Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 01, 2025

Assessment of Bot Detection Approaches Using Behavioral Biometrics and Mouse Dynamics

[1]Prof. G. G. Sayyad, [2]Sahil Kadam, [3]Kishor Kadam, [4]Saneel Godage, [5]Ritesh Zagade

[1 2 3 4 5]Department of Computer Engineering, SPVP's S.B. Patil College of Engineering, Indapur, Pune, India

Email: sahilgorakshanathkadam3@gmail.com, kishorkadam1817@gmail.com, saneelgodage@gmail.com, zagaderitesh2312@gmail.com

Peer Review Information

Submission: 11 Sept 2025

Revision: 10 Oct 2025

Acceptance: 22 Oct 2025

Keywords

Bot Detection, Behavioral Biometrics, Mouse Dynamics, Machine Learning, Deep Learning, Cybersecurity, User Authentication.

Abstract

The proliferation of automated software agents, or ”bots,” presents a significant and evolving threat to web security, data integrity, and user trust. Traditional defense mechanisms, most notably CAPTCHAS, have been systematically defeated by advancements in artificial intelligence, rendering them increasingly ineffective and detrimental to user experience. In response, the field of cybersecurity has shifted its focus towards behavioral biometrics, a paradigm that seeks to distinguish humans from bots based on their intrinsic interaction patterns. This survey provides a comprehensive review of the literature on bot detection with a specific focus on mouse dynamics—the analysis of a user's cursor movement patterns. We trace the evolution of this field from foundational concepts and statistical feature engineering to the adoption of sophisticated deep learning models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs). Furthermore, we examine the critical role of public datasets in advancing research, explore the challenges posed by advanced threats such as session-replay bots and adversarial attacks, and identify key research gaps. This review synthesizes the current state-of-the-art and establishes a clear justification for the development of next-generation, robust, and frictionless bot detection systems.

INTRODUCTION

The digital landscape is engaged in a continuous and escalating arms race between malicious automated agents (bots) and the systems designed to protect web applications. Bots are responsible for a wide range of nefarious activities, including data theft, spam distribution, Distributed Denial-of-Service (DDoS) attacks, and click fraud [1]. For years, the primary line of defense has been the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). However, the foundational premise of CAPTCHA—a task that is easy for humans but difficult for computers—has been fundamentally inverted. Recent studies show that modern AI, leveraging deep learning and advanced Optical Character Recognition (OCR), can solve traditional text and image-based CAPTCHAs with accuracy rates as high as 96%, significantly outperforming the average human success rate of 50-86% [2]. This paradigm shift signals the obsolescence of challenge-response security mechanisms.

In response to the failure of traditional methods, research has pivoted towards more passive and intelligent forms of user verification. Behavioral biometrics has emerged as a leading alternative, offering a way to authenticate users based on their unique, subconscious patterns of interaction with a system [3, 4]. This modality is advantageous because it is unobtrusive, requires no specialized hardware, and analyzes behaviors that are inherently difficult for automated scripts to perfectly replicate [5]. Among the various forms of behavioral biometrics, mouse dynamics has garnered significant attention. It involves capturing and analyzing the rich, high-fidelity data stream generated by a user's cursor movements, including velocity, acceleration, path curvature, and pauses.

This survey paper presents a structured and comprehensive review of the academic literature concerning the use of mouse dynamics for bot detection. The objective is to map the intellectual terrain of this field, from its foundational principles to the current state-of-the-art. We will cover:

- The evolution from classical machine learning techniques reliant on handcrafted features to modern deep learning approaches that enable end-to-end feature extraction.
- The role and limitations of publicly available datasets used for training and benchmarking detection models.
- The emerging challenges posed by sophisticated threats, including session-replay bots and the vulnerability of machine learning models to adversarial attacks.

By synthesizing this body of work, we aim to identify the critical research gaps and unresolved questions, thereby establishing a clear and compelling rationale for the development of a novel, robust, and frictionless bot detection system based on mouse dynamics.

FOUNDATIONAL CONCEPTS AND METHODOLOGIES

Bot detection is a broad field within cybersecurity, encompassing a variety of techniques to identify and mitigate automated threats. These methods can be broadly categorized, providing a context for understanding where mouse dynamics fits within the larger landscape.

A Taxonomy of Bot detection Techniques

Historically, bot detection methods have been classified into several primary categories [1]:

- Signature-Based Detection: This is the most traditional approach, analogous to antivirus software. It relies on identifying bots by matching their network traffic or behavior against a database of known malicious signatures or patterns. While effective against known threats, it is inherently reactive and fails to detect new or polymorphic bots for which no signature exists [6].
- Anomaly-Based Detection: This method establishes a baseline of ”normal” network or user behavior and flags any significant deviations as potentially malicious. It is more effective at detecting novel threats but can suffer from a high rate of false positives if the baseline is not well-defined or if legitimate behavior is highly variable [7].
- DNS-Based and Mining-Based Approaches: These techniques focus on analyzing network-level data, such as DNS queries or traffic flows, to identify patterns indicative of botnet command-and-control (C&C) communication or other coordinated malicious activities [1].

