SEMINAR NASIONAL 2026 Universitas Ma Chung
Prosiding Seminar Nasional Universitas Ma Chung 2026
(Sistem Informasi dan Informatika)
pp. 492-510
E-ISSN: 2808-7771

Recurrent Neural Network for Human Fall Motion Prediction

Bintang Rizqi Pasha[1], Yesy Diah Rosita[2], Andi Prademon Yunus[3*]

[1,2,3]Teknik Informatika, Telkom University, Jl. DI Panjaitan No 128, Banyumas, Indonesia, 53147

***Email korespondensi:** andiay@telkomuniversity.ac.id

Received: 27 05 2026 – Revised: 29 07 2026 - Accepted: 04 08 2026 - Published: 05 08 2026

Abstract. Falls pose a significant health risk, especially for older adults, where one in three people over 65 experiences a fall each year. For those over 85, the consequences of falls can be severe, leading to life-threatening injuries and a marked decline in quality of life. To address the critical need for fall prevention, this study proposes a prediction approach using Recurrent Neural Networks (RNNs) to recognize patterns in human motion that may indicate an impending fall. By utilizing the CAUCA fall dataset—carefully designed to detect abnormal fall movements—we implemented and assessed different RNN architectures, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). Our findings show that the GRU model performed best, achieving an accuracy of 0.716, MPJPAE of 32.921 pixels, MPJVE of 22.457 pixels per frame, and Euclidean distance metric outperforming the other models, with RNN closely following at an accuracy of 0.716, MPJPAE of 41.872 pixels, MPJVE of 21.44 pixels per frame. These promising results suggest that RNN models, particularly GRU, can serve as valuable tools in predicting falls, offering a foundation for future technology that can help prevent falls before they happen.

Kata kunci: Fall, Human fall motion prediction, Human motion prediction, RNN.

INTRODUCTION

Falls have become a significant risk to human viability, particularly among people over 65 years of age, and can be fatal for those over 85. Historical data indicates that approximately 30% of older people over 65 years of age suffer falls each year, which is a leading cause of injury and death in this group of older people (Christiansen et al., 2020; Tamura et al., 2007). The natural aging process often leads to a gradual decline in physical abilities, which can subtly increase the probability of accidental falls. Falls among older adults are frequently associated with a natural decline in physical functions, including reduced leg strength, which can limit stable footing. This loss of stability may leave them more vulnerable to slippage, tripping, or momentary lapses in balance, thus increasing the risk of falls (WHO, 2021). The decline in physical abilities can be caused by various factors, including reduced muscle strength, balance disorders, and slower reflexes (WHO, 2021). Thus, preventing possible falls among the elderly depends on Pilates activities.

According to a report by the World Health Organization, falls are the second most common cause of accidental deaths worldwide and contribute to injuries in several other

cases (WHO, 2021). Due to the biological changes that cause falls to increase greatly with maturity, the elderly face more injuries due to falls (Ozcan et al., 2005). Falls can occur spontaneously and affect individuals in all demographics—older adults, men, teens, and children—often without warning, such as slipping on stairs or tripping while walking.

The recognition of human activities is becoming more crucial as unusual behavior might lead to serious injuries. In recent years, advances in technologies such as machine learning and vision-based algorithms have enabled more effective recognition of human activities, as evidenced by previous studies (Huang & Li, 2023; Lau et al., 2022; Ruiz et al., n.d.; Yunus et al., 2023). Recent works in machine learning and deep learning have significantly improved the ability to predict human activities. Thus, their research aims to develop a predictive model for the motion of human falls, expanding the foundational research to improve our understanding of fall dynamics and improve preventive measures (Lau et al., 2022; Usmani et al., 2021). Their studies have used methods such as CNN, LSTM, GRU, and attention mechanisms, using human pose estimation on images as features to predict falls (Lau et al., 2022). Their approach achieved high accuracy using attention-based mechanisms.

Various methods and techniques for detecting human falls, including sensor-based and non-sensor-based approaches, as well as machine learning methods such as SVM, ANN, and Random Forest, are discussed in numerous literature reviews. Generally, SVM and wearable sensors are commonly employed for the detection and prevention of falls, with various studies indicating strong performance, although limitations are observed in controlled settings (Usmani et al., 2021). Previous studies have demonstrated that current technologies, particularly machine learning and deep learning methods, achieve strong accuracy in detecting human falls. Hence, the purpose of this study is to develop a prediction model for falls, which will be evaluated using certain performance measures to determine its reliability and utility.

Unlike previous studies (Lau et al., 2022; Mankodiya et al., 2022; Ramirez et al., 2021), which focused primarily on fall detection, in this paper we propose RNN-based methods to predict and classify fall human motion through a visual approach. Moreover, RNN is used to motivate the user to be effective in processing sequential data, making it suitable for predicting dynamic human movement. The input data originates from video frames obtained in the CAUCAFall dataset and have been transformed into pose key points

using the YOLOv8 pose model (Gu & Su, 2024; Guerrero et al., 2022). The sliding window technique divides the key data points into three segments, which are subsequently analyzed by RNN models including RNN, LSTM, and GRU. The metrics such as root mean square error (RMSE), L1 norm, mean per joint position absolute error (MPJPAE), mean per joint velocity error (MPJVE), mean per joint localization error (MPJLE), and Euclidean distance 2D are used to evaluate the model's performance. We believe that our research's main contribution is:

1. A baseline for future fall motion prediction.
2. The capability of Recurrent Neural Networks to predict falls.
3. The present study extended the scope of existing research by not only forecasting fall motion but also classifying the fall category.

