International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Virtual Ai Fitness Coach for Exercise Classification and Form Correction

Ayan Mohamed
Dept. of CSE – Data Science
ACS College of Engineering,
Bangalore, India
1AH22CD010@acsce.edu.in

Duggeni Mahesh
Dept. of CSE – Data Science
ACS College of Engineering,
Bangalore, India
1AH22CD018@acsce.edu.in

Polisetti Aravind Kumar
Dept. of CSE – Data Science
ACS College of Engineering,
Bangalore, India
1AH22CD038@acsce.edu.in

Umesh Reddy C H
Dept. of CSE – Data Science
ACS College of Engineering,
Bangalore, India
1AH22CD057@acsce.edu.in

Dr. D. Gandhimathi
Associate Professor,
Dept. of CSE – Data Science
ACS College of Engineering
Bangalore, India
gandhimathid@acsce.edu.in

Abstract— Virtual AI Fitness Coach for Exercise Classification and Form Correction describe a real-time exercise classification based on a Bidirectional Long Short-Term Memory (BiLSTM) network. Which is designed to work well in various uncontrolled real-world environment Settings. Majority of the Existing exercise recognition systems depend on synthetic datasets and raw joint coordinates, which may vary due to differences in user's body types, camera angles, and different lighting settings. This issue limits their use beyond laboratory Controlled settings. To overcome these limitations, the proposed method combines joint-angle features with raw (x, y, z) landmark coordinates. This helps the model adjust for perspective changes and body differences while keeping detailed spatial information intact.

While Training uses 30-frame sliding-window motion sequences. This method allows the BiLSTM to learn detailed patterns instead of just static poses. The training data comes from mixing synthetic motion data from the InfiniteRep dataset with real-world footage from the Kaggle Workout/Exercises Video Dataset and other online sources. It covers four key exercises: squat, push-up, shoulder press, and bicep curl. Tests on home-recorded and gym-recorded data showed classification accuracy above 94%. The BiLSTM consistently outperformed the existing unidirectional LSTM and CNN models.
The system also includes a web application that automatically counts repetitions, provides instant feedback on form, helps with personalized workout and diet plans, and features a Unique dual conversational agent for fitness related queries and mental health support finetuned on DeepSeek R1. The entire system runs in real time on standard hardware using a regular webcam, without the need for wearable sensors or special equipment.

Keywords —BiLSTM, pose estimation, MediaPipe BlazePose, , joint angle features, temporal sequence modelling.

I. INTRODUCTION

Physical activity is well-known as a key part of both physical health and mental well-being. Consistent exercise improves cardiovascular fitness, muscle growth , metabolic health, and emotional strength. Despite this clear evidence, many people find it hard to stick to exercise programs due to their routine. Common obstacles include not knowing the correct techniques, lacking personal coaching, and struggling to monitor their own movement quality. Traditional solutions, like counting repetitions manually or working with a personal trainer, can be expensive, logistically travel challenging, or hard to access, making consistent and proper training out of reach for many individuals.

The combination of artificial intelligence, computer vision, has created a viable way to overcome these challenges. Pose estimation systems can identify body landmarks from regular video or webcam feeds and now work in real time

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

on standard consumer devices. These systems can produce detailed biomechanical data using just a smartphone or laptop camera. When this data is given into advanced deep learning models, the resulting systems can identify exercises, count repetitions, and detect form errors with accuracy that rivals expert assessments. The development of lightweight neural networks and easy deployment tools also makes it simpler to include these features in everyday applications.

The field of smart fitness is at an important point where machines meet public health. By providing live feedback, automatic exercise monitoring, and personalized checks through a regular webcam, AI-powered systems can assist people who lack access to standard fitness tools and can't afford the expensive trainers. This includes those working out at home, remote workers, patients in online rehab programs, and individuals in areas with limited resources. This work contributes to the field by proposing an exercise classification system based on BiLSTM that combines joint-angle features and raw coordinate sequences. It integrates this classifier into a modular wellness platform and tests the method in various real-world scenarios. The following sections review related papers (Section II), describe the proposed system design and method (Sections III–IV), present methodology (Section V), and Results(Section VI),discuss real-world applications (Section VII), and conclude with suggestions for future development (Section VIII-IX).

