

Accessible Input on Small Touchscreens: Evaluating Deep Learning Architectures for Stroke and Motion Gestures

Laura-Bianca Bilius

MintViz Lab, MANSiD Center, Ștefan cel Mare University of Suceava
Suceava, Romania
laura.bilius@usm.ro

ABSTRACT

People with upper-body motor impairments often experience difficulties when interacting with small touchscreens and digital devices. To address this challenge, we present a comparative evaluation of representative deep learning architectures for recognizing touchscreen and mid-air motion input, using an open dataset with gestures collected from people with upper-body motor impairments. Across the evaluated architectures, the 1D-CNN achieved the highest recognition performance, with an accuracy of 94.7% for stroke gestures and 84.0% for motion gestures. Our analysis further revealed that gesture recognition performance, represented by accuracy rates, cannot be fully explained by participants' health conditions or generic WHODAS health and disability scores, suggesting the need for ability-based design approaches.

AUTHOR KEYWORDS

Gesture recognition; stroke gestures; motion gestures; upper-body motor impairments; deep learning.

DOI: 10.37789/icusi.2026.26

INTRODUCTION

An estimated 1.3 billion people worldwide experience significant disabilities, representing approximately 16% of the global population.¹ Consequently, the design of accessible technologies must therefore account for the diverse abilities and needs of users. Despite rapid technological advances, however, many systems and applications remain difficult for people with motor impairments to use [27,29]. Touchscreen stroke and mid-air motion gestures have become a fundamental interaction technique in user interfaces, enabling users to navigate interfaces, manipulate digital content, and invoke commands through simple hand movements [24,25]. Because the articulation of stroke and motion gestures depends on users' individual motor abilities, differences in dexterity, coordination, muscle strength, tremor, and fatigue can alter gesture trajectories, affecting both usability and gesture recognition accuracy [14,24,27,29]. Therefore, designing gesture-based interfaces that accommodate this variability remains a challenge for accessible gesture interaction. Despite significant advances in gesture recognition, there is still limited understanding of how people with motor impairments perform hand gestures [12,14,24,25] and how recognition systems can better support the diversity of their movement patterns.

To address this gap, we present a comparative evaluation of ten representative deep learning architectures for recognizing touchscreen stroke gestures and mid-air motion gestures using gesture data collected from people with upper-body motor impairments. We evaluate the deep learning architectures

under a user-dependent training approach and analyze gesture recognition performance across participants with diverse upper-body motor impairments from a public dataset [25]. The main contributions of this paper are as follows:

- (1) A comparative evaluation of ten representative deep learning architectures for recognizing touchscreen and mid-air gestures performed by users with upper-body motor impairments using a small-screen digital device.
- (2) An analysis of recognition performance across participants with various health conditions and disability scores.

RELATED WORK

Upper-body motor impairments can affect movement speed, muscle strength, coordination, range of motion, and movement precision, which can significantly influence the articulation of both stroke and motion gestures [12,25]. In this direction, several studies have investigated the challenges and opportunities of gesture-based interaction for people with upper-body motor impairments. For example, Malu *et al.* [12] examined smartwatch interaction and found that taps, swipes, and gestures performed near the display edges were particularly difficult to perform. Similarly, Naftali and Findlater [14] explored everyday smartphone use through surveys, interviews, diary studies, and observations. Their findings showed that smartphones support many daily activities and increase independence, while touchscreen interactions remain challenging in real-world situations. Vataavu and Ungurean [25] compared the articulation of stroke and mid-air motion gestures performed by people with upper-body motor impairments using wearables worn on the wrist, finger, and head. Their findings showed that motion gestures were less affected by motor impairments than stroke gestures, which indicates that motion gestures may provide a more accessible interaction technique for users. Research on personalized gesture interaction revealed that there is rarely a single gesture vocabulary that is equally accessible to all users. In this context, Yamagami *et al.* [29] showed that user-defined gesture sets articulated by people with upper-body motor impairments are highly ability-dependent. As a resource for accessible gesture interaction research, Gesture-A11Y [20] provides a public dataset of stroke and motion gestures performed by people with upper-body motor impairments.

