Efficient and Customizable Dynamic Hand Gesture Recognition for Edge Deployment

Victor Teslaru; Gabriel Pojoga
“Gheorghe Asachi” Technical University of Iasi

1. Problem & Scope

Challenges:

- Real-time performance on resource-constrained devices
- Domain-specific vocabularies need custom datasets, no universal solution
- Ambiguous user intent from natural movement hard to capture in the collection phase

Our solution:

- Hierarchical pipeline (palm → landmarks → gesture)
- Active learning, guides data collection to fix weaknesses

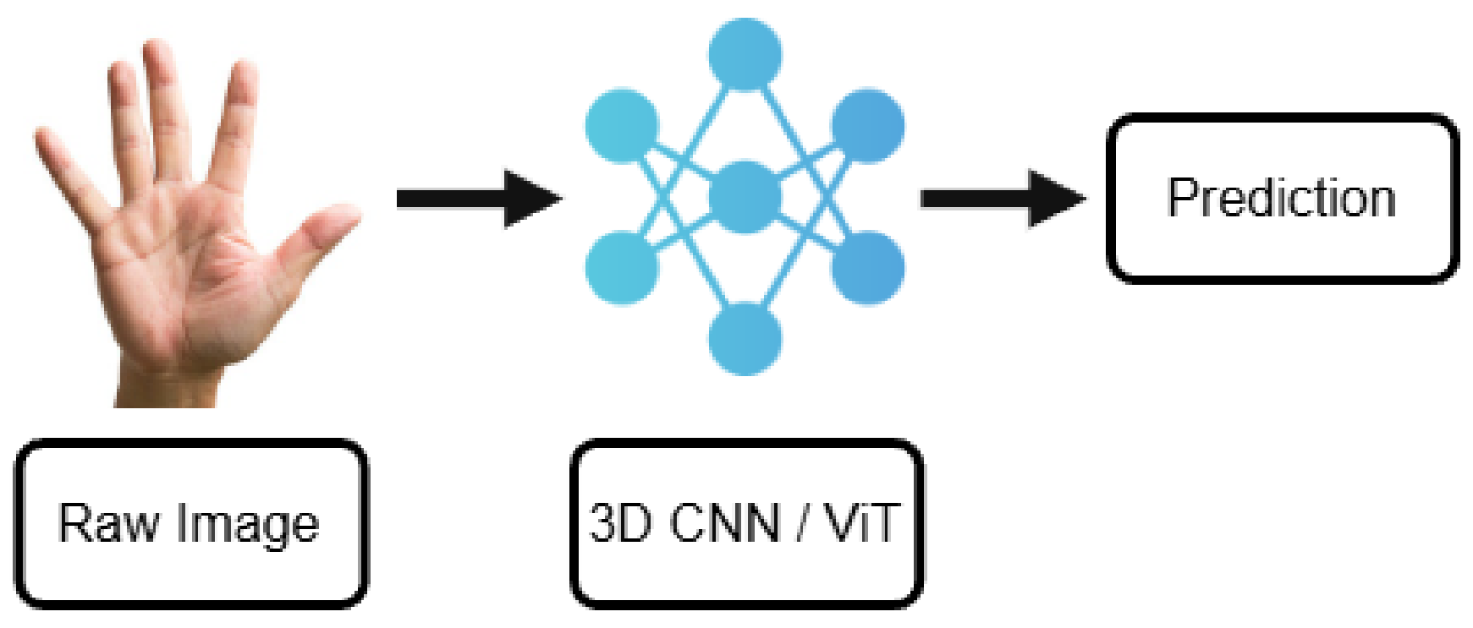


Figure 1. End-to-end approach.

