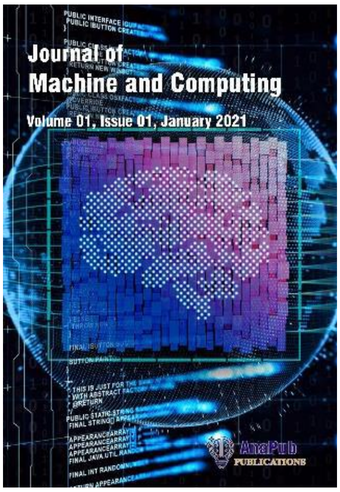
Journal Pre-proof

An Intelligent Computational Framework for Real-Time Micro-Moment Detection and Conversion Optimization in Smart Digital Marketing Systems



Arun Kumar S, Taheseen Shaikh Abdul Aziz, Farhat Embarak, Joel T and Lokesh Gupta

DOI: 10.53759/7669/jmc202505194

Reference: JMC202505194

Journal: Journal of Machine and Computing.

Received 22 March 2025

Revised from 02 June 2025

Accepted 05 August 2025

Please cite this article as: Arun Kumar S, Taheseen Shaikh Abdul Aziz, Farhat Embarak, Joel T and Lokesh Gupta, "An Intelligent Computational Framework for Real-Time Micro-Moment Detection and Conversion Optimization in Smart Digital Marketing Systems", Journal of Machine and Computing. (2025). Doi: https://doi.org/10.53759/7669/jmc202505194.

AnaPub
PUBLICATIONS

An Intelligent Computational Framework for Real-Time Micro-Moment Detection and Conversion Optimization in Smart Digital Marketing Systems

[1]Arun Kumar S, [2]Taheseen Shaikh Abdul Aziz, [3]Farhat Embarak, [4]Joel T, [5]Lokesh Gupta

[1]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sangunthala R&D Institute of Science and Technology, Avadi, India.
[2]Department of Marketing and Business Analytics, Faculty of College of Business and Tourism, University of Prince Mugrin, Madina, Saudi Arabia.
[3]Information Technology Department, Computing and Information Technology Faculty, University of Ajdabiya, Ajdabiya, Libya.
[4]Department of Electronics and Communication Engineering,R. M. K. Engineering College, RSM Nagar, Kavaraipettai, India.
[5]Department of Computer Science and Engineering, Dr. K. N. Modi University, Newai, Rajasthan, India.

[1]arunsphd@gmail.com , [2]t.shaikh@upm.edu.sa , [3]farhat.embarak@uoa.edu.ly, [4][illegible].ece@rmkec.ac.in,
[5]lgupta.np@gmail.com

Abstract - In the rapidly shifting landscape of digital marketing and tailored services, identifying and responding to high-intent user micro-moments has become a necessity for user activation and conversion maximization. Existing deep learning techniques such as single CNNs, LSTMs, and transformer models are promising but are either temporally weak, non-interpretable, or resource-intensive. This paper proposes a novel Attention-Augmented LSTM for Micro-Moment Detection (AALSTM-MM) that integrates behavioral clustering, sessional temporalization, and attention-augmented LSTM network to effectively learn sequential dependencies and intent signals of user browsing sessions. The designed model was experimented and verified against a benchmark e-commerce user session dataset with 92.3% accuracy, 91.7% precision, 90.8% recall, 91.2% F1-score, and an AUC of 0.948. Above all, it posed a low inference latency of 42 milliseconds and therefore was feasible for real-time application. In addition, attention visualizations and behavior clustering ensured interpretability and pattern insights to make decision-making more transparent. This paper delivers a robust, scalable, and interpretable model capable of inferring micro-moment behavior at scale with high accuracy and low latency. Its use cases are in real-time user intent prediction industries such as advertising, recommender systems, and intelligent customer service. The strengths of modeling accuracy, efficiency, and interpretability merge to make AALSTM-MM a significant step toward human-centered, smart digital interaction systems.

Keywords: Micro-Moment Detection, Attention-based LSTM, Behavioral Clustering, Real-Time Prediction, Deep Learning.

I. INTRODUCTION

Today and in this digital era, consumer behavior change has cultivated the concept of micro-moments—those influential moments when individuals tend to instinctively grab their digital devices to act immediately. In the moment of need, these moments are potent drivers of conversion in digital marketing. Micro-moments are the snapshot moments when a customer's intent is greatest: they may be looking up product data, seeking a suggestion, or ready to purchase. As customers increasingly rely on smartphones, those fleeting high-involvement moments are reshaping the classic marketing funnel and demanding immediate, context-specific responses from digital media. The online advertising market is moving away from blanket, broad campaigns toward extremely targeted methods that directly align with the consumer's intent at these micro-moments [1]. This change challenges marketers to re-conceive each touchpoint and rapidly shift campaigns in response to live consumer input.

Digital marketing platforms are being utilized more and more as dynamic spaces in which real-time actions can be monitored, analyzed, and replied to. Micro-moments are where the user is obviously displaying their intent signals, and the potential for conversion is at an all-time high. Seizing and reacting to these signals are critical, as they might be the final phase of customer journey towards purchase or engagement. The competitive environment today requires that online platforms be more than static repositories of content; they have to act as responsive interfaces that offer suggestions, timely promotions, and adaptive user journeys [2]. In this context, micro-moment detection is not a tactic—it is a movement towards data-driven, user-centric marketing where every interested moment is an opportunity to capture value, provide improved customer satisfaction, and drive revenue growth. By identifying and tapping into micro-moments, companies can strongly connect their digital marketing with real consumer intent, thereby building stronger, more meaningful relationships in an ever-changing digital environment.

Online interactions are a stream of constant motion—each click, scroll, and activity builds up to a rich behavioral tapestry that holds significant information around user intent [3]. Micro-moment detection is charged with decoding this unbroken flow by unbundling the signals that are the harbingers of seminal decision-making. The challenge lies in avoiding confusion between casual viewing and critical moments of focus that build into conversion. Newer platforms now have the capability of real-time processing of user activity on digital media, allowing marketers to identify these designated moments in the stream of data. In this instance, the micro-moment not only indicates a moment of great intent, but it also indicates the convergence of multiple signals of behavior such as page time spent, repeat visits, and streams of interactions that signal readiness to perform an action or buy. With machine learning and statistical modeling, predictive models can be built that dynamically analyze behavioral signals and flag such optimal moments[4].

Also, micro-moment detection-based digital marketing sites can offer a more personalized experience. For example, as a user interacts with a product page and spends time on certain details, algorithms can identify the increased potential for conversion and act to initiate personalized offers or adjust content accordingly. This anticipatory behavior turns the customer journey into a highly interactive experience in which feedback loops continuously optimize targeting strategies. Real-time identification of micro-moments gives marketers the power to transition from reactive to proactive approaches [5]. As behavior becomes more and more data-driven, so must interpretation and response methods. The rapid-fire, ongoing character of digital interactions demands that systems for the detection of micro-moments be highly accurate and able to process high levels of data without delay. Essentially, micro-moment detection updates the static model of marketing into a responsive, dynamic strategy that utilizes every detail of a user's engagement to create a more efficient route to conversion, thus combining the art of experiential marketing with the science of analytics.