Recent work has reported on the importance of user-specific data for modeling touch input performance, showing that personalized models can be obtained from relatively small amounts of training data, with performance reaching a plateau after approximately 8 to 24 touch observations [23]. Such findings motivate user-dependent modeling approaches that account for individual differences in motor abilities when designing accessible interaction techniques. From the gesture

¹ <https://www.who.int/news-room/fact-sheets/detail/disability-and-health>

recognition perspective, Vatavu and Ungurean [24] evaluated 5-family recognizers on touchscreen stroke gestures performed by people with upper-body motor impairments using a tablet. Although recognition accuracy remained relatively high (93.0%), gestures articulated by participants with motor impairments were recognized less accurately than those performed by participants without impairments.

In conclusion, prior work shows the need for gesture interfaces that adapt to users' specific abilities rather than requiring users to conform to predefined gesture models. To the best of our knowledge, no previous study has evaluated deep learning architectures for recognizing touchscreen stroke and mid-air motion gestures performed by people with upper-body motor impairments on small-screen devices.

EXPERIMENT

Dataset

For our analysis, we used the public dataset² introduced by Vatavu and Ungurean [25]. That study investigated gestures for smartwatches, smartglasses, and smartrings using both touchscreen stroke and mid-air motion input. The dataset comprises recordings from 28 participants, including 14 people with upper-body motor impairments and 14 participants without impairments serving as a control group. The health conditions and WHODAS [28] scores of the participants with upper-body motor impairments are summarized in Table 1. The dataset² includes 7,290 touchscreen stroke gesture samples and 3,809 mid-air motion gesture samples, both collected from the same 28 participants. Motion gestures were recorded using the smartwatch's built-in three-axis accelerometer,³ comprising 18 gesture types articulated at the arm, finger, and head levels. Stroke gestures were recorded by drawing directly on the smartwatch touchscreen and comprised 12 gesture types articulated at the same three levels. Each participant performed every gesture eight times.

Our analysis considers only the smartwatch data collected from people with upper-body motor impairments, including both touchscreen stroke and mid-air motion gestures. We focused on smartwatch-based interaction because smartwatches have become one of the most widely adopted wearable devices,⁴ due to their health monitoring, communication, and fitness capabilities. Their small touchscreens, however, make conventional touch interaction particularly challenging for people with upper-body motor impairments, compared to larger devices, such as smartphones [23].

It is important to note that not all participants completed both smartwatch gesture categories. Specifically, one participant (P_2) did not perform mid-air motion gestures, while one different participant (P_9) did not perform touchscreen stroke gestures. Participant P_6 was excluded from our analysis because he did not perform the smartwatch gestures at all, although this participant completed the smartring and smartglasses tasks in the Vatavu and Ungurean [25] study. As a result, our analysis was conducted on data collected from 13 different participants with upper-body motor impairments using mid-air motion (573 gestures) and touchscreen stroke gestures (1,150 gestures), a total of 1,723 gestures.

The dataset used in this work includes six mid-air motion gestures (*e.g.*, tapping, shaking, and directional movements performed at two different speeds) and 12 touchscreen stroke gestures, including geometric shapes (*e.g.*, circle, square, and six-point star), symbols (*e.g.*, asterisk, check mark, and heart), letters (A, M, and X), and directional strokes (left and right). A detailed description of the gesture vocabulary is provided by Vatavu and Ungurean [25].

