Differentially Private Integrated Decision Gradients (IDG-DP) for Radar-based Human Activity Recognition

Idris Zakariyya, Linda Tran, Kaushik Bhargav Sivangi, Paul Henderson and Fani Deligianni
School of Computing Science, University of Glasgow, G12 8RZ, Glasgow, United Kingdom

EPSRC
Pioneering research and skills

Abstract

Human motion analysis holds significant promise for healthcare monitoring and early disease detection. Radar-based sensing systems have gained attention for their ability to operate contactlessly and integrate with existing Wi-Fi networks, while being less intrusive than camera-based systems. However, recent studies have revealed that radar gait patterns can accurately identify individuals and their gender, raising privacy concerns. This study addresses these issues by exploring privacy vulnerabilities in radar-based Human Activity Recognition (HAR) systems and introducing a novel privacy-preserving method using Differential Privacy (DP) informed by attributions derived from the Integrated Decision Gradient (IDG) algorithm. We examine Black-box Membership Inference Attacks (MIAs) in HAR settings under varying levels of attacker-accessible information. The proposed IDG-DP method was rigorously evaluated through the development of a DNN-based HAR model, with its robustness against MIAs tested. IDG-DP effectively mitigates privacy attacks while maintaining model utility, particularly in defending against label only MIA.

Background

- Demographic projections estimate that the global population aged 60 and above will reach 1.4 billion by 2030 [1].
- Human motion analysis holds great promise for healthcare monitoring and early disease detection [2].
- Deep Neural Networks (DNNs) have proven highly effective in Human Activity Recognition (HAR) for patient monitoring.
- DNNs are vulnerable to privacy attacks, particularly the Membership Inference Attacks (MIA) which can compromise patient data.
- Developing a DNN based method to mitigate privacy attacks in radar-based HAR systems presents a significant challenge.

Methodology

- By utilizing a transfer-learned ResNet models, we develop a novel privacy preservation approach, namely IDG-DP, that builds on both IDG and Pure-DP method.
- IDG leverages that moving across a gradient path from the decision regions would rapidly affect the output logit.
- Its robustness stems from its strong theoretical guarantees of sensitivity, implementation invariance and completeness.
- Algorithm 1 describe the detail procedure for the development of the IDG-DP method

```
Algorithm 1 Algorithmic description of IDG-DP
    Input: {$\epsilon$, attribution threshold $at$, training data $\mathcal{D}$, number of iterations $\mathcal{T}$, multi-task HAR ($\mathcal{MHAR}$)}
    Output: {IDG-DP}
Function DP($\mathcal{D}$, $\mathcal{MHAR}$)
estimates $sa$                      ▷ $sa$ = subjects attribution feature maps
estimates $aa$                      ▷ $aa$ = activity attribution feature maps
estimates avg of $aa$ and $sa$      ▷ avg = average
estimates std of $aa$ and $sa$      ▷ std = standard deviation
estimates $ns$ based on $at$ ▷ ns = noise indices = indices where avg $< at$
for $i = 0$ to $\mathcal{T}$ do
    $\mathcal{NM} = \mathcal{MHAR}(\epsilon, ns)$      ▷ $NM$ = noise model
    if $\mathcal{NM}$ converges then
        $\mathcal{NM}$ = IDG-DP
        break
    end if
end for
return IDG-DP
End Function
```