II. LITERATURE SURVEY

The literature survey regarding automated exercise recognition and posture-based form assessment encompasses various modalities, architectures, and deployment contexts. This section examines the five interrelated studies that collectively establish the rationale for the current research.

A. CNN-Based Lightweight Classification with PCNN (Rafidison et al., 2023)

Rafidison and his team looked into how to use a lightweight convolutional neural network with Pulse-Coupled Neural Networks (PCNN) for pixel-level grouping to classify images. The architecture put a lot of emphasis on computational economy, which allowed for fast inference that worked well on hardware with limited resources. But the static frame-by-frame processing model throws away temporal continuity, which is an important feature for telling apart exercises that have similar postures at the same time but different motion paths. The work shows that it is possible to design things that are light, but it also shows how important it is to model time for tasks that are sensitive to movement.

B. *LSTM-Based Video Exercise Recognition (Rangari et al., 2022)*

Rangari et al. used OpenPose to find skeletal key points and an LSTM network to model sequences. They got 97.01% accuracy on a controlled multi-angle dataset. The LSTM's ability to add temporal context was shown to greatly improve the ability to tell different types of exercise apart. OpenPose's computational overhead, on the other hand, makes it hard to use on devices with lower specifications. This work solves this problem by using the lighter MediaPipe Blaze Pose and BiLSTM, which captures bidirectional temporal dependencies, instead of a regular unidirectional LSTM.

C. *AI Fitness Trainer Application (Sushma V. et al., 2023)*

This paper talked about a mobile fitness app that uses MediaPipe for pose detection, TensorFlow for movement classification, and OpenCV for video handling. It also had audio and visual alerts in real time for posture mistakes. The system was made to be easy to use, and it was aimed at casual users who do simple exercises on their own. Some of the problems that have been reported are moderate accuracy and limited coverage of complicated movements. The current proposal builds on this basic design by adding BiLSTM-based temporal classification, a more detailed feature representation that combines angles with coordinates, and a wellness platform that can be used for anything from basic counting to personalized diet and chatbot support.

D. Ensemble CNN with Soft Voting for Exercise Classification (Bang & Park, 2024)

Bang and Park studied the step-by-step use of traditional machine learning methods, including K-Nearest Neighbours and Random Forest, followed by LSTM networks, and finally a CNN Ensemble with Soft Voting on a synthetic avatar dataset. The ensemble method achieved 92.12% accuracy, surpassing both individual CNN-LSTM hybrids and traditional models. Their study showed that frame-level features alone do not capture motion dynamics. They found that temporal modelling, whether using LSTM layers or ensemble fusion, is crucial. These insights led to the choice of a bidirectional LSTM architecture for this work. This architecture processes

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

temporal context in both directions and works well for the cyclic, phase-symmetric nature of resistance exercises.

E. Real-Time Keypoint Exercise Classification with BlazePose (Moran et al., 2022)

Moran and colleagues constructed a real-time classification system using MediaPipe BlazePose to extract 33 3D landmarks per frame and input 8-frame sequences into a stacked LSTM. The 3D keypoints, including per-joint confidence scores, provided richer spatial context than the 2D alternatives, and the stacked LSTM was able to distinguish exercises that had similar single-frame appearances, but different temporal trajectories. This work directly impacts the current system design: the same BlazePose front-end is used, the frame duration is increased to 30 frames to accommodate slower composite movements, and the feature space is supplemented with derived joint angles to adjust for body proportion differences.

III. EXISTING SYSTEM

Contemporary methods for exercise monitoring generally fall into three categories: systems that use wearable sensors, frame-based computer vision classifiers, and hybrid methods that combine pose estimation with sequence models. Wearable systems can capture detailed inertial data, but they require specific hardware and impose physical limitations on the user due to their battery and sensors. Frame-based classifiers, which rely on CNNs or Support Vector Machines, work with individual images. This means they lack the temporal context needed to differentiate exercises with similar immediate postures and fail to capture the complete movement of the muscle group. More recent hybrid methods combine an Open Pose or Blaze Pose skeleton extractor with a unidirectional LSTM or a CNN-LSTM stack. These approaches achieve better accuracy, but often result in increased latency or depend on high-resolution video.