RELATED WORKS

Human Fall Motion Prediction

Research in fall motion prediction has advanced through diverse methodologies, yielding notable results. A dual-channel feature integration technique was proposed to classify falls into falling-state and fallen-state, using a sliding window model for feature extraction. This method achieved a classification accuracy of 97.33%, with performance tested on the UR Fall and Le2i datasets, showing an accuracy of 96.91% (Wang et al., 2020).

Another approach utilized bone models and a positive semi-definite (PSD) matrix with symmetric and non-negative eigenvalues for analyzing motion. Dynamic Time Warping (DTW) was employed to measure the similarity of time-series motion sequences, achieving 100% sensitivity on the Charfi and URFD datasets and specificity rates of 87% and 93%, respectively. This method, however, was limited to a single human subject, restricting its scalability (Youssfi Alaoui et al., 2021).

A visual-based fall detection method was developed using a 360-degree camera to capture the entire surroundings. This approach overcame limitations of traditional visual systems and was tested on the Up-Fall and URFD datasets, demonstrating superior performance compared to state-of-the-art methods (Saurav et al., 2022).

In transformer-based fall prediction, a bidirectional transformer GAN was introduced. The method included an encoder-decoder transformer structure with soft-DTW for the

generator and dual discriminators for sequence differentiation. Although challenges were observed in forecasting foot motion during walking, the model effectively reduced stagnant pose generation in long-term predictions. This method achieved a 4% lower average error than other approaches (Zhao et al., 2023).

Human Motion Prediction

Falls often result from irregularities in human motion, which makes prior work in motion prediction critical to understanding this phenomenon.

A self-attention mechanism was used to predict both short-term and long-term human motions. The model reported an error rate of 23.51 pixels for short-term and 10.3 pixels for long-term forecasts using the Human 3.6M dataset. Metrics such as MPJPE and MPJVE were adopted for evaluation, and OpenPose was utilized for pose estimation. The sliding window method facilitated data processing, with long-term predictions outperforming short-term ones by maintaining results within a tolerance of 50 pixels (Yunus et al., 2023).

Further, a dual-attention and multi-granularity temporal convolutional network (DA-MgTCN) focused on capturing temporal evolution for human motion forecasting. Despite limitations in the hidden Markov posture model and its reliance on Gaussian processes and Boltzmann machines, the DA-MgTCN effectively captured spatial linkages and multi-scale temporal information. This method improved long-term predictions through a multi-granularity TCN approach (Huang & Li, 2023).

Another study examined real-world mobility patterns using deep networks, introducing frequency distribution-based metrics to evaluate the alignment between predicted and actual motion. Although effective in modelling authentic movements, the metrics were limited in adaptability to other models due to their specificity (Ruiz et al., n.d.).

A multiscale 3D skeleton model was applied using datasets like Human 3.6M and CMU Mocap. Employing graph neural networks with GRU and MGCU operators, the Dynamic Multiscale Graph Neural Network (DMGNN) demonstrated strong performance. However, limitations were observed in classifying specific actions, with errors of 29.18 ms for short-term (400 ms) and 86.04 ms for long-term predictions (1000 ms) (Li et al., n.d.).

Lastly, a comparative analysis between RNN-LSTM and Kalman Filters revealed that Kalman Filters outperformed RNN-LSTM in motion prediction. Using the COCO dataset and RGB cameras, the study showed that while pose estimations were consistent, RNN-LSTM models lacked stability (Yunus et al., 2022).

Human Fall Motion Prediction using RNN Based

Recurrent Neural Networks (RNN) and their variations have shown great potential in predicting fall motion. One study employed the Up-Fall dataset for fall detection using RNN algorithms, achieving a classification accuracy of 94.99% and an F1-score of 94.49% through skeleton pose estimation (Ramirez et al., 2021).

Another study tackled gradient vanishing issues in long-term learning by utilizing LSTM and Bidirectional LSTM models. Evaluated on the UpFall, SisFall, and UMAFall datasets, the FallNeXt network achieved 96.16% accuracy and a 99.12% F1-score. The skeleton-joint co-attention (SCA) mechanism was implemented on the h3.6m mocap dataset to dynamically learn motion over time, enabling accurate joint-level predictions (Shu et al., 2022).

A simpler yet effective RNN-LSTM approach modelled 2D fall motion dynamics, achieving 99% sensitivity on the FDD and URFD datasets (Hasan et al., 2019). Another ensemble RNN model combined Long Short-Term Memory with ensemble learning techniques such as XGBoost, AdaBoost, and Random Forest. Tested on the Smart-Fall dataset, the Random Forest method outperformed single LSTM models, showcasing enhanced prediction accuracy (Farsi, 2021).