Design and Procedure

To analyze the performance of deep learning models in recognizing touchscreen stroke and mid-air motion gestures, we compared ten neural network architectures, as follows:

- A₁ RNN: Recurrent Neural Networks are a standard recurrent architecture that models sequential data by propagating hidden states across time steps [4];
- A₂ LSTM: Long Short-Term Memory is a RNN architecture designed to capture long-term temporal dependencies in sequential data through memory cells and gating mechanisms [9];
- A₃ GRU: Gated Recurrent Units are a lightweight recurrent architecture that employs update and reset gates to efficiently model temporal dependencies while reducing computational complexity [15];
- A₄ 1D-CNN: One-dimensional Convolutional Neural Networks are a convolution-based architecture that extracts local temporal patterns directly from one-dimensional signals using convolutional filters [3];
- A₅ 2D-CNN: Two-dimensional Convolutional Neural Networks are a convolution-based architecture that learns spatial feature representations from two-dimensional transformations of gesture signals [31];
- A₆ MLP: Multi-Layer Perceptrons are a fully connected feed-forward neural network architecture that performs classification using nonlinear feature transformations [32];
- A₇ AE: Autoencoders are an unsupervised neural network architecture that learns compact latent representations by reconstructing the input data through an encoder-decoder architecture [6];
- A₈ TF: Transformer-based models are an attention-based architecture that captures long-range dependencies in sequential data using self-attention mechanisms instead of recurrence [22];
- A₉ GNN: Graph Neural Networks are designed to capture relationships in graph-structured data by propagating information between connected nodes [19];
- A₁₀ GAN: Generative Adversarial Networks are designed to learn robust feature representations through adversarial training [8].

The A₁–A₁₀ architectures were selected to represent the main neural network paradigms [10] and were chosen based on their previous use in gesture and human movement recognition. The considered architectures range from classical feed-forward models, such as MLP [18], to convolutional and recurrent architectures designed to capture spatial and temporal movement patterns, including 1D-CNN [7], 2D-CNN [21], RNN [5], LSTM [17], and GRU [30]. The comparison also covers architectures based on attention, reconstruction, graph modeling, and adversarial learning, such as Transformers [16], Autoencoders [2], GNNs [13], and GANs [11].

²<https://raduvatavu.usv.ro/projects/MI-GESTURES-WGR-2022>

³<https://www.samsung.com/ro/support/model/SM-R365NZRNR0M/>

⁴<https://www.demandsage.com/smartwatch-statistics/>

All models, $A_1 - A_{10}$ were implemented using the TensorFlow⁵ and Keras⁶ deep learning frameworks. Prior to training, all gesture sequences were resampled to a fixed length of 64 samples, and smoothed using a moving average filter [1], ensuring a consistent input representation across all architectures. Each mid-air motion gesture sample was represented as a $T \times 3$ matrix, where $T \in \mathbb{R}^{64}$ denotes the accelerometer samples and the three columns correspond to acceleration values along the x , y , and z axes. At the dataset level, the input was organized as a three-dimensional tensor of shape $N \times T \times 3$, where N denotes the number of gesture samples per participant. Similarly, each stroke gesture sample was represented as a $T \times 2$ matrix, where $T = 64$ denotes the number of touchscreen samples and the two columns correspond to the x and y coordinates of the touch trajectory. At the dataset level, the stroke gesture input was organized as a two-dimensional tensor of shape $N \times T \times 2$.

All models were evaluated under a user-dependent training approach, with stratified shuffle-split cross-validation using five splits to determine the best hyperparameters for each participant, algorithm and gesture category (*i.e.*, motion and stroke gestures). Hyperparameter optimization was performed with Optuna⁷ using Bayesian optimization [26]. The search space included the number of layers (one to five), dropout rate (0.0, 0.25 and 0.5), learning rate (1e-2, 1e-3, 1e-4, 1e-5), optimizer (adam and rmsprop), and batch size (4, 8, 16, 32 and 4), together with architecture-specific hyperparameters such as the number of hidden units or filters, kernel size, pooling strategy, window size and so on. Overall, the hyperparameter search space comprised approximately 50,000 possible configurations for each of the $A_1 - A_{10}$ models. As an example, Code 1 presents the hyperparameter search space defined for the 1D-CNN architecture. Each model was trained for a maximum of 100 epochs, with early stopping based on validation performance. The performance of each hyperparameter configuration was measured as the average of recognition accuracy across the five validation splits. Hyperparameter optimization was performed for each participant, model, and gesture type, resulting in participant- and gesture-specific hyperparameter configurations for each model. The following section presents the results obtained for each participant and deep learning model using the best-performing hyperparameter configuration.