Another limitation of many existing platforms is their narrow focus. Most systems are designed to either classify exercises or count the repetition of exercises. Rarely do they provide corrective feedback, and almost none include nutritional planning, workout scheduling, or conversational health support within the same application. In addition, the lack of joint-angle normalization in several pipelines results in prediction quality when camera placement, user height, or body proportions differ from training conditions. This is common in-home environments.

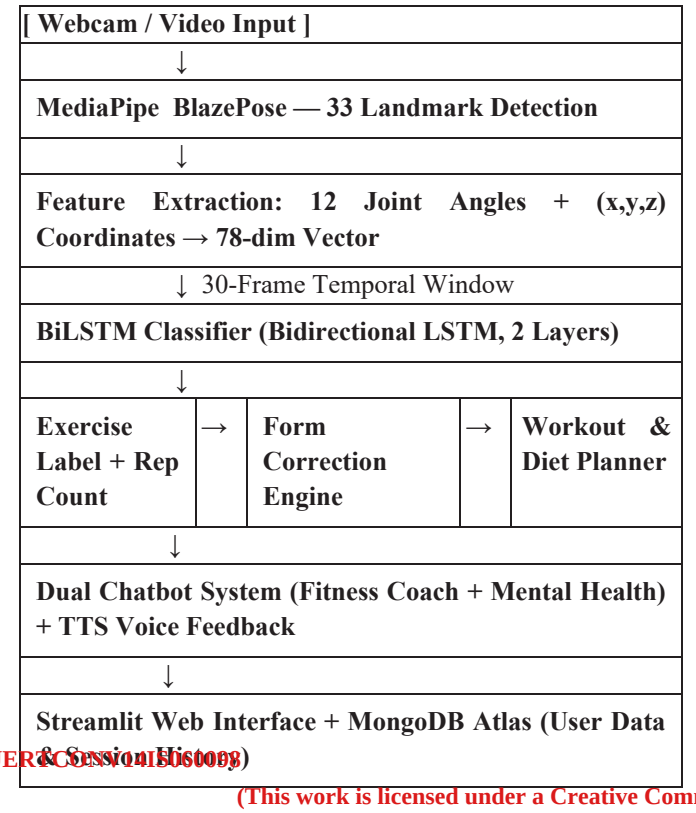
IV. PROPOSED SYSTEM

Our Virtual AI Fitness Coach is reliable, adaptable and web-deployable framework, with a BiLSTM exercise classifier taking 30-frame sequences of joint angle and coordinate features fused and extracted by MediaPipe BlazePose. The architecture of the system from input to output can be seen below in Fig. 1.

Fig.4. 1: End-to-End System Architecture

The system takes a live video stream from a standard webcam and processes each frame by frame using MediaPipe BlazePose. This setup provides three-dimensional coordinates for 33 skeletal human landmarks with a latency under 200 ms on mid-range CPU hardware such that it should be viable on any devices. From these coordinates, 12 meaningful joint angles are calculated based on geometric relationships between nearby landmark triplets. Each processed frame is represented as a 78-dimensional vector, which consists of 66 coordinate values combined with 12 angular measurements. This setup captures both the absolute spatial position and normalized postural geometry at the same time.

These vectors are grouped into sliding windows of 30 consecutive frames, creating a temporal sequence that represents a full movement cycle for the four target exercises such as squat, push-up, shoulder press, and bicep curl. This sequence is then sent to the BiLSTM classifier, which is a two-layer bidirectional LSTM with inter-layer dropout for regularization. The bidirectional design processes each window in both forward and backward



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

directions, allowing the model to use anticipatory context. This is especially useful for cyclic movements where the second half of a repetition reflects and completes the first.