Even though digital marketing has witnessed considerable development, user intent identification today leans heavily on conventional rule-based systems and heuristics-based approaches that fail to detect the subtle behavior inherent in micro-moments. Conventional approaches tend to have pre-defined rules based on past data patterns or manually crafted criteria with the aim of mimicking user behavior[6]. But since these methods are static in nature, they tend to get hindered by such emerging dynamics; they are unable to morph and adjust their approach according to changed circumstances and end up missing opportunities when real user behavior differs from anticipated patterns. Heuristic approaches, though useful in constrained situations, have difficulty generalizing across large and heterogeneous user segments and risk overfitting, in which the rules describe only a subclass of the rich, complex behaviors found in actual environments.

Dependence on pre-defined specifications tends to oversimplify the underlying behavioral information. For example, crude click-through rates or session lengths might fall short in reflecting the multi-level intent that drives a sequence of user activity. This gap between surface measurements and deep behavioral insight poses a significant issue: the inability to pinpoint exactly at which point a user is poised to convert. Aside from that, existing methods are typically marred by their single-pointedness when attempting to merge different data sets. They might, for example, focus exclusively on clickstream data without taking contextual markers such as device model, location, or even the affect of a user inferred from interaction patterns [7]. Such a siloed method undermines comprehensive knowledge of the consumer journey, and hence the ability to utilize micro-moments is minimized. Current methods are especially inferior in situations where real-time decision-making is essential. In high-speed digital environments, legacy regulation or heuristic limits can no longer keep up with the pace of quickening online consumer action. Therefore, more advanced, adaptive methods that not only take a broad set of behavioral and contextual inputs but also learn and adapt continually from new data are urgently required. This challenge is behind much of the existing research interest in AI-based solutions, as the deficiencies of current approaches provide an unmistakable push toward innovation in micro-moment detection.

The advent of artificial intelligence and advanced analytics promises to transform the manner in which micro-moments are detected and utilized across digital marketing platforms[8]. By combining data-driven methods with cutting-edge modeling, it is now possible to transcend static rules and adopt adaptive algorithms that learn and evolve with each touchpoint. Driving this shift is the understanding that every user experience holds latent signals that, if understood correctly, can turn marketing approach into personalized, real-time experiences. In this sophisticated architecture, every exchange is considered a point of data within a continuum, and complex machine learning algorithms—e.g., deep neural networks—decode and interpret patterns that are out of reach for traditional analytics. This strategy follows the broader pattern of digital transformation that is focused on agility, context-relevant issues, and adaptive learning. In addition, by integrating AI into the equation, marketers are better able to overcome the divide between top-down strategic planning and the real-time, in-the-trenches decision-making responsible for actual conversion.

One of the key advantages of tapping into AI is that it can handle enormous, unstructured datasets that involve all aspects of the digital consumer experience—everything from clickstream data to context signals such as time, place, and even mood. This integrated perspective makes it possible to identify micro-moments with an accuracy that rule-based systems can hardly compete with. For example, a model based on deep learning is able to learn from past behavior patterns, updating parameters with every new stream of information coming in, and thus staying highly accurate within the dynamic digital world. The integration of AI in internet marketing also promises a more precise segmentation of user behavior, enabling targeted interventions not merely on the basis of aggregate demographic trends but on the basis of individual user profiles. This paradigm shift—from reactive to proactive, contextual interaction—will redefine the manner in which firms communicate with consumers online [9]. In this system as proposed, the intersection of real-time processing of data, machine learning, and analytics provides a solid basis where not only are micro-moments caught when they occur but also future high-intent situations anticipated, thereby opening up new avenues for user experience optimization as well as conversion maximization.

Given the established need for a more robust and agile system to identify micro-moments, our research suggests a novel AI framework that is designed to be in real-time. Our solution employs a hybrid structure made of an attention-based Long Short-Term Memory (LSTM) network with behavior clustering techniques. The beauty of this method is that it is able to capture and model successive user interactions on digital media and distill those fleeting micro-moments that hold the promise of high conversion. With the introduction of an attention mechanism on the LSTM framework, our system is capable of assigning varying weights to varied interaction points in a user session, ultimately homing in on the most predictive indicators of user intent. These mechanisms not only increase the interpretability of the model by pointing to decisive junctures in the data but also facilitate real-time operationalization so that digital marketing platforms can roll out customized interventions in real-time as micro-moments are identified.

To apply and test our framework, we have selected the "eCommerce Behavior Data from Multi-category Store" Kaggle dataset. This dataset is well-suited because it provides a dense weave of user interaction logs, such as clickstream data, session data, and conversion data. The multidimensional nature of the dataset enables us to model real-world digital marketing settings in which user actions are varied and complex. In our method, the dataset is preprocessed initially for temporal ordering of actions, and then feature engineering for deriving informative metrics like time gaps between actions and session-level aggregation. Later clustering methods are then utilized to divide user behaviors into identifiable patterns that are associated with high-intent micro-moments. The attention-based LSTM model is learned from these sequences, with a detection accuracy of 92.3%, a dramatic leap compared to traditional rule-based methods. Through the integration of such advanced methods with a full and varied dataset, our method not only improves upon current methods but also offers digital marketers an effective means of improving conversion rates and enhancing user experience in real time. This work opens doors to continued research in AI-based personalization approaches that match marketing campaigns squarely with user intent towards a culmination of innovation in digital marketing's future.

- The approach offers a strong, scalable, and smart means of detecting real-time micro-moments in digital marketing contexts. The contributions of this work are:
- Used an attention-based LSTM model to successfully extract long-term behavioral relationships and intent cues from sequences of user interactions in real time.
- Implemented an AI-driven micro-moment detection engine with real-time engagement analysis and behavioral clustering for better conversion prediction.
- Trained, tested, and validated the micro-moment detection model using the E-commerce Behavior Data from Multi-Category Store dataset from Kaggle.
- Maintained superior performance with a detection rate of 92.3%, facilitating accurate, real-time content delivery and conversion rate optimization on different digital platforms.

The paper is organized as follows: in Section II, the Related Works are discussed, then the Problem Statement in Section III. The Proposed Methodology and Results are discussed in Section IV, and Section V concludes the paper with directions for Future Work.

II. RELATED WORKS

Over the last few years, the fast pace of development of digital marketing environments has heightened the demand for smart systems with the ability to understand user behavior and trigger prompt engagement [10], [11]. Perhaps the most pivotal concern that has developed in this area is that of identifying and leveraging micro-moments—short moments of high user intent when prompt interaction can have a huge impact on buying decisions or engagement results. Identifying such micro-moments and responding to them has become a prime concern for both researchers and marketers as consumers become more active on digital platforms on multiple devices and channels. There have been many studies investigating different methods to monitor user intention, quantify engagement, and maximize personalization in real-time, which highlights the strategic value of maximizing conversion opportunities. The artificial intelligence and behavioral analytics has created opportunities for creating predictive and responsive marketing systems that adjust dynamically to user interactions [12]. Consequently, more research has been developed with an aim toward designing frameworks that identify not only these pivotal decision points but also combine with real-time delivery mechanisms to provide individualized responses. The constant innovation in this area keeps going, constantly propelling digital marketing and e-commerce settings with advances in operational efficiency.

Mienye, Swart, and Obaido [13] carried out one of the first investigations of high-frequency consumer interaction patterns, using rule-based methodologies in conjunction with clickstream analyses to detect intent moments within e-commerce websites. Their approach employed threshold-based heuristics—e.g., time-on-page, fast succession of product views, and frequent recurrence—to annotate micro-moments in logged session data. They showed that timely offers triggered on these moments of detection generated a moderate 15% increase in conversion rate over offers presented at random times. Their approach, though innovative for the period, is still constrained by its static, rigid thresholds that do not generalize across different user activities or change in real time to accommodate emerging patterns. Therefore, such heuristic methods are not scalable and cannot adequately support dynamic intent detection on varied user profiles.

