EMBEDDED TRUSTWORTHY AI FOR HEALTHCARE

A Multi-Objective Study of Fairness, Privacy, and Efficiency under TinyML Constraints

Author: Luca Tompea
l.tompea@student.tudelft.nl

Supervisor: Dr. Qing Wang

TU Delft Delft University of Technology

1. INTRODUCTION

- **The Rise of TinyML:** Moving healthcare diagnostics to microcontrollers (Edge AI) [1].
- **The Trust Gap:** High accuracy is insufficient; models must be fair (no demographic bias) and private (protecting sensitive medical data).
- **Hardware Constraints:** Deploying on resource-constrained devices requires model compression (quantization).

2. RESEARCH QUESTION

How do model quantization and privacy-preserving perturbations influence the trade-off between predictive accuracy and group fairness in healthcare TinyML models?

3. METHODOLOGY

- **Datasets:**
  - **Pima Indians Diabetes [2]:** 768 samples, 8 features, age-stratified groups — all fairness, privacy and threshold analyses.
  - **Diabetes 130-US Hospitals [3]:** ~101k samples, 154,817-parameter MLP — quantization efficiency and fairness at scale.
- **Models:** Multi-Layer Perceptron (MLP) and Logistic Regression.
- **Interventions:**
  - **Privacy:** Differential Privacy via Gaussian Noise Injection at levels 0 to 0.5, functioning as a Gaussian mechanism privacy proxy [4].
  - **Efficiency:** Post-Training Quantization (INT8 = 8-bit integer quantization from FP32 = 32-bit floating point) targeting Raspberry Pi 4B deployment [1].
- **Metrics:** Accuracy, Demographic Parity (DP) Gap, and Equal Opportunity(EO) Gap [5].

4.1. BASELINE FAIRNESS-ACCURACY COMPARISON

Performance and Fairness across 10-fold cross-validation MLP vs Logistic Regression on Pima:

1.MLP:
- Accuracy: 74.98% (+/- 4.95%)
- AUC Score: 0.835
- Fairness:
  - DP gap: 0.310
  - EO gap: 0.177

2.LOGISTIC REGRESSION:
- Accuracy: 76.42% (+/- 5.96%)
- AUC: 0.837
- Fairness:
  - DP gap: 0.308
  - EO gap: 0.152

4.2. PRIVACY EFFECTS ON FAIRNESS-ACCURACY USING LOGISTIC REGRESSION

As seen in Table 2, as privacy-preserving noise increases, utility degrades and fairness gaps change:

- **Accuracy Drop:** Accuracy decreases from 76.54% (Noise 0.0) to 72.62% (Noise 0.5) - Figure 1.
- **Fairness Instability:** Noise injection does not affect all groups equally; as noise increases, the model's ability to maintain fair outcomes across demographics fluctuates non-monotonically, creating a "privacy-fairness" tension as shown in Figure 2. Privacy does not directly correlate with fairness

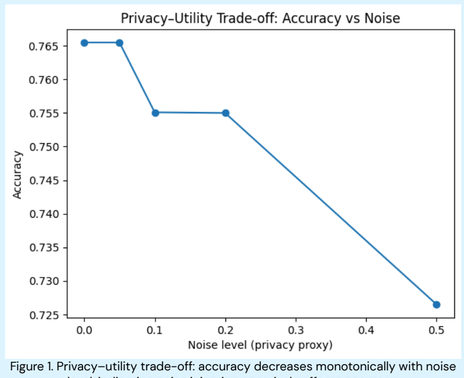

Figure 1. Privacy–utility trade-off: accuracy decreases monotonically with noise level, indicating noise injection negatively affects accuracy.

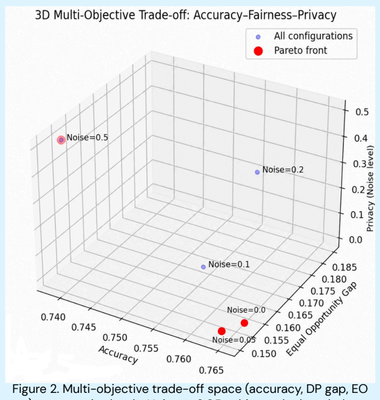

Figure 2. Multi-objective trade-off space (accuracy, DP gap, EO gap) across noise levels. Noise σ=0.05 achieves the best balance across all three objectives.