Code 1: Hyperparameter search space for 1D-CNN.

```
hyperparams = {
  "1D-CNN": {
    "num_layers": [1, 2, 3, 4, 5],
    "batch_size": [4, 8, 16, 32, 64],
    "optimizer": ["adam", "rmsprop"],
    "filters": [32, 64, 128],
    "kernel_size": [3, 5, 7, 9],
    "stride": [1, 2],
    "activation": ["relu"],
    "pooling": ["max", "average"],
    "pool_size": [2, 3],
    "dropout": [0.0, 0.25, 0.5],
    "learning_rate": [1e-2, 1e-3, 1e-4, 1e-5]
  }
}
```

⁵<https://www.tensorflow.org/>

⁶<https://keras.io/>

⁷<https://optuna.org/>

RESULTS AND DISCUSSION

Table 1 summarizes the overall recognition accuracy obtained by the evaluated deep learning architectures under the user-dependent training approach, using the best-performing hyperparameter configuration identified during hyperparameter optimization for each participant and gesture category. Overall, 1D-CNN achieved the highest average recognition accuracy, reaching 84.0% for motion gestures and 94.7% for stroke gestures. These results suggest that local temporal patterns captured by convolutional filters are sufficient to discriminate between both stroke and motion gestures. Among the remaining architectures, 2D-CNN, and MLP also achieved competitive performance, particularly for stroke gesture recognition, with average accuracies of 94.4% and 94.1%, respectively. For motion gestures, LSTM yielded the highest results after 1D-CNN, achieving an average recognition accuracy of 80.0%. In contrast, AE (64.2%) and GAN (69.1%) achieved the lowest average accuracies for motion gesture recognition, whereas GNN (84.1%) and RNN (85.9%) showed the lowest performance on stroke gestures. Overall, stroke gestures were recognized more accurately than motion gestures across all evaluated architectures, whereas recognition performance for motion gestures achieved greater variability across participants, making the recognition task more challenging.

An interesting observation concerns the diversity in recognition performance across participants with different upper-body motor impairments. Differences were observed between participants diagnosed with the same health condition. Participants with the same health condition generally achieved comparable recognition performance of motion gestures, with differences remaining below five percentage points across 1D-CNN architecture. The largest difference was observed for participants with Multiple Sclerosis (P_5 and P_7) and Spinal Cord Injury C4-C5 (P_1 and P_8) during the motion gesture task (80.0% and 90.0%, and 78.0% and 70.0%, respectively). A similar trend was observed for stroke gestures. The largest difference was observed for participants with Spinal Cord Injury C4-C5 level (P_1 and P_8), who achieved recognition accuracies of 100.0% and 84.0%, respectively. In contrast, participants with Multiple Sclerosis (P_5 and P_7), and Spastic Quadriplegia (P_{11} and P_{12}) achieved only minor differences in recognition performance.

Moreover, participants with similar WHODAS scores achieved substantially different recognition accuracies. For example, participants P_8 , P_9 , and P_{11} all reported a WHODAS score of 37.5, yet their motion gesture recognition accuracy using the 1D-CNN architecture ranged from 70.0% to 100.0%. Participants with different WHODAS scores sometimes achieved comparable recognition performance. For instance, participant P_1 (WHODAS = 52.1) and participant P_{13} (WHODAS = 12.5) both achieved 100.0% recognition accuracy for stroke gestures. Table 1 provides the complete participant-level recognition results. No statistically significant correlations were observed between WHODAS scores and 1D-CNN performance for either stroke gestures ($\rho(12) = -.244, p = .445$) or motion gestures ($\rho(12) = .155, p = .630$).

These observations reveal that neither health condition nor the overall health and disability score explains gesture recognition performance in our study. Instead, future work should implement ability-based design principles [27], where interaction techniques and user interfaces are adapted to users'

Table 1: Recognition accuracy rates obtained by the evaluated deep learning architectures under user-dependent training.