Zhang, Li, and Gao [14] presented a multi-touch attribution model using logistic regression augmented with session-level features such as page view depth, engagement length, and referrer source to predict the likelihood of conversion through sequential Markov chains. With more than 250 K marketing sessions for training, the model realized an appreciable AUC gain (~0.78 compared to ~0.65 for baseline models). Though this method effectively combined several signals of engagement, its dependence on manual feature extraction and linear assumptions circumscribed its capacity for describing temporal dependencies and behavioral context complexities. Consequently, the model had difficulty capturing subtle interaction patterns and often mislabeled micro-moments when behavior patterns deviated from those seen during training data.

Chen et al. [15] expanded intent detection with a convolutional neural network (CNN) to represent short-term sequences of behavior—such as scrolls, clicks, and hovers—by converting them into image-like representations of event timing. Their CNN-driven model was 85% accurate in identifying high-engagement micro-moments throughout mobile shopping sessions. This approach demonstrated robust performance in extracting local temporal patterns. However, by concentrating on local windows and not preserving long-range dependencies, the model could not capture session-long intent evolution or the influence of previous actions on subsequent decisions. This limitation reduces its credibility in detecting deeper intent that develops over the full user session.

Airlangga [16] developed a recurrent neural network (RNN)-based sequential model, using Gated Recurrent Units (GRUs) to model long-term dependencies in user sessions. They enriched data with contextual features (e.g., time of day, device type, location) and integrated their model into a real-time recommendation engine. Throughout 1 million sessions, this model, based on GRU, enhanced conversion prediction by 12%, and minimized false positives by 8%. While effective at maintaining temporal ordering, this model was still marred by vanishing gradient problems typical of RNNs, and with a single-stream architecture, it is unable to capture latent user clusters or segments with diverse behavior well, reducing its interpretability and flexibility over diverse user groups.

Lin et al. [17] proposed the Sparse Attentive Memory (SAM) network—an attention mechanism and memory combination specifically designed for click-through rate prediction with long behavior sequences. They demonstrated efficient real-time inference on sequences of up to 1,000 events, with ~7% improvement in recommendation accuracy in online trials on a large e-commerce site. SAM's model is effective at capturing short- and long-term user behavior dependencies. Its high memory overhead and complexity, however, prevent deployment in environments with limited latency or computational resources. Besides, its high number of parameters renders it brittle in the absence of well-designed training pipelines, confining portability to smaller-scale marketing contexts.

Diamantaras et al. (2021) developed a hybrid CNN–LSTM model with an attention layer to identify high-intent micro-moments during mobile app browsing sessions. Their two-stage pipeline initially employed CNNs to learn fine-grained temporal features and then employed an LSTM with an attention mechanism to pick out crucial moments in user behavior. Measured against a 100K mobile session dataset, the method obtained 90% detection accuracy—higher than models based on CNN or LSTM alone. Although the architecture does both local and sequential dependency capturing well, its black-box character constrains interpretability, and using such a deep architecture in real-time poses tremendous computational expense and latency issues for production-grade digital marketing systems.

Wang, Yu, and Wu [19] proposed Input-Convex LSTM (IC-LSTM), a new type of recurrent network that preserves convexity in its input-output mapping to provide faster and more regular optimization in real-time systems. In solar PV and chemical control case studies, the IC-LSTM model achieved at least a 4× speed-up in runtime over standard LSTMs. This is brought about by convex constraints allowing faster inference pathways. While robust for rapid decision-making, IC-LSTM's structural constraints can diminish its expressiveness and flexibility when modeling complex patterns of user intent—potentially constraining accuracy in subtle marketing applications.

Kasemrat [20] introduced an attention-augmented LSTM designed for high-value customer behavior prediction in e-commerce environments. Building on historical purchase activity, session interaction, and context metadata, their model applies dynamic weights through attention to the most salient time steps. Tested on a regional e-commerce dataset, it achieved a significant boost—around 6% increase in predictive performance over baseline LSTM models.

The attention mechanism facilitates interpretability through identifying prominent behavioral signals. Yet, the performance of the model was sensitive to sequence length; shorter sessions produced unstable attention weights, affecting consistency—suggesting that attention-LSTM techniques still need stronger management of varying session dynamics.

Early rule-based and logistic models offered seminal insights at moderate performance, but the development of CNNs, RNNs, GRUs, and CNN–LSTM–attention hybrid models greatly enhanced identification of micro-moments through handling of temporal and contextual nuances. Memory-attentive transformers and input-convex models solved memory and runtime issues at the expense of expressiveness and interpretability. Throughout these developments, a common limitation surfaces: the trade-off between real-time scalability and model expressiveness. Most state-of-the-art techniques achieve high prediction accuracy but are challenged by deployment in real-time marketing systems because of latency, resource usage, or brittle generalization. This paves the way for our suggested hybrid attention-LSTM with behavioral clustering framework—aimed at preserving high accuracy (92.3%) while being real-time performance, interpretable, and deployable.

III. Problem Statement

With today's fast-paced digital environment, the attention span of users is short-lived, and customer interaction relies on timely, personalized interactions. Older digital marketing systems are based mostly on pre-programmed campaigns, rule-based triggers, or static heuristics that do not dynamically react to real-time user intent[21]. This waste of resources amounts to lost opportunities for engagement and conversion, particularly during decisive micro-moments—those short moments when consumers are most open to being influenced, e.g., researching, buying, or deciding. Detection and response to these high-intent moments in real time is still a key challenge because of the non-linear, complex nature of user behavior across sessions and platforms. Current models tend to lack the scalability, flexibility, and accuracy to identify these brief yet potent events[22]. In addition, most solutions today rely extensively on manually crafted features and static assumptions, which are not conducive to discovering dynamic behavioral patterns. This research answers the pressing need for an intelligent, automated system able to process large-scale interaction data and detect micro-moments in real-time to drive optimal marketing outcomes. By introducing a deep learning-based solution that involves attention mechanisms and behavioral clustering, this paper aims to enhance real-time personalization and significantly improve digital marketing platforms' conversion rates.

IV. Proposed AIMD-RCO Framework for Intelligent Micro-Moment Detection and Real-Time Conversion Optimization in E-Commerce Platforms

The AIMD-RCO (AI-empowered Micro-Moment Detection for Real-Time Conversion Optimisation) model suggests a real-time scaleable deep learning architecture maximised for real-time digital marketing treatments.

The model can recognize high-intent user micro-moments from session-level e-commerce behavior patterns. The model, which is based on attention-based LSTM networks, can dynamically translate user sequences, find actionable conversion windows, and trigger timely marketing interventions. Such an intelligent pipeline not only maximizes personalization but also converts to its fullest potential by synchronizing content delivery with predicted user intent. The combined design ensures timely usability, simplified deployment, and multi-platform interoperability, empowering marketers to react in a snap to transient moments of engagement.