- **The Mechanism**: Injecting random noise (δ) obscures individual data points.
- **The Math**: Higher noise mathematically lowers the privacy budget (ε), yielding stronger privacy.
- **The Reality**: Our tested noise levels act as a practical, measurable proxy for privacy, though they remain too high (ε >> 1) for certified differential privacy.
  - ε ranges from 96.9 (σ=0.05) to 9.7 (σ=0.5)

Noise	Accuracy	DP Gap	EO Gap
0.00	0.765499	0.308710	0.152812
0.05	0.765482	0.303116	0.141297
0.10	0.755058	0.303586	0.173662
0.20	0.755041	0.326728	0.148815
0.50	0.726452	0.302985	0.166698

Table 2. Effect of Gaussian noise injection on accuracy and fairness (Logistic Regression, 10-fold CV, median age split).

4.3. QUANTIZATION EFFECTS ON FAIRNESS-ACCURACY USING MLP

Quantization efficiency and fairness effects are scale-dependent, summarised in Table 3 and detailed for Pima in Table 4 under the effect of noise.

Small scale (Pima, 221 params):
- **Efficiency:** INT8 TFLite (3.10 KB, 0.079 ms) is larger and 6.5× slower than FP32 (2.93 KB, 0.012 ms); metadata overhead exceeds weight savings. Accuracy drops ~2% across all noise levels.
- **Fairness:** INT8 amplifies EO gap under noise. At σ=0.5, INT8 EO gap (0.369) is 67% higher than FP32 (0.221).

Large scale (Hospital, ~155k params):
- **Efficiency:** INT8 achieves 3.87× size reduction (613 → 159 KB) and 3.6× speedup (0.091 → 0.025 ms) on Pi 4.
- **Fairness:** INT8 reduces EO gap by 61% (0.041 → 0.016) at no accuracy cost, the opposite direction from Pima.

Dataset	Params	FP32 TFLite (KB)	INT8 TFLite (KB)	Reduction	FP32 Pi (ms)	INT8 Pi (ms)	Speedup
Pima	221	2.93	3.10	0.94×	0.012	0.079	0.15×
Hospital	154,817	613.05	158.50	3.87×	0.091	0.025	3.6×

Table 3. Scale-dependent quantization efficiency (FP32 and INT8 TFLite, measured on Raspberry Pi 4B). Reduction = FP32/INT8 size; Speedup = FP32/INT8 latency. Values < 1× indicate INT8 is larger/slower.

σ	FP32 Acc	FP32 DP	FP32 EO	INT8 Acc	INT8 DP	INT8 EO
0.00	0.754	0.310	0.177	0.732	0.324	0.162
0.05	0.754	0.320	0.180	0.741	0.327	0.180
0.10	0.754	0.323	0.192	0.745	0.349	0.245
0.20	0.759	0.329	0.220	0.732	0.351	0.279
0.50	0.749	0.244	0.221	0.719	0.352	0.369

Table 4. MLP fairness and accuracy under quantization across noise levels. INT8 consistently widens both DP and EO gaps relative to FP32.

5. CONCLUSION

- **Viability is scale-dependent:** At Pima scale (221 params), quantization provides no efficiency benefit and amplifies fairness disparities. At Hospital scale (~155k params), INT8 achieves 3.87× compression, 3.6× speedup, and a 61% EO gap reduction — simultaneously efficient and fairer.
- **Privacy Cost:** Noise injection degrades accuracy by ~4% (76.5% → 72.6%) and creates fairness instability: EO gap fluctuates non-monotonically, revealing a privacy–fairness tension.
- **Future Work:** Validating the parameter threshold at which INT8 gains reliably materialise across diverse clinical datasets, and replacing heuristic noise proxies with formal DP-SGD for rigorous privacy–fairness characterisation.

REFERENCES:

[1] P. Warden and D. Situnayake, TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. O'Reilly Media, 2019.
[2] UCI Machine Learning Repository. Pima indians diabetes database. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database/data, 1990
[3] Strack et al. (2014). Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. BioMed Research International, 2014, 781670.
[4] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.
[5] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in Advances in Neural Information Processing Systems (NeurIPS), vol. 29, 2016.