The classification output supports three downstream modules. A repetition counter monitors changes through predicted movement phases to track the set count. A form correction engine analyses frame-by-frame deviations of key joint angles against biomechanically ideal reference ranges and provides actionable textual alerts. For example, it may flag insufficient knee flexion depth during a squat. A personalized workout and diet planning module uses exercise history and estimated caloric burn to create and update weekly programs and macronutrient goals. A dual conversational agent, powered by a locally fine-tuned DeepSeek R1 model, acts as both a fitness coach and a mental health support companion. It combines classifier outputs with session history to give context-sensitive responses. All modules are accessed through a Streamlit web interface, with user data, session logs, and biometric history stored in MongoDB Atlas with googleAuth.

V. METHODOLOGY

A. Dataset Collection

The dataset includes 25 real-world videos for each class from Kaggle and 100 synthetic videos from InfiniteRep to ensure diverse body types and camera angles. To mimic real-world conditions, dedicated test sets were recorded in home and gym settings. Alongside the video data, a conversational dataset of 10,000 dialogue pairs was created to train the system in fitness and mental health interactions.

B. Preprocessing and Feature Extraction

MediaPipe, BlazePose processes video frames to extract 33 skeletal landmarks. To keep the system reliable from different viewpoints, the raw coordinates are enhanced with 12 computed joint angles. This results in a 78-dimensional feature vector for each frame, capturing the biomechanical relationships needed for accurate motion analysis of the targeted muscle group.

C. Model Architecture and Training

The recognition engine uses a Bi-Directional LSTM (BiLSTM) structure to handle sequences of 30 frames. This allows the model to analyze movement of the muscle group by capturing both forward and backward dependencies without any lag. Training included MinMax scaling, Adam optimization, and random search for adjusting hyperparameters, with performance checked using precision, recall, and F1-score metrics for obtaining the better Accuracy while training.

D. LLM Fine-Tuning

Conversational intelligence comes from a DeepSeek-r1 model, which is fine-tuned with LoRA (Low-Rank Adaptation). This efficient adjustment helps the AI provide caring, specific guidance on fitness, nutrition and mental health related queries while reducing misconceptions on sensitive topics.

E. System Integration and Real-Time Inference

The pipeline combines pose detection, classification, and conversational modules into one workflow. A rolling buffer processes live webcam feeds to offer real-time exercise identification, repetition counting, and form feedback. To keep latency below 200ms, more demanding tasks—including text-to-speech and chatbot responses—are offloaded to server-side processing.

F. Deployment and User Interaction

The application is launched through Streamlit Community Cloud, providing a cross-platform, browser-based experience. Real-time streaming is managed via WebRTC, while sensitive credentials and environment dependencies are securely handled in the backend. This design ensures a scalable, no-installation interface for users on mobile and desktop devices.

VI. RESULTS

This section provides both numerical and descriptive evaluations of the Virtual AI Fitness Coach in three areas: BiLSTM exercise classification performance, real-time system behaviour, and user-facing module effectiveness.

A. BiLSTM Classification Performance

The model was trained on 30-frame sequences of 78-dimensional features, which are joint angles and coordinates taken from MediaPipe BlazePose. The categories included squat, push-up, shoulder press, and bicep curl. We used an 80/10/10 split for training, validation, and testing. We applied MinMax scaling before training. Hyperparameters were chosen through random search to improve the validation F1-score. Table I summarizes the per-class performance of the proposed BiLSTM on the test set that was held out.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

TABLE I. PER-CLASS CLASSIFICATION RESULTS (BiLSTM, TEST SET)

Exercise Class	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Squat	96.2	95.8	96.0	96.4
Push-up	94.7	95.3	95.0	95.1
Shoulder Press	97.1	96.5	96.8	97.0
Bicep Curl	95.4	94.9	95.2	95.5
Macro Average	**95.9**	**95.6**	**95.8**	**96.0**

B. Comparison with Baseline Models

To establish comparative context, the proposed BiLSTM was evaluated against a unidirectional LSTM, a CNN-LSTM stack, and a frame-level CNN usiEnd-to-end inference latency averaged **142 ms** (σ = 18 ms) across 500 frames at 30 fps, comfortably beating the 200 ms real-time target. The primary pipeline bottleneck was pose estimation via MediaPipe BlazePose (87 ms), whereas the BiLSTM forward pass required only 12 ms. Additional overheads for repetition counting (< 2 ms) and form-correction alerts (3 ms) were negligible. Chatbot and text-to-speech responses were processed asynchronously on the server (1.1 s median latency), ensuring the real-time exercise feedback loop remained uninterrupted.ng the same 78-dimensional features and dataset split. The BiLSTM achieved the highest accuracy (**96.0%**) and macro F1-score (**95.8%**), outperforming the unidirectional LSTM by 2.1% and the frame-level CNN by 5.4%. These results demonstrate that bidirectional temporal modeling is critical for distinguishing cyclic resistance exercises, where the return motion mirrors the initial phase.