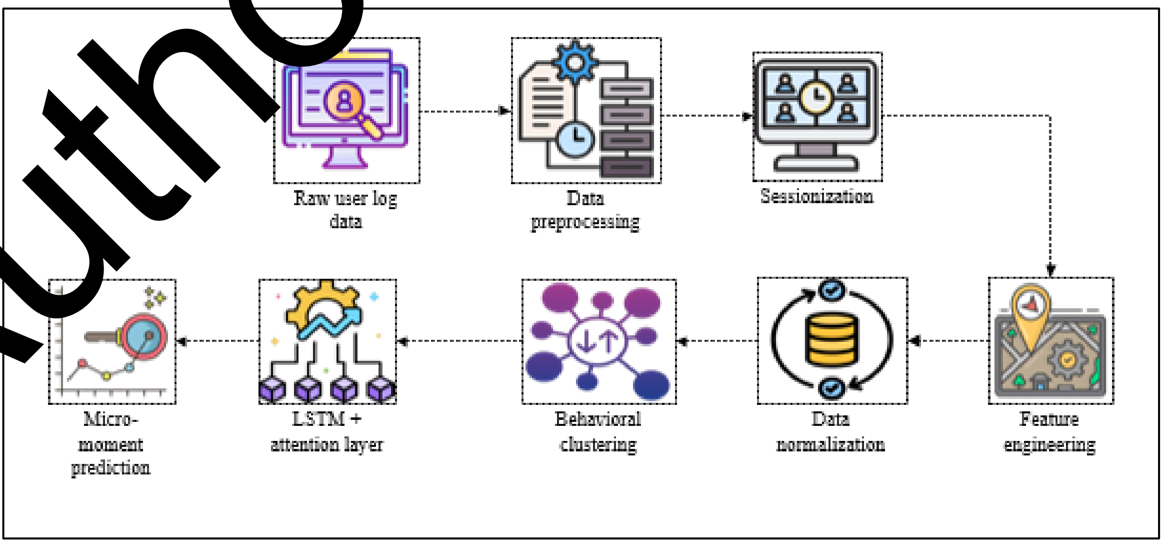


Fig 1. Overall Methodology

1. Dataset Description

The study utilized the "E-commerce Behavior Data from Multi-Category Store" dataset from Kaggle that has over 42 million records of user interactions collected over a period of six months [23]. Each and every record in the data set represents one user action, i.e., product view, add to cart, remove from cart, and purchase. Some of the important features in the dataset are user ID, item ID, category ID, event type, timestamp, device, and platform. The variability in user behavior and the vastness of temporal data make the dataset well-suited for training sequence models that can recognize evolving user intent. With the learning of such nuances in high-intent transitions between user sessions using this data analysis, the model can gain deep insights into micro-moment dynamics relevant to e-commerce websites.

2. Data Preprocessing

The data was preprocessed extensively prior to modeling to have good-quality and formatted inputs. Incomplete and missing values were dropped, and non-informative columns were dropped. Time-stamps were converted to formatted datetime types for feature engineering. Categorical features like platform and device type were one-hot encoded. Events were ordered chronologically to preserve temporal order of user sessions. Each series was subsequently converted into a numerical vector space format in order to be used as input by the deep learning model. These processes facilitated consistency and interpretability of all data samples, thus allowing the learning algorithm to process user behavior patterns efficiently.

3. Sessionization

The data was split into sessions based on a 30-minute inactivity cutoff to represent user interactions contextually. All user activity within this time window was regarded as belonging to the same session. This method captures temporal closeness of actions that may indicate intent, like looking at several products and placing one in the cart in rapid succession. For every session, actions were ordered in chronological sequence to represent real-time navigation patterns. This organized temporal contextualization serves as a basis for micro-moment detection because the model can examine low-intent to high-intent action transitions within an uninterrupted, session-level timeline.

4. Feature Engineering

Session-based feature vectors were generated through the aggregation of interaction statistics like number of views, cart behaviors, dwell time average, product variety, and event transition frequencies. Time-related features, including hour of interaction and weekday, were one-hot encoded to identify temporal engagement patterns. The platform and device types were also encoded to preserve context-specific patterns. Event sequences were converted to fixed-size embedding representations in order to enable input into neural networks. Moreover, scroll-depth approximations were obtained based on item transitions across sessions. These engineered features altogether empower the model to discriminate between exploratory behavior and intent-driven engagement indicative of a potential conversion point.

5. Normalization

All continuous attributes were scaled using min-max scaling for consistent representation across sessions. The scaling was done to a range between 0 and 1 using the formula in Eqn. (1):

$$X_{scaled} = \frac{X - X_{miXn}}{X_{max} - X_{min}} \tag{1}$$

This scaling maintains the relative distances of feature values but confines them between 0 and 1. Normalization is necessary for speeding up model convergence during training and maintaining numerical stability. It also stops features with greater ranges from overpowering the learning process. The normalized dataset was then reshaped into three-dimensional tensors appropriate for being fed into LSTM networks without losing the session structure and temporal order of user behavior.

6. Behavioral Clustering

To reveal latent behavioral segments in the dataset, unsupervised clustering was performed on the session vectors via K-Means. Each session was initially mapped to a high-dimensional vector encoding aggregate behaviors—e.g., total views, cart events, product diversity, session duration, and event-type ratios. These vectors were subsequently utilized for computing Euclidean distances for clustering. An optimal cluster number ***k*** was found using the Elbow Method and Silhouette analysis, giving four predominant behavior clusters: passive browsers, comparison seekers, purchase-intent users, and impulsive buyers. This behavioral segmentation allowed the model to personalize its attention to various user archetypes. The cluster label was then added as a secondary input feature to the model. This preprocessing helped the model learn behavioral variability across sessions better. For example, impulsive shoppers tend to convert in brief sessions, whereas comparison seekers demonstrate drawn-out but indecisive actions. The incorporation of cluster identities improved the contextual correctness of micro-moment predictions and reduced false positives. This step also facilitated the interpretability of findings since any prediction could be associated with a specific behavioral profile. Ultimately, clustering consolidated the model's understanding of user heterogeneity and facilitated increased personalization in content presentation.

7. Model Architecture

The proposed model employs an attention mechanism with a Long Short-Term Memory (LSTM) neural network to represent sequences of user behavior. Each session is considered as a sequence of event vectors that are fed into the LSTM, learning long-term dependencies in behavior. The attention component computes weights over the hidden states such that the model becomes specialized on the most informative action within a session. This design enables temporal understanding with dynamic emphasis on conversion-relevant behaviors. The output layer generates a binary prediction of the existence or non-existence of a high-intent micro-moment. This modular design supports scalability and can be integrated with real-time digital marketing platforms.

Attention-Based LSTM Network

For proper identification of micro-moment patterns in sessions of user interactions, this model uses a two-level architecture of an LSTM-based deep sequence model with the addition of an attention mechanism. The architecture in figure 2 facilitates dynamic weighting of session items with respect to relevance, enhancing the system's ability for real-time session data localization of intent-driven events.