Gesture Type	No.	CR [†]	WHODAS [‡]	LSTM	1D-CNN	2D-CNN	GRU	AE	TF	MLP	GNN	GAN	RNN
Motion	P ₁	100%	52.1	74.0%	78.0%	80.0%	74.0%	68.0%	84.0%	76.0%	82.2%	77.8%	77.8%
Motion	P ₃	100%	41.7	96.0%	100%	80.0%	90.0%	70.0%	94.0%	70.0%	86.7%	80.0%	88.9%
Motion	P ₄	95.8%	39.6	86.0%	94.0%	76.0%	78.0%	82.0%	80.0%	86.0%	80.0%	84.4%	86.7%
Motion	P ₅	97.9%	52.1	72.0%	80.0%	78.0%	80.0%	64.0%	68.0%	74.0%	73.3%	62.2%	77.8%
Motion	P ₇	100%	39.6	90.0%	90.0%	90.0%	84.0%	70.0%	90.0%	78.0%	77.8%	64.4%	75.6%
Motion	P ₈	100%	37.5	54.0%	70.0%	72.0%	64.0%	46.0%	66.0%	58.0%	66.7%	62.2%	73.3%
Motion	P ₉	100%	37.5	84.0%	82.0%	74.0%	70.0%	64.0%	78.0%	74.0%	64.4%	73.3%	71.1%
Motion	P ₁₀	100%	14.6	70.0%	70.0%	66.0%	66.0%	46.0%	68.0%	52.0%	60.0%	57.8%	60.0%
Motion	P ₁₁	100%	37.5	96.0%	100%	88.0%	92.0%	88.0%	100%	92.0%	95.6%	91.1%	88.9%
Motion	P ₁₂	100%	33.3	82.0%	72.0%	64.0%	74.0%	54.0%	56.0%	56.0%	57.8%	51.1%	57.8%
Motion	P ₁₃	100%	12.5	74.0%	84.0%	76.0%	62.0%	52.0%	66.0%	58.0%	57.6%	55.6%	71.1%
Motion	P ₁₄	100%	16.7	82.0%	88.0%	66.0%	76.0%	66.0%	64.0%	76.0%	62.2%	68.9%	77.8%
Overall Motion		99.5%	31.9	80.0%	84.0%	75.8%	75.8%	64.2%	76.2%	70.8%	73.5%	69.1%	75.6%
Stroke	P ₁	100%	52.1	96.0%	100.0%	99.0%	100.0%	98.0%	97.0%	100.0%	94.7%	96.8%	92.6%
Stroke	P ₂	100%	45.8	83.0%	82.0%	87.0%	86.0%	80.0%	88.0%	84.0%	88.4%	79.0%	79.0%
Stroke	P ₃	100%	41.7	96.0%	99.0%	98.0%	98.0%	95.0%	100.0%	97.0%	100.0%	94.7%	86.3%
Stroke	P ₄	100%	39.6	100.0%	100.0%	100.0%	99.0%	99.0%	100.0%	99.0%	90.5%	99.0%	97.9%
Stroke	P ₅	100%	52.1	90.0%	96.0%	94.0%	91.0%	90.0%	93.0%	94.0%	80.0%	93.7%	89.5%
Stroke	P ₇	100%	39.6	90.0%	95.0%	96.0%	90.0%	90.0%	93.0%	97.0%	72.6%	89.5%	76.8%
Stroke	P ₈	100%	37.5	61.0%	84.0%	82.0%	81.0%	72.0%	67.0%	84.0%	44.2%	85.3%	62.1%
Stroke	P ₁₀	99.0%	14.6	97.9%	99.0%	97.9%	96.8%	95.8%	99.0%	96.8%	89.5%	95.8%	96.8%
Stroke	P ₁₁	100%	37.5	93.0%	94.0%	94.0%	93.0%	90.0%	93.0%	93.0%	90.5%	93.7%	95.8%
Stroke	P ₁₂	100%	33.3	86.0%	92.0%	96.0%	87.0%	90.0%	92.0%	90.0%	79.0%	90.5%	79.0%
Stroke	P ₁₃	99.0%	12.5	90.5%	100.0%	96.8%	94.7%	93.7%	94.7%	96.8%	83.2%	93.7%	85.3%
Stroke	P ₁₄	100%	16.7	91.0%	96.0%	92.0%	94.0%	93.0%	90.0%	98.0%	96.8%	93.7%	89.5%
Overall Stroke		99.8%	32.5	89.5%	94.7%	94.4%	92.5%	90.5%	91.7%	94.1%	84.1%	92.1%	85.9%

