Supplementary material

No Author Given

No Institute Given

1 Proof of Proposed Loss Function

When all H hospitals exhibit identical training loss, the optimal solution for the fairness term emerges. To demonstrate this, we minimize L_{fair}. Subsequently, we elaborate on the mathematical representation of the fairness term. Further, we discuss the enhancement in performance resulting from integrating the fairness term into the local training loss, as described by Eq. 1.

$$L_{fair} = \sum_{h=1}^{H} (F_h(w) - F(w))^2 \tag{1}$$

$$\begin{aligned} L_{fair} &= \sum_{h=1}^{H}(F_h^2(w) + F(w)^2 - 2F_h(w)F(w)) \\ &= \sum_{h=1}^{H}(F_h^2(w) - 2F_h(w)F(w)) + HF(w)^2 \end{aligned} \tag{2}$$

To minimize the loss function, we computed the derivative with respect to $F_j(w)$ for a given hospital $j \in 1, 2, \cdots, H$. Setting it to zero, as represented in Eq. 3, enables us to identify the minima.

$$\begin{aligned} \frac{dL_{fair}}{dF_j(w)} &= 2F_j(w) = 0 \\ F_j(w) &= 0 \end{aligned} \tag{3}$$

The minimization of L_{fair} implies that $F_h(w) = 0$ holds true for all hospitals, which is equivalent to the condition $F_h(w) = F(w)$ for all $h \in 1, 2, 3, ..., H$. This indicates that all M hospitals will converge to the same local loss, ensuring fairness. We have shown our proposed loss function by plotting hospital 1's loss (F_1) while fixing the total loss. The plot (refer to Fig. 1) confirms the quadratic penalty effect - the loss grows with the divergence between F1 and the fair 0.5 value. Importantly, the plot also shows that the global minimum is precisely when $F_1 = F_2$. The equal losses satisfy the condition for fairness we derived mathematically in the proof. That clearly means all clients will converge to the same local loss, satisfying fairness across all H hospitals.

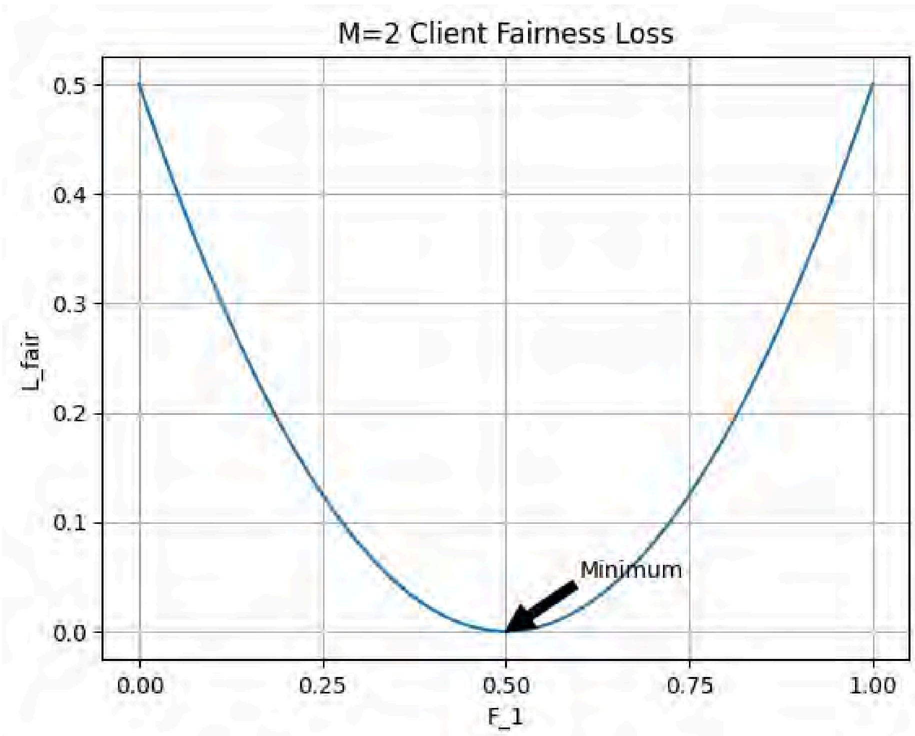

Fig. 1. The plot of the proposed loss function considering the 2 clients. It is clear from the graph that the minimum of the loss function occurs when both clients have the same loss value i.e. tending to zero