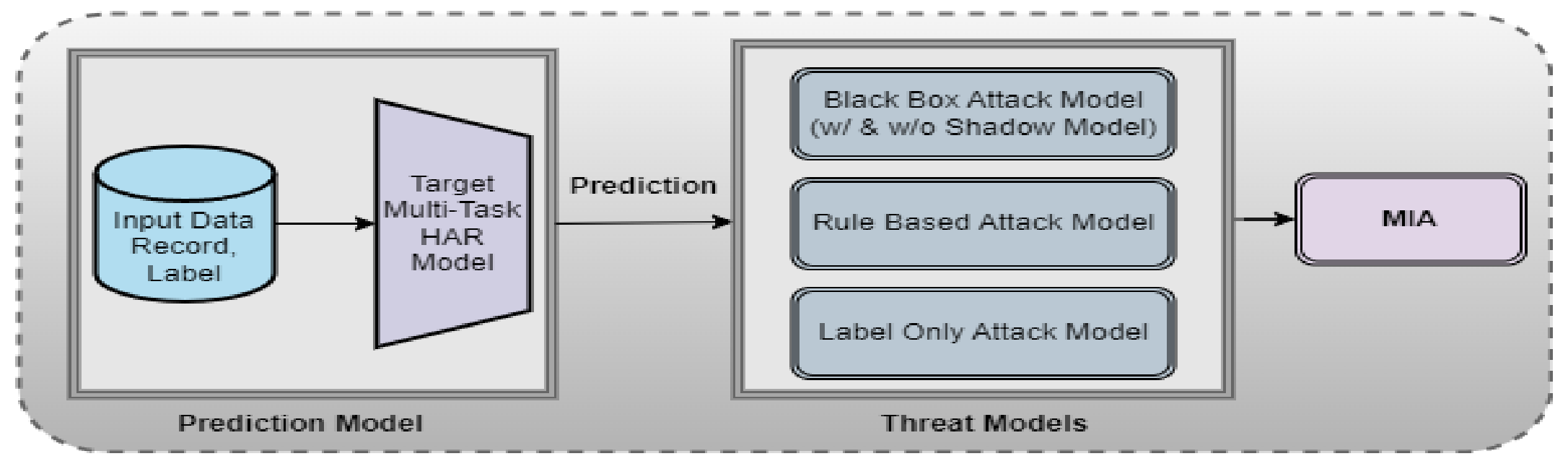
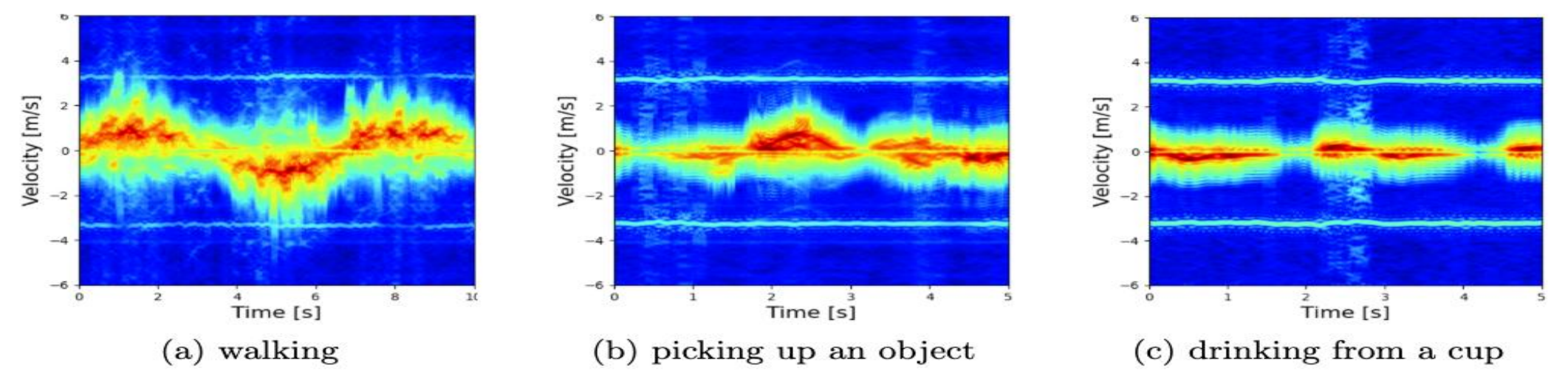


Fig.1 IDG-DP Threat Models.

Table 1: Data Distribution .

Procedure	Training	Testing
HAR	60	30
Blackbox MIA	25	25
Rule-based MIA	25	25
Blackbox MIA with 3 shadow model	60	30
Blackbox MIA with 10 shadow model	25	25
Label-Only 25	25	25
Label-Only 10	20	20
Label-Only 20	40	30

*Evaluation samples for Label-Only 10 = 10, and 20 for Label-Only 20.