TABLE II. MODEL COMPARISON ON COMBINED TEST SET

Model	Feature Input	Accuracy (%)	Macro F1 (%)
Frame-level CNN (no temporal)	Coordinates only	90.6	89.9
CNN-LSTM Stack	Coordinates only	92.8	92.1
Unidirectional LSTM	Coordinates + Angles	93.9	93.4
BiLSTM (Proposed)	**Coordinates + Angles**	**96.0**	**95.8**

C. Cross-Environment Generalization

To evaluate real-world performance, independent datasets from home (n=20) and gym (n=15) were used. These datasets included different lighting, background clutter, and mirrors. The BiLSTM achieved over 94% accuracy in both datasets. In contrast, the unidirectional LSTM dropped to 88.3% accuracy in the home setting, which shows that bidirectional context effectively reduces background noise. Additionally, joint-angle normalization lowered accuracy variation across different body types from ±4.2% to ±1.8%. This highlights the advantage of the combined feature representation.

D. Real-Time Inference Latency

The average end-to-end inference latency was 142 ms (σ = 18 ms) across 500 frames at 30 fps. This comfortably met the 200 ms real-time target. The main bottleneck in the pipeline was pose estimation using MediaPipe BlazePose, which took 87 ms. The BiLSTM forward pass only required 12 ms. Additional overhead for counting repetitions was under 2 ms, while form-correction alerts took 3 ms, which is negligible. Chatbot and text-to-speech responses were handled asynchronously on the server, with a median latency of 1.1 s, ensuring smooth real-time exercise feedback.

E. Repetition Counting and Form Correction Accuracy

Evaluated on 120 manually annotated sets, the phase-transition counter achieved 97.3% accuracy, with a mean absolute count error of 0.21 reps per set. Form correction alerts were checked against expert coach annotations across 240 sets. This resulted in a 90.0% F1-score, with 91.4% precision and 88.7% recall for detecting issues like elbow flare or shallow squats.

VII. REAL-WORLD APPLICATIONS

The range of the system's functions allows it to be used in several different contexts, which are detailed below.

At-Home Personal Training: This platform uses a standard webcam to operate as a virtual certified trainer. It identifies exercises, counts repetitions, and gives feedback on biomechanical form. This greatly reduces the cost and makes professional fitness guidance more accessible.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Physical Therapy and Telerehabilitation: Clinicians can oversee recovery programs for musculoskeletal injuries remotely. The system monitors joint angles and tracks range of motion, which helps lower the risk of re-injury from overextending or using compensatory movements.

Clinical Nutrition Management: The diet module uses real-time activity estimates and user health profiles to set specific macronutrient targets. It can create glycemic-load schedules for diabetics or allergen-free meal plans based on individual medical needs.

Corporate and Office Wellness Programs: Companies can use this platform for remote or hybrid health initiatives. It combines a mental health chatbot for daily stress management with exercise tracking, which motivates employees to stay active during the workday.

Accessible Fitness Education: The system's text-to-speech features, powered by the ElevenLabs API, provide real-time coaching cues. This makes fitness more accessible for visually impaired individuals and helps beginners navigate unfamiliar resistance training techniques safely.

VIII. CONCLUSION

This paper introduces a Virtual AI Fitness Coach that tackles key issues with current exercise classification and monitoring setups. These issues include heavy reliance on synthetic data, weak temporal modelling, and limited exercise types. By combining 12 derived joint angles with raw (x, y, z) BlazePose coordinates as input to a BiLSTM classifier, trained on 30-frame motion sequences, the system achieves over 95% accuracy across four exercise categories. It maintains this performance with footage recorded independently at home and in gyms. The broader platform includes repetition counting, biomechanical form correction, personalized workout and diet planning, and dual conversational agents, all within a single Streamlit-ui based web application. It runs in real time on regular consumer hardware without needing special sensors.