Integrating OpenPose skeletal data with LSTM/GRU models, another study improved baseline performance by 9.3%, achieving 98.2% accuracy in fall detection through sequential learning of keypoints (Lin et al., 2020).

A visual-based method utilized monocular video to extract keypoints for motion prediction. Using a sequence-to-sequence (seq2seq) architecture, the model forecasted future postures for fall classification. Experiments showed high accuracy, with an F1-score of 94.4% and an accuracy of 97.8% on RGB-based evaluations (Hua & Lian, n.d.).

RNN-based methods, incorporating techniques like skeleton-joint co-attention and ensemble learning, have proven effective for fall prediction. Building on this foundation, the proposed study utilizes YOLO-based pose estimation with the CAUCAFall dataset. Keypoints extracted from images are processed using a sliding window strategy for training RNN models, including RNN, LSTM, and GRU. The model evaluation highlights the effectiveness of RNN-based approaches in enhancing fall motion prediction.

PRELIMINARIES

Dataset

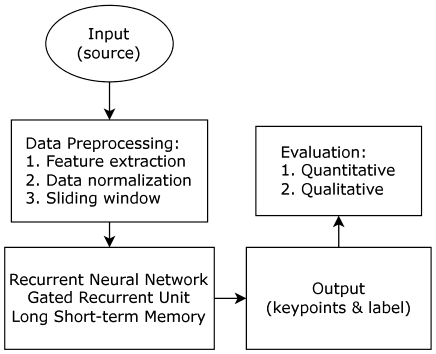
The CAUCAFall dataset is designed explicitly for recognizing human falls, with its distinguishing feature being its creation under uncontrolled home environment conditions. This dataset includes various challenging scenarios such as occlusions, lighting changes, diverse participant clothing, background movement, differing floor and room textures, multiple fall angles, varying distances between the camera and the fall, and participants of different ages, weights, heights, and dominant legs. These factors play a significant role in accurately recognizing provided segmentation labels for each frame, enabling the implementation of human fall recognition methods. CAUCAFall includes five types of falls—forward falls, backward falls, falls to the left, falls to the right, and falls arising from sitting. Additionally, the dataset was chosen for this research as it uniquely ally; it incorporates five types of activities of daily living (ADLs): walking, hopping, picking up an object, sitting, and kneeling. Frames were labelled as “fall” or “1” for fall-related activities and “no-fall” (or “0”) for ADLs. The segmentation labels distinguishing “fall” and “no-fall” were manually created using a text editor (Guerrero et al., 2022).

Pose Estimation: YOLOv8 Pose

YOLO is a state-of-the-art object detection model. In general, the YOLO model was built to detect common objects like furniture, animals, and buildings since the model improved the performance of each version. YOLO can detect the joints in the human body pose or is mainly known for pose estimation. Our work used revised versions of previously developed models which aim at pose estimation, being specific, this study utilized YOLOv8 pose as it`s a redesigned, faster detecting model with smaller layers compared to other models of its kind (Gu & Su, 2024).

PROPOSED METHOD

In this section, we describe the method used in this research. We propose predicting human motion to determine if a fall will occur in the future, with both quantitative and qualitative evaluations. Gambar 1 illustrates the overall general flow.



Gambar 1. Overview system flow.

Data Preprocessing

The YOLOv8 pose estimation model helps in future extraction through a sequence of images by capturing the human body pose. The explained body pose is normalized in the scale of 0 and 1, and then a sliding window is used to make a subset (T_P) from the entire dataset (T) in length. The details of each step in data preprocessing will be explained in points below 1, 2, and 3, respectively.

Pose Estimation by YOLOv8 Pose

In this study, we have employed the advanced version of YOLO version 8, which is called YOLOv8 pose. YOLOv8 pose can detect several persons within the image, estimating 17 joints for each individual.

Data Normalization

Normalization is performed to reduce the scale of the input and the output within the implemented boundaries of the model such that the prediction is not over these boundaries. Given the YOLOv8 detected image shape of (448×640), which has a maximum range of (448) called width or (640) called height, normalization is also made concerning the shape dimensions of the data as stated in Eq.(1).

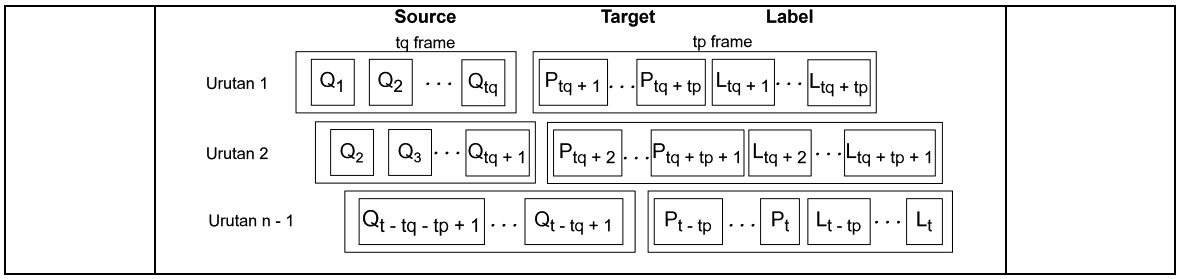
$$x_i = \frac{x_i - \min(\text{dimension})}{\max(\text{dimension}) - \min(\text{dimension})}, \quad (1)$$

where the x_i represents the data with indexing of i-th order in the sequential form. In this case, the dimension of dim is transformed to the cartesian coordinate form. When the joint's position coordinates are x_i coordinate of the joints, and the maximum dimension x is 448, the maximum y coordinate is 640. Dimension, on the other hand, has a lower limit of 0. With

Eq.(1), the data that has the upper limits of 448 and 640 can be represented in a range from 0 to 1.