[†]CR represent task completion rate achieved by participants during the experiment. [‡]WHODAS 2.0 health and disability score [28]. The participants presented a diverse range of health conditions, including Friedreich's ataxia (P₁₀), Multiple Sclerosis (P₅ and P₇), Osteogenesis Imperfecta (P₆), Parkinson's disease (P₉ and P₁₃), Phocomelia (P₁₄), Spinal Cord Injury C4-C5 (P₁ and P₈), Spinal Cord Injury T7 (P₃), Spina Bifida (P₄), Spastic quadriplegia (P₁₁ and P₁₂), and Traumatic Brain Injury (P₂).

specific motor abilities rather than their health condition or self-reported impairments to generic disability questionnaires.

CONCLUSION

This paper presented a comparative evaluation of ten deep learning architectures for recognizing touchscreen stroke and mid-air motion gestures collected from people with upper-body motor impairments. Among the evaluated models, the 1D-CNN architecture achieved the highest recognition accuracy for both gesture types. Our results also revealed large recognition variability, highlighting the importance of personalized recognition approaches, contributing to future work in designing more accessible gesture-based interaction techniques for people with upper-body motor impairments.

ACKNOWLEDGMENTS

This work was supported by a grant of the Ministry of Education and Research, CCCDI-UEFISCDI, project no. PN-IV-P7-7.1-PTE-2024-0434, within PNCDI IV.

References

- [1] Laura-Bianca Bilius, Ștefan Gheorghe Pentiuc, and Radu-Daniel Vatavu. 2023. TIGER: A Tucker-based instrument for gesture recognition with inertial sensors. *Pattern Recognition Letters* 165 (2023), 84–90. <https://doi.org/10.1016/j.patrec.2022.11.028>
- [2] Yidan Cao, Qingshan Wang, Qi Wang, and Peng Liu. 2023. Can Same-right-and-different-left Gestures Be Recognized with Only Right-hand Signals? *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 22, 9, Article 226 (Sept. 2023), 19 pages. <https://doi.org/10.1145/3617370>
- [3] Shumei Chen, Jianbo Yu, and Shijian Wang. 2022. One-dimensional convolutional neural network-based active feature extraction for fault detection and diagnosis of industrial processes and its understanding via visualization. *ISA Transactions* 122 (2022), 424–443. <https://doi.org/10.1016/j.isatra.2021.04.042>
- [4] Susmita Das, Amara Tariq, Thiago Santos, Sai Kantareddy, and Imon Banerjee. 2012. *Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and*

Introduction to Influential Research. Springer, 117–138. https://doi.org/10.1007/978-1-0716-3195-9_4