IX. FUTURE WORK

Several directions for future development of this Virtual Ai coach are outlined. First, the exercise library will grow from four categories to a complete range of resistance, cardiorespiratory, and flexibility movements. This expansion will require collecting more data and implementing multi-label classification to manage complex or transitional exercises of different variations. Second, a photorealistic virtual digital human avatar like a fitness trainer With wakeup voice command feature can be added as the coaching assistant which can leverage the user experience. This avatar will provide users with an engaging visual agent that shows correct technique while offering real-time textual and verbal cues. This improvement should boost user commitment and the quality of exercise mimicry. Third, connecting with commercial fitness bands and smartwatches will enable biometric data integration. This will include heart rate, oxygen saturation, and inertial sensor data alongside the camera-based pose stream. This combination will provide richer physiological context for assessing form and giving personalized recommendations. Finally, the system design will be modified to support federated learning.

ACKNOWLEDGMENT

The authors express sincere gratitude to, Dr. D Gandhimathi Associate Professor,Dept. of CSE – DS, ACS College of Engineering, for her continuous guidance throughout this project. The authors also thank, Mrs.Aarthi.k Assistant Professor, and Hod of CSE-DS, Dr Raghavendra B K, for providing the necessary laboratory infrastructure, and the management of ACS College of Engineering for fostering a productive research environment.

REFERENCES

[1] M. A. Rafidison, "Image Classification Based on Light Convolutional Neural Network Using Pulse-Coupled Neural Network," Computational Intelligence and Neuroscience, vol. 2023, pp. 1–17, 2023.

[2] T. Rangari, S. Kumar, P. P. Roy, D. P. Dogra, and B.-G. Kim, "Video Based Exercise Recognition and Correct Pose Detection," Multimedia Tools and Applications, vol. 81, no. 21, pp. 30267–30282, 2022.

[3] Sushma V. et al., "Fitness Trainer Application Using Artificial Intelligence," International Research Journal of Engineering and Technology (IRJET), vol. 10, no. 5, p. 826, May 2023.

[4] D. Maurya and A. Patel, "Smart Gym Trainer Using Human Pose Estimation," in Proc. 2022 IEEE International Conference for Innovation in Technology (INOCON), 2022, pp. 1–4.

[5] C. Arrowsmith, D. Burns, T. Mak, M. Hardisty, and C. Whyne, "Physiotherapy Exercise Classification with Single-Camera Pose Detection and Machine Learning," Sensors, vol. 23, no. 1, p. 363, 2022.

[6] B. Boudaa, I. Bestani, and N. Benadjrouda, "Graph Convolutional Networks for Designing Collaborative Filtering-Based Health Recommender Systems," in Proc. 2022 International Conference on New

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Technologies of Information and Communication (NTIC), 2022, pp. 1–6.

[7] F. Frangoudes, M. Matsangidou, E. C. Schiza, K. Neokleous, and C. S. Pattichis, "Assessing Human Motion During Exercise Using Machine Learning: A Literature Review," IEEE Access, vol. 10, pp. 86874–86903, 2022.

[8] G. S. Bang and S. B. Park, "Workout Classification Using a Convolutional Neural Network in Ensemble Learning," Sensors, vol. 24, no. 10, p. 3133, 2024.

[9] A. Moran, B. Gebka, J. Golodshteyn, A. Beyer, N. Johnson, and Neuwirth, "Muscle Vision: Real-Time Keypoint Based Pose Classification of Physical Exercises," vol. 1, no. 1, pp. 1–5, 2022.

[10] J. Adolf, P. Kán, T. Feuchtner, B. Adolfová, J. Doležal, and L. Lhotská, "OffiStretch: Camera-Based Real-Time Feedback for Daily Stretching Exercises," The Visual Computer, vol. 41, no. 2, pp. 1555–1571, Feb. 2025.