Sliding Window

The frames in the dataset consist of a considerable amount of $F_i \in \mathbb{R}^{2K}$, which are long sets of the x and y coordinates of the K joints. A sliding window technique is often used to extract a smaller portion from the sequence of frames. This work used the sliding windows technique with source and target labelling schemes for the prediction task. As it was stated by Gambar 2, the scheme of sliding windows is visualized in our research.



Gambar 2. The sliding window scheme involves dividing the input data (source) and the expected model output (target) into overlapping segments or windows. Each row corresponds to a window consisting of a source sequence of length q (query) and a prediction sequence of length p. The subsequent window is created by shifting the start of the previous window by a predefined step size. This process continues until the final subset of the data, with the total length denoted as t. Additionally, the sliding window for the label L is appended to the target for alignment with the model output.

In this manner, the data is divided into three sections: source, target, and label, using a sliding window technique. All three of them are included in one window. Hence, these three data parts are utilized for different scenarios. The input to the model will be called the source, while the model's expected output will be both target and label.

Model

The prediction model uses Recurrent Neural Network (RNN) architectures, particularly LSTM and GRU variants, which are suitable for sequential data processing. The sliding window technique is applied to segment the data for input into the model.

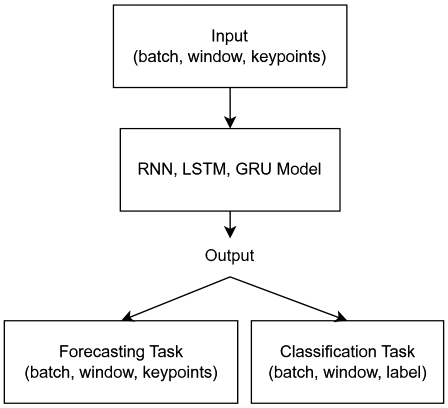
RNNs are specialized neural networks that excel in processing sequential data by utilizing their internal memory to handle inputs of varying lengths. This capability makes

them particularly suitable for tasks like time-series prediction and human motion forecasting (Huang & Li, 2023; Ruiz et al., n.d.; Yunus et al., 2023). However, traditional RNNs often encounter challenges with vanishing and exploding gradients during training, which can impede their performance on longer sequences.

Long Short-Term Memory networks (LSTMs) were introduced to address these limitations, featuring memory cells and gating mechanisms that enhance their ability to learn long-term dependencies (Mankodiya et al., 2022; Usmani et al., 2021). These components allow LSTMs to retain information over extended periods and selectively forget irrelevant data, enhancing their ability to learn from sequential inputs. This makes LSTMs particularly effective for tasks that require understanding long-range dependencies.

Similarly, GRUs offer a streamlined alternative to LSTMs by combining the forget and input gates into a single update gate. This simplification reduces computational complexity while maintaining the ability to capture temporal dependencies, resulting in faster training times (Mankodiya et al., 2022; Usmani et al., 2021). GRUs have been effectively applied in various sequential data tasks, including human motion prediction and fall detection using wearable devices (Lau et al., 2022; Mankodiya et al., 2022).

We utilize the sliding window technique to facilitate the input into our models, which segments the data into overlapping frames. This method allows us to create multiple input sequences from the original dataset, effectively expanding the training set and enabling the model to learn from various temporal contexts. For our specific application, the models are designed to predict the successive ten frames, structured as (batch, windows, and 1D tensors of size 34 for forecasting and 1 for classification). To address this need, the classification output of the model using the activation function sigmoid, as stated by Gambar 3 the expected output model is visualized in our research. This configuration allows for a comprehensive representation of the sequential data, thereby enhancing the predictive capabilities of the RNN, LSTM, and GRU architectures.



Gambar 3. Forecasting and classification task expected output from the models

Loss Function

The prediction result from the model training is computed by comparing the distance of prediction and ground truth. The prediction result is separated into two parts: forecasting and classification. We calculate the loss function for the forecasting task with RMSE and classification task with L1 norm as described in Equations (2) and (3).

$$\mathcal{L}_1 = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(P_i - \hat{P}_i)^2}, \tag{2}$$

$$\mathcal{L}_2 = \frac{1}{N}\sum_{i=1}^{N}\left|L_i - \hat{L}_i\right|, \tag{3}$$

where P_i represents the ground truth forecasting data in the index (i-th), ($\hat{P}_i$) is the prediction result for the forecasting task, N is the total number of the data, (L_i) is the ground truth label and ($\hat{L}_i$) is the prediction result for falling or not falling down classification task.

$$\mathcal{L}_{Total} = \mathcal{L}_1 + \mathcal{L}_2, \tag{4}$$

Then, this loss is combined as described in Eq. (4) for the backpropagation loss score in the next iteration. Where $\mathcal{L}_{Total}$ is the sum of loss RMSE and loss L1 norm.