2. Data Pipeline & Dataset

Collection:

- Dynamic Time Warping [1] similarity to prevent redundancy
- Interactive pose rotation augmentation with user validation
- Linear interpolation for ≤2 missing frames, discard otherwise

Dataset:

- 8 gestures: Swipe (4), Zoom (2), Rotate (2)
- 10-frame sequences, 21 3D landmarks
- Wrist and hand span normalized (position/scale-invariant)

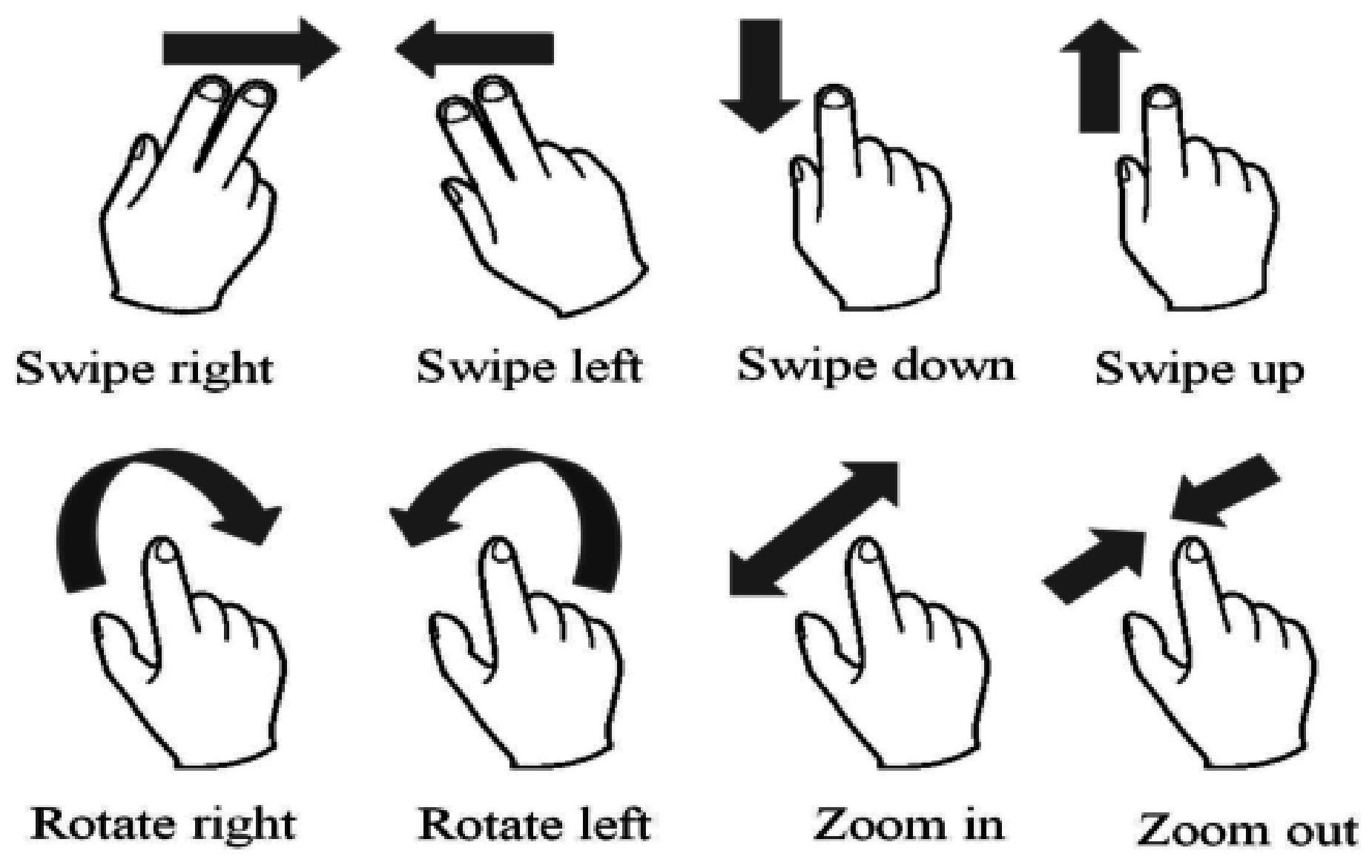


Figure 2. Dynamic hand gestures captured (taken from [2]).