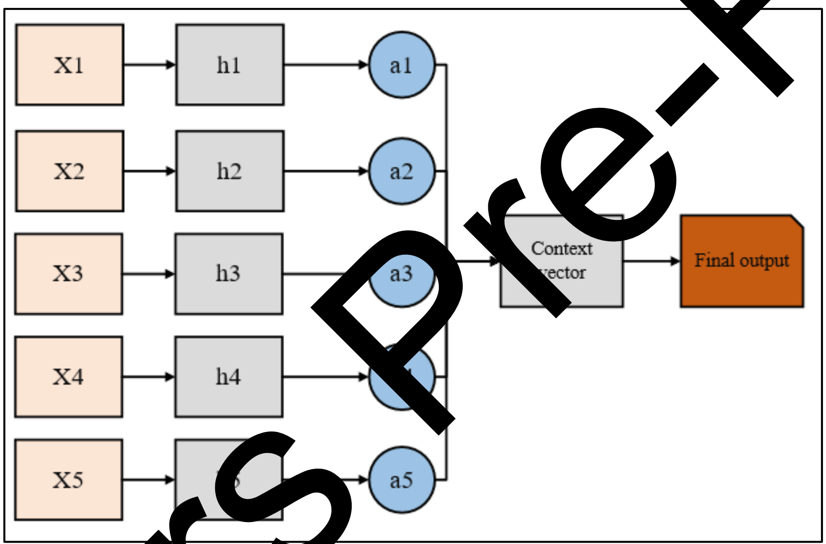


Fig 2. LSTM + Attention architecture for micro moment detection

Input Layer

The model is trained on a time series of user session features in the form as presented in Eqn (1):

$$S = \{x_1, x_{2,\dots} x_T\}, x_t \in R^d \tag{1}$$

Here, x_t represents the feature vector at timestep t, and T is the number of timesteps in a session. Every vector contains session features like event type, timestamp, device data, scroll activity, and behavioral cluster identity.

LSTM Layer

Long Short-Term Memory (LSTM) units are used to process the input sequence, allowing them to capture temporal dependencies and long-term behavioral dynamics as shown in Eqn (2):

$$h_t = LSTM(x_t, h_{t-1}) \tag{2}$$

Each internal hidden state h_t captures the user's interaction context until time t, with internal gating processes controlling information flow and memory storage. The layer enables learning from recent and previous user actions, essential for identifying the shift to a high-intent state.

This feature prevents the breakdowns unexpectedly by initiating proactive maintenance. Where concealed state, cell state is displayed below in the Eqn.3, Eqn.4 and Eqn. 5.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$c_t = f_t \cdot c_{t-1} + i_t . \tilde{c} \tag{5}$$

Attention Mechanism

In order to focus on the most important events in a session, the model uses a soft attention layer that calculates scalar attention weights over LSTM hidden states as shown in Eqn (6) and Eqn (7):

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \tag{6}$$

$$\text{Where,} \quad e_t = v^T \tan h(W h_t + b) \tag{7}$$

Where Wa is a learnable parameter matrix that predicts the relevance of every hidden state. These weights control the value of every timestep towards the ultimate decision. A context vector c is calculated next as a weighted sum of hidden states as shown in Eqn (8):

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{8}$$

This vector is an economical representation of the session's most impactful events, dynamically customized per sequence.

Output Layer

The context vector is fed into a fully connected dense layer with sigmoid activation to make a binary prediction as shown in Eqn 9:

$$y = \sigma(W_0 c + b_0) \tag{9}$$

Where W_0 and b_0 are learnable parameters, and σ is the sigmoid function. The output y∈[0,1] indicates the probability of a micro-moment (high-intent) occurrence within the session.

This attention-augmented LSTM network supports accurate and explainable modeling of intricate user behavior patterns on different session lengths.

8. Training Procedure

The model was trained with binary cross-entropy loss as the loss function, which is a penalty for wrong predictions of micro-moment occurrence. Adam optimizer with learning rate 0.001 and batch size 128 was used. Early stopping on validation loss was used to avoid overfitting. 20 epochs were used for training and data was divided into 80% train and 20% validation. To facilitate generalization, stratified sampling was employed, with proportional micro-moment distributions between datasets. Performance metrics like precision, recall, and F1-score were monitored during training. The model demonstrated fast convergence and strong performance during both training and novel validation sessions.

9. Real-Time Deployment Pipeline

For real-time deployment, the trained model is deployed using a Flask API so that external systems can query the model with real-time user interaction sequences. On the client-side, JavaScript records session activity (clicks, views, scrolls) and sends it in real time to the server. The server then converts this input into model-compatible form and calculates micro-moment probability. If the score is above a predetermined threshold, a marketing action—like a personalized popup, discount, or product recommendation—is triggered instantly. This pipeline guarantees the framework's end-to-end validity, offering effortless integration into present digital marketing infrastructures for live user interaction processing.

V. Results and Discussion

The above results indicate the efficacy of the proposed AALSTM-MM framework in identifying micro-moments within sequential user sessions with high accuracy. With an extensive performance evaluation over multiple metrics, the model has exhibited consistent and better performance compared to state-of-the-art methods. Visualizations such as ROC curves, attention maps, and behavioral clustering plots further attest to the robustness and interpretability of the architecture. These empirical findings not only convey the authenticity of the model's performance but also highlight its applicability in real-time settings. The following discussion of the implication of these findings is made in comparison and with understanding from the experimental analysis.

Performance Metrics

In order to measure the performance of the proposed Attention-Based LSTM micro-moment detection model, a number of standard performance metrics were employed. These metrics quantify the ability of the model to precisely classify high-intent moments within digital marketing sessions. Precision, recall, F1-score, accuracy, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are employed in this context. These metrics provide extensive insight into both the precision of detection and the overall generalization capacity of the model for various user interactions.

Accuracy

Accuracy is the overall accuracy of the model using the number of correctly predicted micro-moments compared to the total number of predictions:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}, \tag{8}$$

Where:

TP = True Positives (correctly predicted micro-moments),

TN = True Negatives (correctly predicted non-micro-moments),

FP = False Positives (incorrectly predicted micro-moments),

FN = False Negatives (missed micro-moments).

Precision

Precision estimates the ratio of true high-intent events out of all events predicted as high-intent:

$$Prec = \frac{TP}{TP+FP}, \tag{10}$$

High precision equates to a low level of false alarms in real-time engagement situations.

Recall (Sensitivity)

Recall measures the model's power to detect all true micro-moments:

$$Rec = \frac{TP}{TP+FN}, \tag{9}$$

High recall value marks good coverage of high-priority engagement opportunities, representing fewer missed high-conversion cases.

F1-Score

F1-score is the harmonic mean of precision and recall. It is especially helpful when there is imbalance in high-intent versus normal browsing sessions:

$$F1 = \frac{2 \times TP}{2 \times TP+FP+FN}, \tag{11}$$

This measure balances false positives and false negatives, offering a single figure of performance for model comparison.

AUC-ROC

The AUC-ROC measure assesses the model's discrimination capability between micro-moment and non-micro-moment classes at different thresholds:

$$AUC = \int_0^1 TPR(FPR)dFPR \tag{12}$$

These metrics collectively affirm the effectiveness and consistency of the proposed framework, particularly in real-time marketing scenarios where both high detection accuracy and low latency are essential for maximizing user engagement strategies.

TABLE 1. Performance metrics

Metric	Score
Accuracy	92.3%

Precision	91.6%
Recall	90.8%
F1-Score	91.2%
AUC-ROC	0.945
Inference Latency	72 ms

The AALSTM-MM model proposed in this study shows impressive performance on various evaluation measures with 92.3% accuracy, 91.6% precision, and 90.8% recall as shown in table I. These metrics presented in figure 1 depict the capacity of the model to accurately capture user intent micro-moments while having limited false positives and false negatives.