While these network-centric approaches are valuable, they often struggle to detect sophisticated bots that are designed to mimic legitimate traffic patterns. This limitation has driven the adoption of host-level analysis, focusing on the interaction between the user (or bot) and the client machine.

The Rise of Behavioral Biometrics

Behavioral biometrics represents a paradigm shift from what a user knows (password) or what a user has (token) to who a user is, based on their actions. It involves measuring unique, idiosyncratic patterns in human activities for identification and authentication [3]. Unlike physiological biometrics (e.g., fingerprint, iris scan), behavioral biometrics are dynamic and can be monitored continuously and unobtrusively.

Mouse dynamics is a particularly potent form of behavioral biometrics for several reasons [5]:

- Ubiquity: The mouse is a standard input device for most desktop and laptop interactions, requiring no additional hardware for data collection.
- Unobtrusiveness: Data can be collected passively in the background without interrupting or altering the user's natural workflow.
- Rich Data Stream: A single user session can generate thousands of data points (e.g., x-y coordinates, timestamps, click events), providing a rich dataset for analysis.
- Difficulty of Forgery: The subconscious neuromuscular patterns that govern fine motor control are highly individualized and difficult for both humans and machines to replicate perfectly.

These characteristics make mouse dynamics an ideal candidate for building a security layer that is both highly effective and transparent to the end-user, directly addressing the usability failures of traditional CAPTCHAS.

EVOLUTION OF MOUSE DYNAMICS ANALYSIS

The application of mouse dynamics to user verification and bot detection has evolved significantly, mirroring broader trends in machine learning from classical, feature-based models to complex, end-to-end deep learning architectures.

Early Approaches: Statistical Features and Classical Machine Learning

Initial research in mouse dynamics focused on identifying and engineering a set of discriminative features from the raw stream of cursor data. These features were designed to capture the statistical properties of human movement and were then fed into classical machine learning classifiers like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), or Random Forests [8, 9].
Commonly extracted features include [9]:

- Kinematic Features: Metrics related to the physics of the movement, such as velocity, acceleration, and jerk (the rate of change of acceleration).
- Geometric Features: Characteristics of the trajectory's path, including curvature, straightness, and the number of inflection points.
- Temporal Features: Metrics related to timing, such as the duration of movements, the length and frequency of pauses, and click speed.
- Aggregate Statistics: Statistical measures (mean, variance, standard deviation, etc.) applied to the above features over a given time window or entire session.

While these methods demonstrated the viability of mouse dynamics for authentication, they were heavily dependent on the quality of the handcrafted features. This ”feature engineering” step is often domain-specific and can be brittle; as bots become more sophisticated, they can be programmed to mimic these specific statistical distributions, thereby evading detection [8].

The Deep Learning Revolution: End-to-End Models

The advent of deep learning provided a powerful alternative to manual feature engineering. By processing raw or minimally-processed sequential data directly, deep neural networks can learn complex, hierarchical feature representations automatically. This has led to significant performance improvements in bot detection.

Sequence Learning with LSTMs

Mouse movement data is fundamentally a time-series of (x, y, t) coordinates. Recurrent Neural Networks (RNNs), and particularly Long Short-Term Memory (LSTM) networks, are exceptionally well-suited for this type of data. LSTMs are designed to learn long-range dependencies and temporal patterns in sequential data, making them ideal for capturing the subtle, continuous flow of human cursor movement [8, 10, 11]. Several studies have successfully applied LSTMs to classify mouse trajectories, demonstrating their ability to outperform models based on static, handcrafted features [11].

Pattern Recognition with CNNs

While commonly associated with image analysis, Convolutional Neural Networks (CNNs) have also been adapted for mouse dynamics in two primary ways:
1. 1D-CNNs for Time-Series: A one-dimensional CNN can be applied directly to the sequence of mouse data to extract local temporal patterns. The convolutional filters act as motif detectors, identifying characteristic short-term patterns in velocity, acceleration, or direction that are indicative of human or bot behavior [8].
2. 2D-CNNs for Trajectory Images: A novel and highly effective approach involves converting a mouse trajectory into a 2D image. This image can encode not only the spatial path (x-y coordinates) but also kinematic information like velocity or acceleration through pixel intensity or color channels. This representation allows powerful, pretrained 2D-CNN architectures (like VGG or ResNet) to be used for feature extraction and classification, effectively transforming the time-series problem into an image recognition task [12]. This method has shown remarkable success in detecting bots with advanced statistical attack capabilities [12].