(a) walking (b) picking up an object (c) drinking from a cup

Fig.2 Radar HAR Activities.

Results

Table 1 presents HAR performance comparison where IDG-DP demonstrate better HAR performance compared with all tested method. It excel in detecting MIA attacks, showing a stronger robustness in shadow model attacks setting and label only adversarial attacks. These results demonstrates the effectiveness of IDG-DP method in protecting HAR systems against MIA privacy attacks, especially in radar settings where privacy is a great concerns.

Table 1. Performance evaluation comparison of HAR for the base-line and various multi-task DP based activity models. ϵ = 1.20.

Model	Accuracy↑ (%)	Precision↑ (%)	Recall↑ (%)	F1-Score↑ (%)
Baseline	83.30	83.30	83.30	82.90
Base-DP	90.00	90.00	90.00	90.00
Optics	63.30	63.30	63.30	56.10
IG-DP	93.30	93.30	93.30	93.30
GradC-DP	83.30	83.30	83.30	82.20
Sal-DP	80.00	80.00	80.00	80.00
IIG-DP	90.00	90.00	90.00	89.80
ISG-DP	90.00	90.00	90.00	90.30
IDG-DP	**96.70**	**96.70**	**96.70**	**96.70**

Performance evaluation comparison for the baseline and various HAR DP based activity models with ϵ = 1.20 against black box MIA using 10 shadow models described in Table 1.

Model	Clean Test Accuracy (%)↑	Total Attack Accuracy (%)↓	Total Attack Precision (%)↓	Total Attack Recall (%)↓
Baseline	83.33	64.00	62.07	72.00
Base-DP	90.00	56.00	55.17	64.00
Optics	63.33	66.00	75.00	48.00
IG-DP	93.33	44.00	41.18	28.00
GradC-DP	80.00	58.00	57.14	64.00
Sal-DP	80.00	42.00	44.44	64.00
IIG-DP	90.00	46.00	43.75	28.00
ISG-DP	90.00	48.00	46.67	28.00
IDG-DP	**96.70**	**40.00**	**35.29**	**24.00**