Evaluation Metrics

We calculate the evaluation metrics for the forecasting task with MPJPAE, MPJVE, classification task with accuracy, and Euclidean distance in 2D as described in Equations (6)

to (10). The same evaluation metrics are in use in research of human motion and human fall motion (Ruiz et al., n.d.; Yunus et al., 2023).

Accuracy is calculated by computing the sum of label ground truth equal to label prediction. The evaluation metric of Accuracy in a label is defined in Eq. (5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}, \tag{5}$$

Tabel 1. Confusion Matrix

	Predicted Falls	**Predicted Non-Falls**
Actual Falls	True Positive (TP)	False Negative (FN)
Actual Non-Falls	False Positive (FP)	True Negative (TN)

where TP is the number of true positive samples, TN is the number of true negative samples, FN is the number of false negative samples, and FP is the number of false positive samples as shown in the Tabel 1.

MPJPAE is calculated by computing the squared Euclidean distance between the ground truth and the prediction concerning the treated joints. The evaluation metric of E_{MPJPAE} is defined as:

$$E_{\text{MPJPAE}} = \frac{1}{N}\sum_{T_p=1}^{N} \left| P_{T_p} - \widehat{\boldsymbol{P}}_{\boldsymbol{T_p}} \right|, \tag{6}$$

where N is the total number of frames, while the ground truth and the predicted position of the $\mathbf{T_P}$-th joint in the frame $\mathbf{T_P}$ are represented by $\boldsymbol{P}_{\boldsymbol{T_p}}$ and $\widehat{\boldsymbol{P}}_{\boldsymbol{T_p}}$, respectively. $(|\ldots|)$ is the Euclidian distance between the ground truth and predicted joint positions is a measurement of the error between them. Therefore, the mean of the Euclidian distance between the predicted joint positions and the corresponding ground truth joint positions for all frames is calculated. This provides an indication of the extent to which the model's predicted joint locations correspond to the actual ground truth.

The evaluation metric of (E_{MPJVE}) in the frame $(\mathbf{T_P})$ is defined by:

$$E_{\text{MPJVE}} = \frac{1}{N}\sum_{T_p=1}^{N} |\, V_{T_p} - \hat{V}_{T_p} |, \tag{7}$$

where $\boldsymbol{T_p}$ is the frame to start until N, which is the total of frames, while $(\boldsymbol{V_{T_p}})$ is the velocity of ground truth and the $\widehat{\boldsymbol{V}}_{\boldsymbol{T_p}}$ is the predicted velocity in the frame of $\boldsymbol{T_p}$. $\boldsymbol{V_{T_p}}$ and $\widehat{\boldsymbol{V}}_{\boldsymbol{T_p}}$ are defined by:

$$V_{T_p} = P_{T_{p+1}} - P_{T_p}, \tag{8}$$

$$\hat{V}_{T_p} = \hat{P}_{T_{p+1}} - \hat{P}_{T_p}, \tag{9}$$

where $\boldsymbol{V_{T_p}}$ and $\widehat{\boldsymbol{V}}_{\boldsymbol{T_p}}$ are ground truth and predicted velocity in the given frame $\boldsymbol{T_p}$ then calculated by the difference between the position of keypoints in the one-frame difference of the position keypoints between the predicted and ground truth. The evaluation metric of the predicted Euclidean distance is defined by Eq. (10)

$$D_k = \sqrt{\frac{1}{N \cdot J} \sum_{i=1}^{N} \sum_{j=1}^{J} \left(P_{i,j} - \widehat{\boldsymbol{P}}_{\boldsymbol{i,j}}\right)^2}, \tag{10}$$

where in this context, the term $\widehat{\boldsymbol{P}}_{\boldsymbol{i,j}}$ and $\boldsymbol{P_{i,j}}$ refers to the predicted and the true position of the i-th window and j each joint in the Euclidean space, respectively. The euclidean distance computes the difference between the squared ground truth and predicted positions of each joint j, summed over the entire data of N frames and J joint samples. The result is then normalized by the product of N and J, also the square root of this quantity is taken to obtain the final Euclidean distance measure D_k. This provides an evaluation of the overall accuracy of the predicted joint positions compared to the true positions in the Euclidean space.

Finally, qualitatively, we plot the frame and sequence skeleton data to visually compare the prediction with the ground truth.

EXPERIMENT AND RESULTS

Dataset

This study uses the CAUCAFall dataset, which contains annotated video frames specifically designed to capture daily living and fall-related activities. The dataset is processed to extract human keypoints for fall motion analysis using YOLOv8 pose model. The sample of the dataset is shown in Gambar 4.

Prosiding Seminar Nasional Universitas Ma Chung 2026
(Sistem Informasi dan Informatika)
pp. 492-510
E-ISSN: 2808-7771

Gambar 4. CAUCAFall dataset sample of falling down motion

Results

In this section, we provide both the quantitative and the qualitative evaluations. For evaluation, we first compare the positions from the dataset with pose estimates from YOLOv8 to show the model performance in unstable data environments. The model is trained solely on the captured position joint data derived from the dataset provided by YOLOv8, and the evaluation is based on this data alone.