3. Model Design & Edge Optimization

Hierarchical architecture:

- SSD-based palm detector [3], which reduces search space
- 21 key-points extraction for structured representation
- NPU offloading for vision tasks
- Quantized 1D CNN for temporal modelling [4]

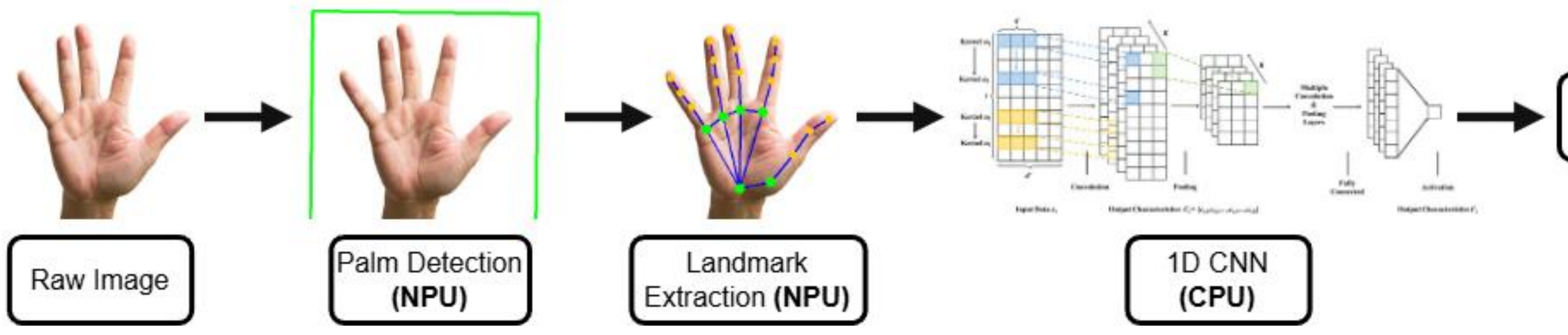


Figure 3. Hierarchical architecture pipeline.

4. Deployment & Inference

Sliding window continuous inference:

- Capture 10-20 frame sequences (user-paced)
- Extract up to 7 overlapping inverse sliding windows, prioritizing recency
- Classify each window independently using uncertainty thresholding
- Filter out ambiguities through majority voting across valid predictions