Hybrid models that combine CNNs for local feature extraction and LSTMs for learning temporal dependencies (CNN+LSTM) have also been proposed, often achieving state-of-the-art results by leveraging the strengths of both architectures [8, 13].

DATASETS, CHALLENGES, AND FUTURE DIRECTIONS

The progress in mouse dynamics research is intrinsically linked to the availability of highquality data and the continuous evolution of bot capabilities. This section reviews the datasets that fuel research and the advanced challenges that define the field’s future.

Public Datasets for Benchmarking

The development and validation of new models depend on public, standardized datasets. Several key datasets have been instrumental in the field:

- The Balabit Dataset: Released in 2016, this was one of the earliest and most widely used datasets. It contains mouse movement data from users performing their regular daily activities over a remote desktop, making it a valuable resource for studying natural, ”in-the-wild” behavior [5].
- The SapiMouse Dataset: This dataset, introduced in 2020, includes data from 120 subjects across multiple sessions, providing a larger and more structured collection for user authentication research [14].
- The ReMouse Dataset: Published in 2023, this dataset uniquely addresses the challenge of session-replay bots. It is the first public dataset to contain multiple, repeated sessions from the same users performing the same task. This allows researchers to study intrauser variability and develop models that can distinguish between a human legitimately repeating a task and a bot replaying a recorded session [5, 15, 16].

Despite these resources, there remains a need for larger, more diverse datasets that include telemetry from a wider range of devices and capture the behavior of the latest generation of sophisticated bots.

Advanced Threats: The Evolving Bot Landscape

As detection methods improve, so do the bots designed to evade them. The research frontier is now focused on two major challenges.

Session-Replay Bots

The latest generation of sophisticated bots no longer relies on simple, programmatic movements. Instead, they employ ”session replay” attacks, where they capture a genuine human’s interaction with a website and then replay that exact mouse trajectory to bypass security [5]. These are exceptionally difficult to detect because the replayed trajectory is, by definition, a perfect replica of human behavior. Research in this area, spurred by datasets like ReMouse, focuses on the hypothesis that genuine human movements are never perfectly repeatable. By analyzing the subtle, stochastic differences between a user’s own repeated sessions, models can learn to identify the unnatural exactness of a replayed one [5, 16].

Adversarial Machine Learning

A more fundamental threat to any machine learning-based security system is the concept of adversarial attacks. Adversarial machine learning is a field dedicated to fooling ML models by introducing maliciously crafted input data [17]. In the context of mouse dynamics, an attacker could generate a synthetic mouse trajectory that is only slightly perturbed from a standard bot movement but is specifically designed to be misclassified as human by the detection model [18]. These attacks can be ”white-box” (if the attacker has full knowledge of the model’s architecture and parameters) or ”black-box” (if the attacker can only query the model and observe its output). Research has shown that it is possible to generate adversarial mouse trajectories that successfully evade even deep learning-based authentication systems [18]. This represents a critical vulnerability, and future work must focus on developing robust defense mechanisms, such as adversarial training, where the model is explicitly trained on adversarial examples to make it more resilient [19].

SYNTHESIS AND IDENTIFICATION OF RESEARCH GAPS

The literature demonstrates a clear and successful progression in using mouse dynamics for bot detection. The field has matured from simple statistical models to powerful deep learning architectures capable of achieving high accuracy. However, this survey also reveals several critical gaps and open research questions that must be addressed:

1. Robustness to Adversarial Attacks: While the vulnerability of mouse dynamics systems to adversarial attacks has been demonstrated, the development of effective and practical defense mechanisms is still in its early stages. Most current models are not explicitly designed to be robust against such targeted manipulations.

2. Scalability and Real-Time Performance: For a system to be practical, it must operate in near real-time with minimal latency and computational overhead. Many academic models are computationally expensive, and there is a need for research into more efficient architectures (e.g., model quantization, lightweight networks) suitable for large-scale web deployment.

3. Data Scarcity and Diversity: Despite the existence of public datasets, there is a continuous need for larger and more diverse data. This includes data from different demographics, devices (e.g., trackpads, high-DPI mice), and, most importantly, data generated by the latest generation of human-mimicking and adversarial bots.

4. **Template Aging and Concept Drift:** A user's biometric signature is not static; their mouse behavior can change over time due to factors like practice, fatigue, or even a new mouse. This phenomenon, known as ”template aging,” can lead to increased false rejections. Systems need to incorporate mechanisms for continuous learning and adaptation to handle this ”concept drift.”

5. **Multi-Modal Fusion:** While mouse dynamics is powerful, it is not the only behavioral signal available. Future systems could achieve even greater accuracy and robustness by fusing mouse data with other behavioral biometrics, such as keystroke dynamics or scroll patterns, creating a more comprehensive user profile.