Quantitative Evaluation

Tabel 2 presents the loss class, loss forecast, and accuracy on ground truth testing for the CAUCAFall dataset, comparing our results with the RNN, LSTM, and GRU methods in 400ms as short-term predictions. The GRU model demonstrates the lowest loss forecast value of 0.1655, indicating better performance in forecasting compared to the RNN and LSTM models, which have loss forecast values of 0.1812 and 0.184, respectively. However, all three models show an identical accuracy of 0.716. This suggests that while the GRU model has a slight edge in forecasting, the overall accuracy remains consistent across the models.

Tabel 2. The following categories are to be considered: loss class, loss forecast, and accuracy

Model	Loss Class	Loss Forecast	Accuracy
RNN	0.2241	0.1812	0.716
LSTM	0.2241	0.184	0.716
GRU	0.2241	**0.1655**	0.716

Tabel 3. Comparison of MPJPAE and MPJVE on RNN, GRU, and LSTM

Model	MPJPAE	MPJVE
RNN	41.872	**21.44**
LSTM	43.04	21.688
GRU	**32.921**	22.457

Tabel 3 shows the MPJPAE and MPJVE comparing our method's result with the RNN, LSTM, GRU in the short-term prediction task. The GRU model achieves the lowest MPJPAE value of 31.921, outperforming the RNN and LSTM models, which have MPJPAE values of 41.872 and 43.04, respectively. In terms of MPJVE, the RNN model shows the best performance with a value of 21.44, compared to the LSTM and GRU models, which have MPJVE values of 21.68 and 22.45, respectively. This indicated that the GRU model is more effective in minimizing joint position error, while the RNN model excels in reducing joint velocity error.

Furthermore, in Tabel 4, we considered the distance value of model on average compare each joint key points in the testing data short-term prediction 500ms. In Tabel 4, the GRU model generally shows the lowest distance values for most key points, indicating better performance in predicting joint positions with an average of 38.731 pixels. Based on the table, the GRU model has the lowest distance value for all key points in the example of the nose at 42.9617, the left shoulder at 38.5483, left knee and right knee at 31.7653, and 32.0717, respectively.

Tabel 4. The RNN, LSTM, and GRU euclidean distance 2D models were selected 10 windows each samples, presented with 17 keypoints at each distance obtained by the model

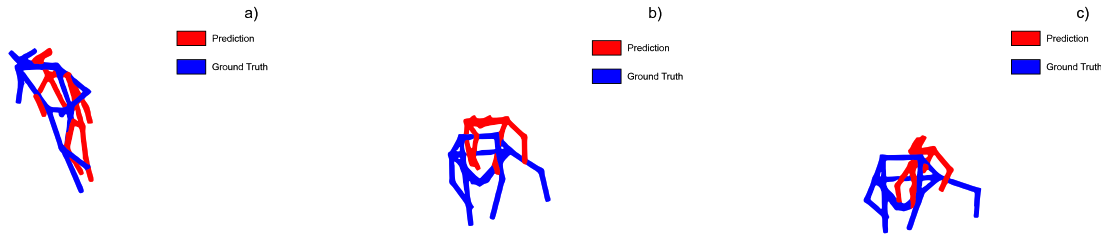
Keypoints	GRU	LSTM	RNN
Nose	**42.9617**	48.6890	47.0255
Left-eye	**43.1590**	49.3964	47.6454
Right-eye	**43.0783**	49.2418	48.7652
Left-ear	**41.4600**	49.3372	48.4919
Right-ear	**41.3155**	48.7127	49.0471
Left-shoulder	**38.5483**	50.8231	49.9608
Right-shoulder	**39.2697**	46.9218	49.2394
Left-elbow	**41.3961**	56.3931	54.8513

Right-elbow	**41.3178**	48.5589	50.5153
Left-wrist	**44.6747**	59.1749	59.4262
Right-wrist	**44.5421**	53.0201	55.2595
Left-hip	**32.8424**	44.3582	44.5212
Right-hip	**33.9649**	45.8150	45.5589
Left-knee	**31.7653**	50.3208	44.9196
Right-knee	**32.0717**	51.1771	46.0909
Left-ankle	**32.0562**	53.1223	47.0840
Right-ankle	**33.9978**	55.7452	49.0292
Average	**38.7310**	**50.6360**	**49.2610**

Overall, the GRU model demonstrates superior performance in most metrics, particularly in minimizing joint position and velocity errors, as well as accurate keypoints localization. Moreover, the RNN and LSTM models show strengths in specific areas, highlighting the importance of model selection based on the specific requirements of the task.