- [5] Rhett Dobinson, Marc Teyssier, Jürgen Steimle, and Bruno Fruchard. 2022. MicroPress: Detecting Pressure and Hover Distance in Thumb-to-Finger Interactions. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction (SUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 4, 10 pages. <https://doi.org/10.1145/3565970.3567698>
- [6] Ganggang Dong, Guisheng Liao, Hongwei Liu, and Gangyao Kuang. 2018. A Review of the Autoencoder and Its Variants: A Comparative Perspective from Target Recognition in Synthetic-Aperture Radar Images. *IEEE Geoscience and Remote Sensing Magazine* 6, 3 (2018), 44–68. <https://doi.org/10.1109/MGRS.2018.2853555>
- [7] Naveen Gehlot, Rajesh Kumar, Surender Hans, and Pilani Nkomozezi. 2025. 1D Convolutional Neural Architecture Search for sEMG Hand Gesture Recognition. *SN Comput. Sci.* 6, 7 (Aug. 2025), 19 pages. doi:10.1007/s42979-025-04324-3
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Communications of the ACM* 63, 11 (2020), 139–144. <https://doi.org/10.1145/3422622>
- [9] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
- [10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep Learning. *Nature* 521 (2015), 436–444. <https://doi.org/10.1038/nature14539>
- [11] Sijie Li, Lvqing Yang, Sien Chen, Jianwen Ding, Wensheng Dong, Bo Yu, Qingkai Wang, and Menghao Wang. 2023. ARFG: Attach-Free RFID Finger-Tracking with Few Samples Based on GAN. In *Advanced Intelligent Computing Technology and Applications: 19th International Conference, ICIC 2023 Proceedings, Part II*. Springer-Verlag, Berlin, Heidelberg, 773–784. https://doi.org/10.1007/978-981-99-4742-3_64
- [12] Meethu Malu, Pramod Chundury, and Leah Findlater. 2018. Exploring Accessible Smartwatch Interactions for People with Upper Body Motor Impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, USA, 1–12. <https://doi.org/10.1145/3173574.3174062>
- [13] Riktim Mondal, Debadyuti Mukherjee, Pawan Kumar Singh, Vikrant Bhatija, and Ram Sarkar. 2021. A New Framework for Smartphone Sensor-Based Human Activity Recognition Using Graph Neural Network. *IEEE Sensors Journal* 21, 10 (2021), 11461–11468. <https://doi.org/10.1109/JSEN.2020.3015726>
- [14] Maia Naftali and Leah Findlater. 2014. Accessibility in context: understanding the truly mobile experience of smartphone users with motor impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on*