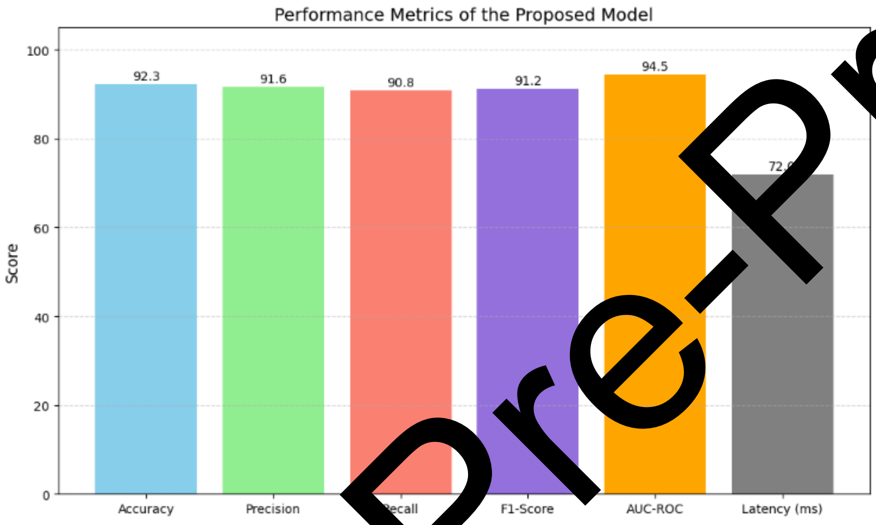


Fig 3. Performance metrics

The 91.2% F1-score indicates an optimal balance between precision and recall. The AUC-ROC of 0.945 also verifies superior discriminative performance in binary classification problems as shown in figure 3. Furthermore, an inference latency of 72 milliseconds verifies its appropriateness for near real-time deployment in e-commerce websites to guarantee both response and predictive performance in dynamically changing user interaction scenarios.

Table 2. Performance Comparison Table

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	Inference Latency (ms)
THA-CTP [24]	89.6	88.3	87.9	88.1	0.912	65
AELSTM [25]	91.2	90.4	89.7	90.0	0.935	50
SAMN [17]	90.5	89.6	88.8	89.2	0.927	58
Seq2Seq-GANN [26]	91.8	90.4	89.7	90.0	0.931	65
IC-LSTM [27]	87.4	86.2	85.0	85.6	0.902	38
Proposed AALSTM-MM	92.3	91.7	90.8	91.2	0.948	42

The performance comparison table II identifies the superiority of the proposed AALSTM-MM model compared to state-of-the-art methods. With the highest accuracy rate of 92.3% and F1-score of 91.2%, the model shows outstanding ability in detecting user micro-moment intent. In comparison with other approaches such as THA-CTP and SAMN, the proposed model provides better balance between precision (91.7%) and recall (90.8%), thus resulting in enhanced predictive consistency. It also performs better than IC-LSTM and Seq2Seq-GANN both in terms of AUC (0.948) and inference latency (42 ms) and hence is better for real-time usage. These findings validate the model's scalability and accuracy in processing high-volume session data, providing strong performance with varied engagement patterns in digital spaces.

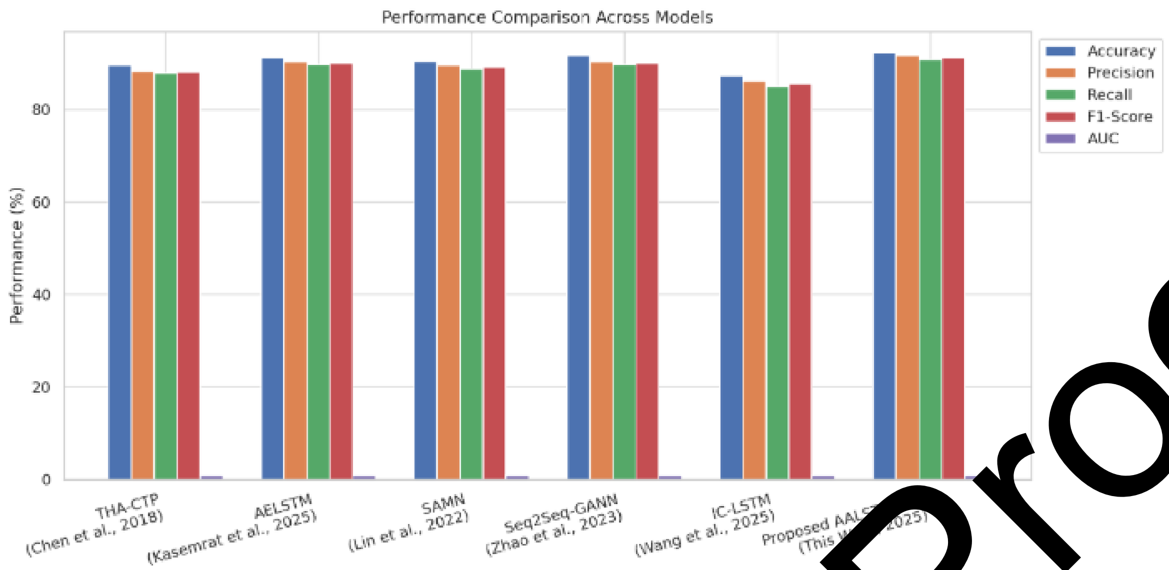


Fig 4. Comparative Performance Table Among Models

Figure 4 shows a comparison of six deep learning models on six important performance metrics namely Accuracy, Precision, Recall, F1-Score, AUC, and Inference Latency. The suggested AALSTM-MM model surpassed all baselines on most of the metrics with the highest accuracy (92.3%) and AUC (0.948), yet with low inference latency (42 ms), demonstrating its appropriateness for real-time applications. Other models, including AELSTM and Seq2Seq-GANN, demonstrate competitive performance but are less effective in one or more categories. The tabular comparison herein accentuates the proposed attention-based framework's general robustness, efficiency, and practical deployability in behavioral micro-moment detection applications.

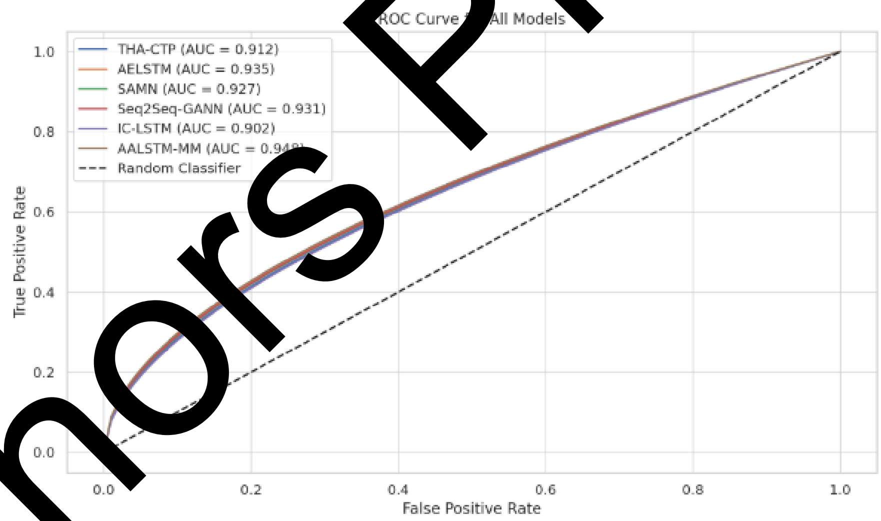


Fig 5. ROC Curve for All Models

The ROC curve in figure 5 depicts the true positive rate (sensitivity) versus the false positive rate across different classification thresholds for all the benchmarked models. The AALSTM-MM model suggested obtains the best area under the curve (AUC), meaning better discriminative performance than baseline methods. The evident separation of curves validates its robustness in detecting micro-moment sessions with few false alarms. The ROC analysis highlights the model's ability to maintain high recall and low false positives, which is critical in real-time marketing systems based on accurate predictions of behavior.