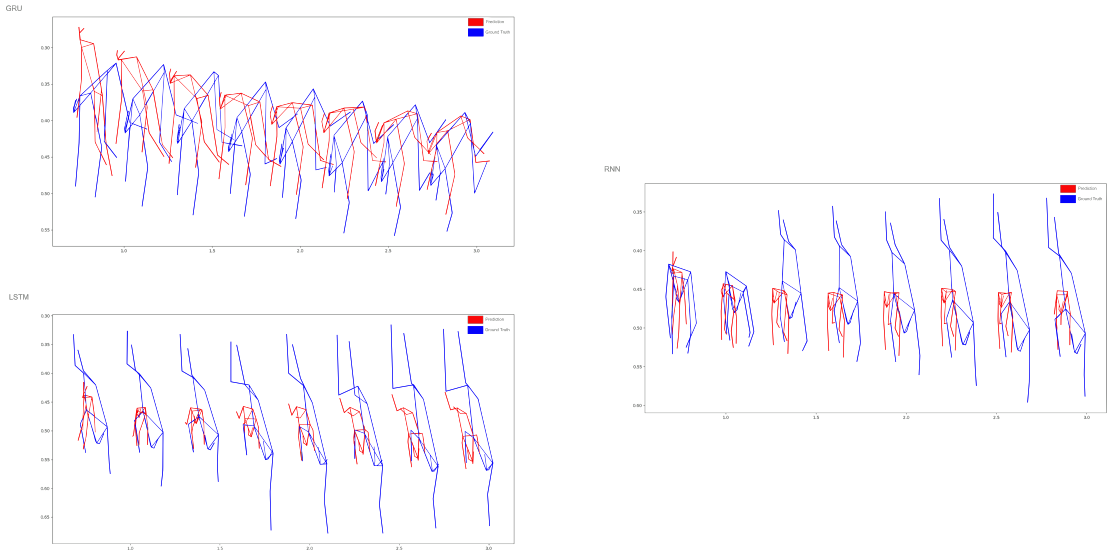
Qualitative Evaluation

In Gambar 5, a) on the left illustrates the RNN prediction, b) in the middle image illustrates the GRU, c) in the right image illustrates the LSTM prediction. The image is taken randomly in the visualization results between the frames of daily living activities and fall-related activities and it can be seen that the GRU frame has keypoints that are not so far from the ground truth and it looks like GRU can well predict keypoints.



Gambar 5. The qualitative frame-by-frame retrieval process utilizes a), b), c), respectively, with the implementation of RNN, GRU, and LSTM. As illustrated, the red keypoints represent the predicted values, while the blue keypoints denote the ground truth values.

In Gambar 6 the sequence result image, a qualitative evaluation experiment is also randomly conducted. This experiment involves a comparison between the upper left side of the image, which is a GRU image, the lower left side of the image, which is an LSTM image, and the left side of the image, which is an RNN image. The figure illustrates that when the distance between red and blue keypoints is minimal, it suggests that the model possesses the capability to predict human motion in activities of daily living or in contexts related to falls.



Gambar 6. The following investigation utilizes recurrent neural networks (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) to sequence keypoints qualitatively.

The red human skeleton image keypoints represent the model prediction results, while the blue human skeleton image keypoints correspond to the ground truth keypoints.

CONCLUSION AND DISCUSSION

This research successfully developed an RNN-based model for predicting human fall motion. Our quantitative evaluation shows that the GRU model generally outperforms the RNN and LSTM models in terms of loss forecast, MPJPAE, MPJVE, and spatial euclidean distance metrics. Our qualitative shows that GRU is the closest to the ground truth, but the RNN and LSTM have pose boundaries that follow the ground truth. The GRU model's ability to minimize joint position and velocity errors, as well as its accurate key point localization, makes it a strong candidate for human motion in activity daily living and human fall motion forecasting tasks. However, the RNN and LSTM models additionally have their strengths,

and the choice of model should be based on the specific needs and constraints of the application. Our work provides a comprehensive baseline for 2D human fall motion forecasting using the CAUCAFall datasets with MPJPAE, MPJVE, and Euclidean Distance 2D as the comparison metrics. Future work could improve the model's robustness, and extend its application to real-time fall prediction systems, and it can be posited that this can be regarded as a baseline model for future fall motion prediction, particularly in the case of GRU models.

REFERENCES

Christiansen, T. L., Lipsitz, S., Scanlan, M., Yu, S. P., Lindros, M. E., Leung, W. Y., Adelman, J., Bates, D. W., & Dykes, P. C. (2020). Patient Activation Related to Fall Prevention: A Multisite Study. *The Joint Commission Journal on Quality and Patient Safety*, *46*(3), 129–135. https://doi.org/10.1016/j.jcjq.2019.11.010

Farsi, M. (2021). Application of ensemble RNN deep neural network to the fall detection through IoT environment. *Alexandria Engineering Journal*, *60*(1), 199–211. https://doi.org/10.1016/j.aej.2020.06.056

Gu, K., & Su, B. (2024). A study of human pose estimation in low-light environments using YOLOv8 model. *Applied and Computational Engineering*, *32*(1), 136–142. https://doi.org/10.54254/2755-2721/32/20230200

Guerrero, J. C. E., España, E. M., Añasco, M. M., & Lopera, J. E. P. (2022). Dataset for human fall recognition in an uncontrolled environment. *Data in Brief*, *45*, 108610. https://doi.org/10.1016/j.dib.2022.108610

Hasan, M. M., Islam, M. S., & Abdullah, S. (2019). Robust Pose-Based Human Fall Detection Using Recurrent Neural Network. *2019 IEEE International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)*, 48–51. https://doi.org/10.1109/RAAICON48939.2019.23

Hua, M., & Lian, S. (n.d.). *Falls Prediction Based on Body Keypoints and Seq2Seq Architecture Yibing Nan CloudMinds Technologies*.