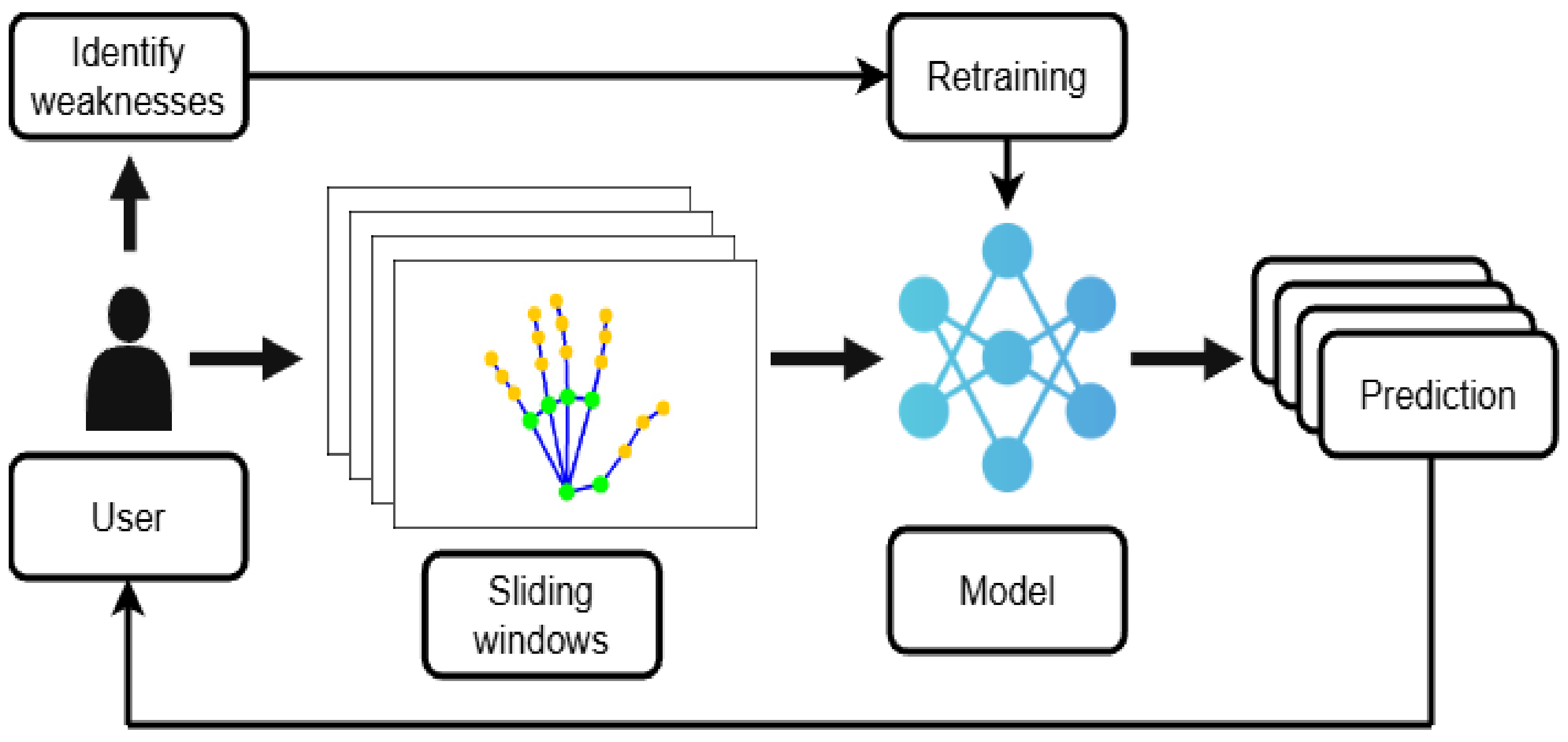


Figure 4. Active learning workflow.

5. Active Learning

Failure driven improvement:

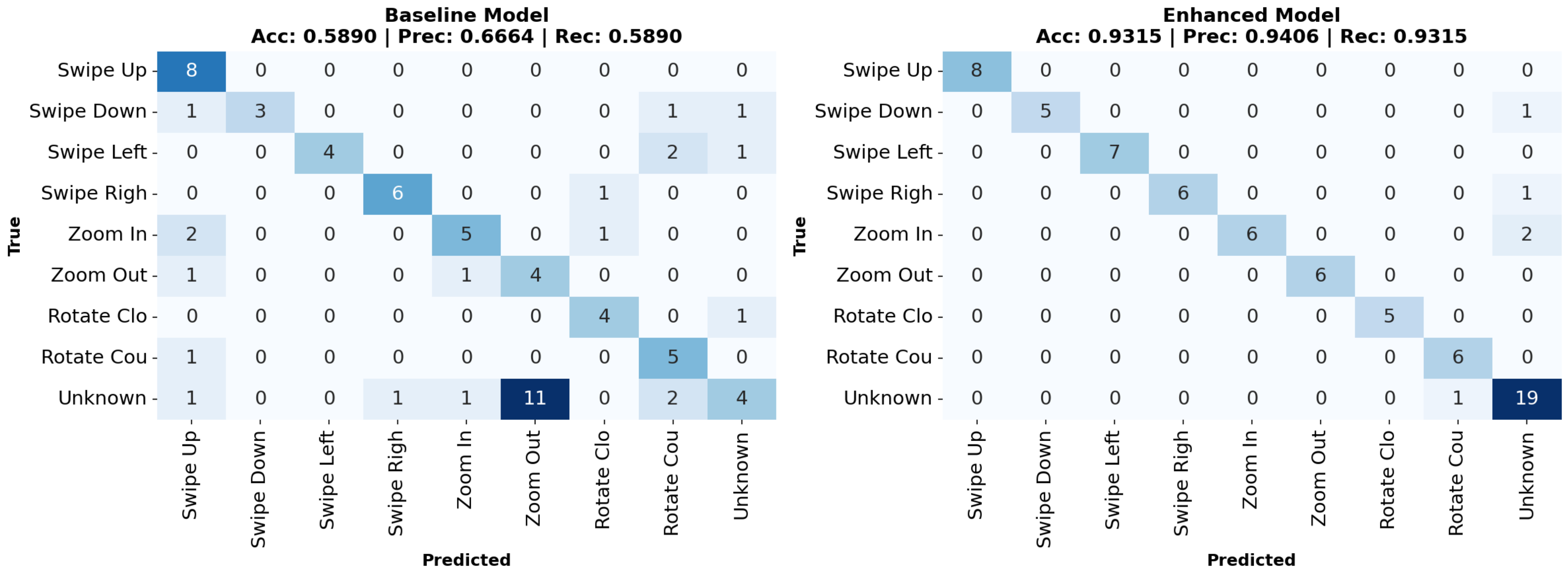
- Actual failure points from source (not hypothetical)
- Multi-window anomalies caught automatically alongside natural movement
- Can adapt to new user variation