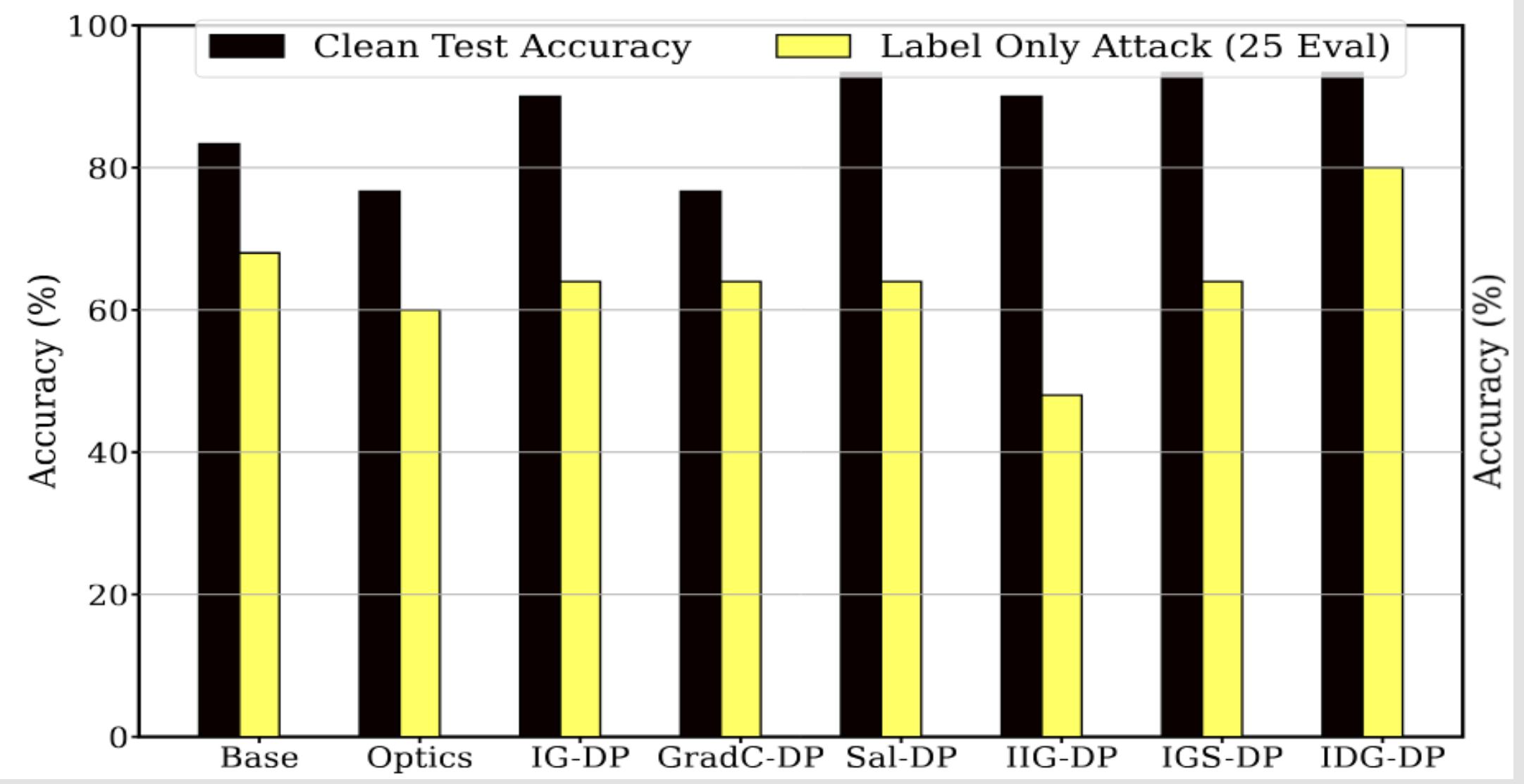


Fig.2 HAR DP-based models performance comparison against Label only 25 MIA attacks in Table 1.

Conclusion

Our study breaks new ground in addressing privacy preservation for radar-based human motion analysis models. %To the best of our knowledge, this study is among the first to address privacy preservation in radar-based human motion analysis models. We introduce a novel IDG-DP method, which exploits the solid theoretical guarantees of Differential Privacy (DP) with a principle way driven by the Integrated Decision Gradients (IDG) of a multi-task model to preserve the utility of the dataset. We assess IDG-DP's robustness against various black-box membership inference attacks (MIA), including label-only attacks. Our findings demonstrate that DP, when applied with carefully selected attributions effectively mitigates MIA. IDG-DP maintain better HAR performance and MIA robustness. IDG-DP's accuracy surpasses that of the tested benchmarks, making it an optimal choice for balancing privacy protection and utility in activity recognition. This research provides a solid foundation for developing privacy-preserving techniques in radar-based human motion analysis, paving the way for more secure a HAR applications in healthcare monitoring.

References

1. World Health Organisation. Ageing and health. https://www.who.int/news-room/fact sheets/detail/ageing-and-health, 2022. Ac-cessed: 2023-11-02.
2. Lazzaro di Biase, Alessandro Di Santo, Maria Letizia Caminiti, Alfredo De Liso,Syed Ahmar Shah, Lorenzo Ricci, and Vincenzo Di Lazzaro. Gait analysis inparkinson's disease: An overview of the most accurate markers for diagnosis and symptoms monitoring. sensors, 20(12):3529
3. Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership in-ference attacks against machine learning models. In 2017 IEEE symposium on security and privacy (SP), pages 3–18. IEEE, 2017.

Acknowledgements

We acknowledge funding from EPSRC(EP\W01212X\1), the Royal Society (RGS\R2\212199) and Academy of Medical Sciences (NGR1\1678).