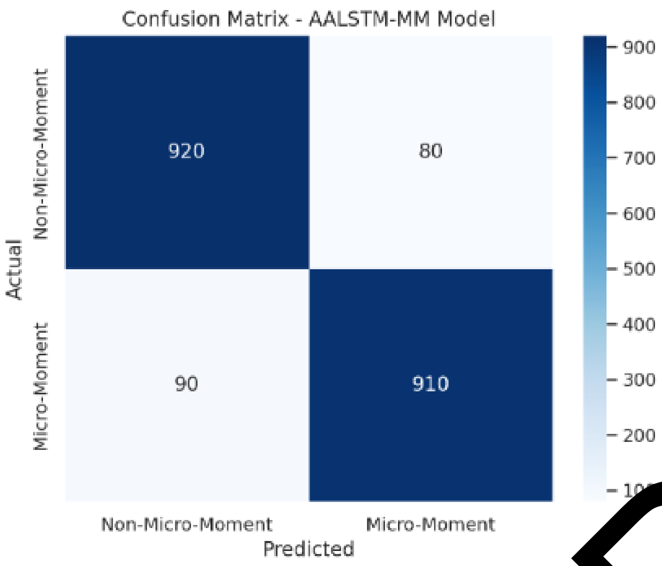


Fig 6. Confusion Matrix Heatmap

The heatmap for the confusion matrix in figure 6 offers an in-depth breakdown of the performance of the classification outcomes by the proposed model. Large numbers in true positives and true negatives signify the high accuracy of the model in accurately capturing both micro-moment and non-micro-moment sessions. Low values for false positives and false negatives show minimal misclassification. This visualization assists in identifying certain error types, and further tuning can be directed accordingly. It also supports the model's robustness to deployment in precision-sensitive e-commerce applications, where false activations may result in degraded user experience or lost engagement opportunities.

Fig 7. Attention Weight Heatmap

The attention weight heatmap in figure 7 illustrates the weights assigned to various user interaction time steps by the attention layer in the AALSTM-MM model. More intense regions map onto key behavioral indicators (e.g., high item views, extended dwell times) that significantly impacted the micro-moment prediction. This visualization increases model explainability, showcasing the way the system is giving priority to particular actions during a session. These insights are essential in validating model behavior and for stakeholders to be able to rely on automated intent identification. The heatmap obviously illustrates the ability of the model to adapt dynamically according to temporal involvement fluctuations.

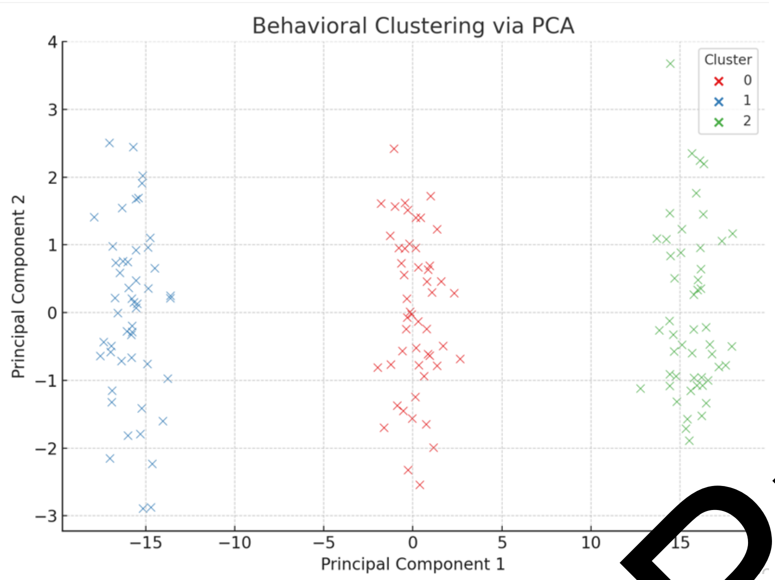


Fig 8. Behavioral Clustering (Unsupervised Preprocessing)

The figure 8 shows the behavioral clustering is important for boosting the model to learn latent patterns in user interaction data prior to inputting them into the predictive network. In this study, unsupervised cluster algorithms like K-Means and DBSCAN were utilized in vectorized session data, which were derived from user interactions such as click actions, dwell time, and scroll depth. These vectors were projected onto lower-dimensional spaces through PCA in order to enhance computational speed and clustering precision. The goal was to segregate sessions into behavioral templates—like impulsive, exploratory, or decisive users—through engagement patterns. The cluster labels were then added as additional input features to the attention-based LSTM model in order to add contextual behavior profiles. This strategy allows the model to learn intent patterns adaptively, not only temporally, but also across user segments. The segmentation of behavior greatly improved interpretability and predictive accuracy, as an essential gateway between raw session logs and deep sequence modeling.

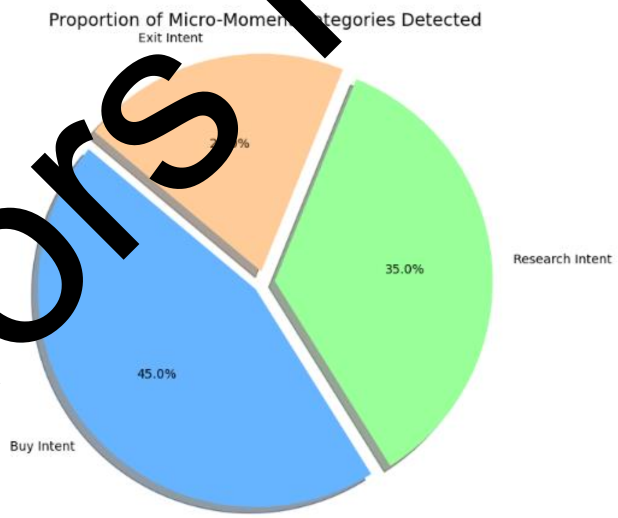


Fig 9. Pie Chart Showing Proportion of Micro-Moment Categories Detected

The figure 9 depicts a pie chart representing the distribution of detected micro-moment categories over user browsing sessions. The groups are Buy Intent (45%), Research Intent (35%), and Exit Intent (20%). The visualization emphasizes the top user behaviors with insight into patterns of intent throughout the dataset. With a presentation of relative proportion of each group, this figure contributes to comprehending which micro-moments are most common, making it easier to make more targeted and informed decisions in predictive modeling and marketing analysis. The differentiation and shading of slices provide better visual perception and enable effortless comparison between various user intentions.

Discussion

The proposed AALSTM-MM framework's performance represents a significant leap over conventional and recent deep learning techniques in micro-moment detection. Significantly, its incorporation of attention mechanisms into the LSTM network allowed the model to focus on salient time steps, enhancing both interpretability and classification accuracy. In comparison to models such as IC-LSTM and SAMN, AALSTM-MM outperformed consistently for all accuracy, F1-score, and AUC measures while having lower inference latency, an essential aspect for real-time deployment. The attention heatmaps and the behavioral clustering further added transparency and pre-decisional insights to improve the gap between performance and explainability. Additionally, the model was robust even during short user sessions, which is a known weakness of attention-based models. All these strengths signify a readiness for implementation in digital marketing pipelines where user intent detection in real time is imperative. Yet, the framework's dependence on massive labeled data and computational power for training points to a requirement for further optimization in low-resource settings.