Huang, B., & Li, X. (2023). Human Motion Prediction via Dual-Attention and Multi-Granularity Temporal Convolutional Networks. *Sensors*, *23*(12), 5653. https://doi.org/10.3390/s23125653

Lau, X. L., Connie, T., Goh, M. K. O., & Lau, S. H. (2022). Fall Detection and Motion Analysis Using Visual Approaches. *International Journal of Technology*, *13*(6), 1173. https://doi.org/10.14716/ijtech.v13i6.5840

Li, M., Chen, S., Zhao, Y., Zhang, Y., Wang, Y., & Tian, Q. (n.d.). *Dynamic Multiscale Graph Neural Networks for 3D Skeleton-Based Human Motion Prediction*. Retrieved http://mocap.cs.cmu.edu/

Lin, C.-B., Dong, Z., Kuan, W.-K., & Huang, Y.-F. (2020). A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. *Applied Sciences*, *11*(1), 329. https://doi.org/10.3390/app11010329

Mankodiya, H., Jadav, D., Gupta, R., Tanwar, S., Alharbi, A., Tolba, A., Neagu, B.-C., & Raboaca, M. S. (2022). XAI-Fall: Explainable AI for Fall Detection on Wearable Devices Using Sequence Models and XAI Techniques. *Mathematics*, *10*(12), 1990. https://doi.org/10.3390/math10121990

Ozcan, A., Donat, H., Gelecek, N., Ozdirenc, M., & Karadibak, D. (2005). The relationship between risk factors for falling and the quality of life in older adults. *BMC Public Health*, *5*(1), 90. https://doi.org/10.1186/1471-2458-5-90

Ramirez, H., Velastin, S. A., Meza, I., Fabregas, E., Makris, D., & Farias, G. (2021). Fall Detection and Activity Recognition Using Human Skeleton Features. *IEEE Access*, *9*, 33532–33542. https://doi.org/10.1109/ACCESS.2021.3061626

Ruiz, A. H., Gall, J., & Moreno-Noguer, F. (n.d.). *Human Motion Prediction via Spatio-Temporal Inpainting*.

Saurav, S., Saini, R., & Singh, S. (2022). A dual-stream fused neural network for fall detection in multi-camera and $$360^{\circ }$$ videos. *Neural Computing and Applications*, *34*(2), 1455–1482. https://doi.org/10.1007/s00521-021-06495-5

Shu, X., Zhang, L., Qi, G.-J., Liu, W., & Tang, J. (2022). Spatiotemporal Co-Attention Recurrent Neural Networks for Human-Skeleton Motion Prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(6), 3300–3315. https://doi.org/10.1109/TPAMI.2021.3050918

Tamura, T., Yoshimura, T., & Sekine, M. (2007). A preliminary study to demonstrate the use of an air bag device to prevent fall-related injuries. *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3833–3835. https://doi.org/10.1109/IEMBS.2007.4353168

Usmani, S., Saboor, A., Haris, M., Khan, M. A., & Park, H. (2021). Latest Research Trends in Fall Detection and Prevention Using Machine Learning: A Systematic Review. *Sensors*, *21*(15), 5134. https://doi.org/10.3390/s21155134

Wang, B.-H., Yu, J., Wang, K., Bao, X.-Y., & Mao, K.-M. (2020). Fall Detection Based on Dual-Channel Feature Integration. *IEEE Access*, *8*, 103443–103453. https://doi.org/10.1109/ACCESS.2020.2999503

WHO. (2021, April 26). *Falls*. Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/Falls.

Youssfi Alaoui, A., Tabii, Y., Oulad Haj Thami, R., Daoudi, M., Berretti, S., & Pala, P. (2021). Fall Detection of Elderly People Using the Manifold of Positive Semidefinite Matrices. *Journal of Imaging*, *7*(7), 109. https://doi.org/10.3390/jimaging7070109

Yunus, A. P., Morita, K., Shirai, N. C., & Wakabayashi, T. (2023). Time Series Self-Attention Approach for Human Motion Forecasting: A Baseline 2D Pose Forecasting.

Journal of Advanced Computational Intelligence and Intelligent Informatics, *27*(3), 445–457. https://doi.org/10.20965/jaciii.2023.p0445

Yunus, A. P., Shirai, N. C., Morita, K., & Wakabayashi, T. (2022). Comparison of RNN-LSTM and Kalman Filter Based Time Series Human Motion Prediction. *Journal of Physics: Conference Series*, *2319*(1), 012034. https://doi.org/10.1088/1742-6596/2319/1/012034

Zhao, M., Tang, H., Xie, P., Dai, S., Sebe, N., & Wang, W. (2023). Bidirectional Transformer GAN for Long-term Human Motion Prediction. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *19*(5), 1–19. https://doi.org/10.1145/3579359