CONCLUSION

This literature survey confirms that bot detection based on mouse dynamics is a vibrant and highly promising field of research. The failure of traditional CAPTCHA systems has created an urgent need for a new security paradigm that is both effective against AI-driven threats and seamless for legitimate users. Behavioral biometrics, and specifically mouse dynamics, perfectly fits this requirement.

The evolution from handcrafted statistical features to end-to-end deep learning models like LSTMs and CNNs has enabled the development of systems with remarkable classification accuracy. However, the battle for web security is dynamic. The emergence of sophisticated threats like session-replay bots and the looming challenge of adversarial attacks mean that accuracy alone is no longer sufficient. The next generation of bot detection systems must be designed with an explicit focus on robustness, scalability, and adaptability.

The research gaps identified in this survey—particularly the need for adversarial robustness and the creation of more comprehensive datasets—provide a clear and compelling direction for future work. The project on ”Bot Detection with Mouse Dynamics” is well-positioned to address these gaps by developing a state-of-the-art sequence learning model, rigorously testing it against sophisticated bot behaviors, and exploring its resilience to adversarial manipulation. By doing so, this work can make a significant contribution to creating a more secure and userfriendly web.

References

A. Survey of Botnet and Botnet Detection Methods. *International Journal of Engineering Research and Technology*, 2013. CyberPeace Foundation, ”Who is Winning the War with AI: Bots vs. CAPTCHA,” 2024.
[Online]. Available:
https://www.cyberpeace.org/resources/blogs/who-is-winning-the-war-with-ai-bots-vs-captcha.

S. Hashia et al., ”Mouse Dynamics Behavioral Biometrics: A Survey,” ResearchGate, 2022.

F. Fourati et al., ”Nonintrusive behavioural biometrics for computer system security,” 2020.

S. Sadeghpour and N. Vlajic, ”ReMouse Dataset: On the Efficacy of Measuring the Similarity of Human-Generated Trajectories for the Detection of Session-Replay Bots,” *Journal of Cybersecurity and Privacy*, vol. 3, no. 1, pp. 95-117, 2023.

Radware, ”4 Botnet Detection Techniques,” [Online]. Available: https: //www.radware.com/cyberpedia/hot-management/4-botnetdetection-techmques/.

A. Survey on Botnet Detection Techniques, ResearchGate, 2020.

H. Niu, J. Chen, Z. Zhang, and Z. Cai, ”Mouse Dynamics Based Bot Detection Using Sequence Learning,” in *Biometric Recognition, CCBR 2021*, Shanghai, China, 2021, pp. 49-56.

User Authentication Based on Mouse Dynamics, ResearchGate, 2019.

M. A. A. Al-qaness et al., ”Using Deep Learning for Trajectory Classification,” in *International Conference on the Quality of Information and Communications Technology*, 2021.

G. Kramida et al., ”Mouse behavior classification using deep learning,” in *IEEE International Conference on Image Processing*, 2016.

A. Wei, Y. Zhao, and Z. Cai, ”A Deep Learning Approach to Web Bot Detection Using Mouse Behavioral Biometrics,” in *Lecture Notes in Computer Science*, 2019.

L. Zhao et al., ”A Hybrid CNN-LSTM Model for Trajectory Prediction,” *Sensors*, 2022.

M. Antal, ”SapiMouse a new dataset for Mouse Dynamics,” GitHub Repository, 2020. [Online]. Available:
https://github.com/margitantal68/sapimouse.

S. Sadeghpour and N. Vlajic, ”ReMouse Dataset: On the Efficacy of Measuring the Similarity of Human-Generated Trajectories for the Detection of Session-Replay Bots,” dblp, 2023.

S. Sadeghpour and N. Vlajic, ”RanABD: MTD-Based Technique for Detection of Advanced Session-Replay Web Bots,” 2023.

Center for Long-Term Cybersecurity (CLTC), UC Berkeley, ”Adversarial Machine Learning.” [Online]. Available:
https://cltc.berkeley.edu/anil/.

Y. X. M. Tan et al., ”Adversarial Attacks on Remote User Authentication Using Behavioural Mouse

Dynamics,” arXiv preprint arXiv:1905.11831, 2019.

Wikipedia, ”Adversarial machine learning.” [Online]. Available: https://en.wikipedia.org/wiki/Adversarial_machine_learning.

P. Singh. ”Mouse Movement Behavioral Patterns Can Reliably Tell Bots from Humans,” 2025.

I. Pozzana and E. Ferrara, ”Measuring Bot and Human Behavioral Dynamics,” *Frontiers in Physics*, 2020.

”Redefining Security: Unveiling the Vulnerabilities of Captcha Mechanisms Using Deep Learning,” ResearchGate, 2024.