- Computers & Accessibility (ASSETS '14)*. ACM, USA, 209–216. <https://doi.org/10.1145/2661334.2661372>
- [15] Fathi M. Salem. 2022. *Gated RNN: The Gated Recurrent Unit (GRU) RNN*. Springer, Cham, 85–100. https://doi.org/10.1007/978-3-030-89929-5_5
- [16] Su Yeon Shim, Younho Nam, Jinsung Yoon, Wooyeol Jeong, and Young-Joo Suh. 2026. AcouLetter: Enabling Fine-Grained Mid-Air Alphabetic Thumb Gesture Recognition on Smartphones with Acoustic and Inertial Sensing. In *Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI '26)*. Association for Computing Machinery, New York, NY, USA, 1772–1786. <https://doi.org/10.1145/3742413.3789093>
- [17] Leonardo Silva, Fabrizio Soares, Juliana Felix, Luciana Cardoso, Renan Vinicius Aranha, and Thamer Horbylon Nascimento. 2024. Gesture Recognition in Smartwatches Using LSTM for Interaction in Low-Cost Virtual Environments. In *Proceedings of the 26th Symposium on Virtual and Augmented Reality (SVR '24)*. Association for Computing Machinery, New York, NY, USA, 284–288. <https://doi.org/10.1145/3691573.3691615>
- [18] Gargi Singh, Saanika Gautam, Yugnanda Puri, Jolly Parikh, and Gargi Mishra. 2024. Novel System based on Amalgamation of Embedded systems and Edge AI for Hand Gesture Recognition: An Edge AI Solution for Hand Gesture Recognition. In *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing (IC3-2024)*. Association for Computing Machinery, New York, NY, USA, 432–439. <https://doi.org/10.1145/3675888.3676089>
- [19] James Tanis, Chris Giannella, Adrian Mariano, and Daoud Meerzaman. 2026. Introduction to Graph Neural Networks for Machine Learning Engineers. *Comput. Surveys* 58, 13, Article 340 (2026), 37 pages. <https://doi.org/10.1145/3816725>
- [20] Mihail Terenti, Laura-Bianca Bilius, Ovidiu-Ciprian Ungurean, and Radu-Daniel Vatavu. 2025. Empowering Accessible Gesture Input Design with Gesture-A11Y. In *Proceedings of the 22nd International Web for All Conference (W4A '25)*. ACM, NY, USA, 74–79. <https://doi.org/10.1145/3744257.3744267>
- [21] Imen Trabelsi, Jules Francoise, and Yacine Bellik. 2022. Sensor-based Activity Recognition using Deep Learning: A Comparative Study. In *Proceedings of the 8th International Conference on Movement and Computing (MOCO '22)*. Association for Computing Machinery, New York, NY, USA, Article 20, 8 pages. <https://doi.org/10.1145/3537972.3537996>
- [22] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (USA) (NIPS'17)*. Curran Associates Inc., NY, USA, 6000–6010. <https://dl.acm.org/doi/10.5555/3295222.3295349>
- [23] Radu-Daniel Vatavu, Irina Petrariu, Tudor Horomnea, and Ovidiu-Ciprian Ungurean. 2026. Modeling Touch Input for Users with Motor Impairments: Empirical Insights into Training Size Requirements. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. ACM, NY, USA, Article 473, 6 pages. <https://doi.org/10.1145/3772363.3798487>
- [24] Radu-Daniel Vatavu and Ovidiu-Ciprian Ungurean. 2019. Stroke-Gesture Input for People with Motor Impairments: Empirical Results & Research Roadmap. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, NY, USA, 14 pages. <https://doi.org/10.1145/3290605.3300445>
- [25] Radu-Daniel Vatavu and Ovidiu-Ciprian Ungurean. 2022. Understanding Gesture Input Articulation with Upper-Body Wearables for Users with Upper-Body Motor Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, NY, USA, Article 2, 16 pages. <https://doi.org/10.1145/3491102.3501964>
- [26] Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. 2023. Recent Advances in Bayesian Optimization. *ACM Comput. Surv.* 55, 13s, Article 287 (July 2023), 36 pages. <https://doi.org/10.1145/3582078>
- [27] Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-Based Design: Concept, Principles and Examples. *ACM Trans. Access. Comput.* 3, 3(9) (2011). <https://doi.org/10.1145/1952383.1952384>
- [28] World Health Organization. 2010. Measuring Health and Disability Manual for WHO Disability Assessment Schedule WHODAS 2.0. [https://www.who.int/publications/i/item/measuring-health-and-disability-manual-for-who-disability-assessment-schedule-\(whodas-2.0\)](https://www.who.int/publications/i/item/measuring-health-and-disability-manual-for-who-disability-assessment-schedule-(whodas-2.0))
- [29] Momona Yamagami, Alexandra A Portnova-Fahreeva, Junhan Kong, Jacob O. Wobbrock, and Jennifer Mankoff. 2023. How Do People with Limited Movement Personalize Upper-Body Gestures? Considerations for the Design of Personalized and Accessible Gesture Interfaces. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (NY, USA) (ASSETS '23)*. ACM, USA, Article 1, 15 pages. <https://doi.org/10.1145/3597638.3608430>
- [30] Huaigang Yang, Dong Zhang, Ping Xie, and Xiaoling Chen. 2025. DSC-GRUNet: A lightweight neural network model for multimodal gesture recognition based on depthwise separable convolutions and GRU. *Pattern Recogn. Lett.* 190, C (April 2025), 35–44. <https://doi.org/10.1016/j.patrec.2025.02.008>
- [31] Xijun Ye, Yongjie Cao, Airong Liu, Xinwei Wang, Yinghao Zhao, and Nan Hu. 2023. Parallel convolutional neural network toward high efficiency and robust structural damage identification. *Structural Health Monitoring* 22 (03 2023), 3805–3826. <https://doi.org/10.1177/14759217231158786>
- [32] Chunwei Zhang, Tianpeng Li, Ying Dai, Li Sun, and Ardashir Mohammadzadeh. 2026. *Multilayer Perceptron (MLP) Neural Networks*. Springer, Cham, 13–34. https://doi.org/10.1007/978-3-032-14746-2_2