Workflow:

- Multiple window predictions
- User reviews (labels / discards)
- Diversity check + augmentation
- Incremental dataset update

6. Evaluation metrics

- Benchmark: 73 gestures with various poses and sliding windows length
- Baseline struggles with unknown class based only on thresholding
- Active learning helps the model better define decision boundaries
- Unknown gesture recall: 20% → 95% (4.75× improvement)



Baseline Model
Acc: 0.5890 | Prec: 0.6664 | Rec: 0.5890

True \ Predicted	Swipe Up	Swipe Down	Swipe Left	Swipe Righ	Zoom In	Zoom Out	Rotate Clo	Rotate Cou	Unknown
Swipe Up	8	0	0	0	0	0	0	0	0
Swipe Down	1	3	0	0	0	0	0	1	1
Swipe Left	0	0	4	0	0	0	0	2	1
Swipe Righ	0	0	0	6	0	0	1	0	0
Zoom In	2	0	0	0	5	0	1	0	0
Zoom Out	1	0	0	0	1	4	0	0	0
Rotate Clo	0	0	0	0	0	0	4	0	1
Rotate Cou	1	0	0	0	0	0	0	5	0
Unknown	1	0	0	1	1	11	0	2	4

Enhanced Model
Acc: 0.9315 | Prec: 0.9406 | Rec: 0.9315

True \ Predicted	Swipe Up	Swipe Down	Swipe Left	Swipe Righ	Zoom In	Zoom Out	Rotate Clo	Rotate Cou	Unknown
Swipe Up	8	0	0	0	0	0	0	0	0
Swipe Down	0	5	0	0	0	0	0	0	1
Swipe Left	0	0	7	0	0	0	0	0	0
Swipe Righ	0	0	0	6	0	0	0	0	1
Zoom In	0	0	0	0	6	0	0	0	2
Zoom Out	0	0	0	0	0	6	0	0	0
Rotate Clo	0	0	0	0	0	0	5	0	0
Rotate Cou	0	0	0	0	0	0	0	6	0
Unknown	0	0	0	0	0	0	0	1	19

Figure 5. Model evaluation before and after active learning.

7. Privacy & Safety

Privacy preserving design:

- Only 3D landmarks collected — No identity exposure
- 100% local processing (edge/desktop)
- User retains full data control

Safety Features:

- Explicit "Unknown" class for avoiding accidental triggers
- Multi-window ambiguity detection

Contact

Victor Teslaru
Gabriel Pojoga
“Gheorghe Asachi” Technical University of Iasi
Email: victor.teslaru@student.tuiasi.ro
gabriel.pojoga@student.tuiasi.ro

References

1. H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 1, pp. 43-49, February 1978, doi: 10.1109/TASSP.1978.1163055.
2. Tumuganti, Nagakarthik & Ahn, Eun & Bae, Yun & Choi, Jun. (2013). TCAM based pattern matching technique for hand gesture recognition. 368-369. 10.1109/ISOCC.2013.6864052.
3. Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214.
4. TensorFlow Team. "Post-training quantization," TensorFlow Lite Documentation, 2023. https://www.tensorflow.org/lite/performance/post_training_quantization