A. Alaimo, A. Esposito, P. Faraci, C. Orlando and G. D. Valenti, "Human Heart-Related Indexes Behavior Study for Aircraft Pilots Allowable Workload Level Assessment," in IEEE Access, vol. 10, pp. 16088-16100, 2022, doi: 10.1109/ACCESS.2022.3145043.