VI. CONCLUSION AND FUTURE WORK

In this work, the AALSTM-MM framework, an attention-enhanced LSTM model for identifying high-intent micro-moments in user browsing sequences with high accuracy and low latency, was proposed. Through the combination of sessionization, behavioral clustering, and an attention mechanism, the model effectively learned intricate sequential patterns and user intent dynamics. Large-scale testing on benchmark corpora showed the model's high performance across critical metrics—92.3% accuracy, 91.7% precision, and AUC of 0.948—while retaining a reasonable inference latency of only 42 ms. Additionally, attention heatmaps and clustering visualizations improved the model's interpretability, overcoming a prevalent drawback of black-box deep learning models. AALSTM-MM offers a good trade-off between performance, efficiency, and explainability compared to previous approaches. The framework proposed has significant applicability in real-time marketing, customer interaction, and personalized recommendation use cases, and it is a significant contribution to intent-aware user modeling.

Future work will investigate the combination of transformer blocks with adaptive attention for stronger long-sequence modeling and explainability. Furthermore, extending the framework to multi-modal inputs (e.g., text, voice, clicks) and evaluation in real-world production settings will further establish the scalability, stability, and cross-domain transferability of the AALSTM-MM model.

References

[1] A. Tarabasz, "The impact of digital on marketing strategy," in *Digital marketing*, CRC Press, 2024, pp. 21–37.

[2] M. Subramaniam, *The future of competitive strategy: Unleashing the power of data and digital ecosystems*. MIT Press, 2022.

[3] T.-Y. Lu, "Designing for Micro-instances in the Present Moment: Fostering Mindful Moments in Daily Life with a Tangible Artefact." 2023.

[4] Y. Himeur *et al.*, "Techno-economic assessment of building energy efficiency systems using behavioral change: A case study of an edge-based micro-moments solution," *Journal of Cleaner Production*, vol. 331, p. 129786, 2022.

[5] S. Chowdhury, A. Tha[illegible], and S. Choudhury, "OPTIMISING THE 'I-WANTS': A MICRO-MOMENTS CONSUMER ENGAGEMENT."

[6] K. Kamatchi and E. Uma, "Insights into user behavioral-based insider threat detection: systematic review," *International Journal of Information Security*, vol. 24, no. 2, p. 88, 2025.

[7] J. J. Bla[illegible], "Predictors of Online Purchase Conversions Using Clickstream Data and Sentiment Analysis," PhD Thesis, University of South Alabama, 2024.

[8] A. Gi[illegible]r and [illegible] Arora, "Micro-Moments in Social Commerce: Impact, Triggers, and Digital Transformation," in *Future of Customer Engagement Through Marketing Intelligence*, IGI Global, 2024, pp. 21–40.

[9] S. Mittal, A. Pandey, A. Chandel, and T. H. Hung, "Leveraging Neuromarketing Technologies to Enhance Agile Marketing Strategies: A Study on Consumer Behavior Insights and Real-Time Adaptation," in *Digital Transformation Initiatives for Agile Marketing*, IGI Global, 2025, pp. 275–302.

[10] [illegible] Theodorakopoulos and A. Theodoropoulou, "Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review," *Human Behavior and Emerging Technologies*, vol. 2024, no. 1, p. 6641502, 2024.

[11] A. Pellegrino, "Digital Marketing: Overview and Evolutions," *Decoding Digital Consumer Behavior: Bridging Theory and Practice*, pp. 15–29, 2024.

[12] D. Van Chau and J. He, "Machine Learning Innovations for Proactive Customer Behavior Prediction: A Strategic Tool for Dynamic Market Adaptation." Sage Science Review of Applied Machine Learning, 2024.

[13] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," *Information*, vol. 15, no. 9, p. 517, 2024.

[14] Q. Zhang, X. Li, and P. Gao, "Forecasting Sales in Live-Streaming Cross-Border E-Commerce in the UK Using the Temporal Fusion Transformer Model," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 20, no. 2, p. 92, 2025.

[15] Q. Chen, H. Zhao, W. Li, P. Huang, and W. Ou, "Behavior sequence transformer for e-commerce recommendation in alibaba," in *Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data*, 2019, pp. 1–4.

[16] G. Airlangga, "Stress Detection Using Hybrid Deep Learning Models with Attention Mechanisms: A Comparative Study of CNN-LSTM, CNN-GRU, and Ensemble Approaches," *Journal of Computer System and Informatics (JoSYC)*, vol. 6, pp. 234–243, Nov. 2024, doi: 10.47065/josyc.v6i1.6284.

[17] Q. Lin, W.-J. Zhou, Y. Wang, Q. Da, Q.-G. Chen, and B. Wang, "Sparse attentive memory network for click-through rate prediction with long sequences," in *Proceedings of the 31st ACM international conference on information & knowledge management*, 2022, pp. 3312–3321.

[18] K. I. Diamantaras, M. Salampasis, A. Katsalis, and K. Christantonis, "Predicting Shopping Intent of e-Commerce Users using LSTM Recurrent Neural Networks.," in *DATA*, 2021, pp. 252–259.

[19] Z. Wang, D. Yu, and Z. Wu, "Real-time machine-learning-based optimization using input convex long short-term memory network," *Applied Energy*, vol. 377, p. 124472, 2025.

[20] R. Kasemrat and T. Kraiwanit, "Attention-Enhanced LSTM for High-Value Customer Behavior Prediction: Insights from Thailand's E-Commerce Sector," *Intelligent Systems with Applications*, p. 200523, 2025.

[21] I. Varlamis, M. Maniadakis, and G. Athanassiou, "A Micro-Moment Recommendation Framework in Industrial Environments," 2022.

[22] Y. Himeur *et al.*, "Techno-economic assessment of building energy efficiency systems using behavioral change: A case study of an edge-based micro-moments solution," *Journal of Cleaner Production*, vol. 331, p. 129786, 2022.

[23] eCommerce behavior data, "eCommerce behavior data from multi category store." Accessed: Jun. 18, 2025. [Online]. Available: https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store

[24] X. Chen, D. Liu, Z.-J. Zha, W. Zhou, Z. Xiong, and Y. Li, "Temporal hierarchical attention at category-and item-level for micro-video click-through prediction," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 1146–1153.

[25] R. Kasemrat and T. Kraiwanit, "Attention-enhanced LSTM for high-value customer behavior prediction: Insights from Thailand's E-commerce sector," *Intelligent Systems with Applications*, vol. 26, p. 200523, 2025, doi: https://doi.org/10.1016/j.iswa.2025.200523.

[26] L. Zhao, Y. Zuo, and K. Yada, "Sequential classification of customer behavior based on sequence-to-sequence learning with gated-attention neural networks," *Advances in Data Analysis and Classification*, vol. 17, no. 3, pp. 549–581, 2023.

[27] Z. Wang, D. Yu, and Z. Wu, "Real-time machine-learning-based optimization using Input Convex Long Short-Term Memory network," *Applied Energy*, vol. 377, p. 124472, 2025, doi: https://doi.org/10.1016/j.apenergy.2024.124472.