Model Performance Comparison

TFCL vs. Supervised Deep Models:

- Pure CNN: TFCL significantly outperforms pure CNN. TFCL achieves an F1 score of 0.9791 on the DLR dataset, while CNN only reaches 0.7512. CNN struggles with long-range temporal dependencies, which TFCL addresses by combining CNNs and LSTMs, enabling the model to better capture temporal patterns.

- Pure LSTM: While LSTM is strong at capturing temporal dependencies, it still yields an F1 score of 0.8265, lower than TFCL's. The lack of biometric data integration in the LSTM model limits its personalization capabilities, which gives TFCL an edge in accurately recognizing individual activity patterns.

TFCL vs. Classical Machine Learning Models:

- SVM (RBF Kernel): The SVM with an RBF kernel achieves an F1 score of 0.6157, but is significantly outperformed by TFCL. SVM's inability to model temporal dependencies in sensor data limits its effectiveness for HAR tasks.

- Random Forest: Random Forest obtains an F1 score of 0.7384, which is an improvement over SVM, but still below TFCL's 0.9791. The model struggles to capture the temporal dynamics crucial for activity recognition tasks.

- K-Nearest Neighbors (k=5): KNN scores an F1 of 0.6485, which is lower than TFCL and other machine learning models. Its high computational cost and inability to handle time-dependent features make it less suitable for HAR tasks.

- Logistic Regression: Logistic Regression shows the poorest performance, with an F1 score of 0.5213. Its linear nature and lack of temporal modeling prevent it from effectively recognizing complex human activities.

Overall, **TFCL** achieves the highest F1 score and shows superior generalization and personalization capabilities, making it the top performer for human activity recognition.

F1 Score and Accuracy Comparison

This shows the comparison of **F1 scores** and **accuracy** across the various models:

TFCL outperforms all other models in F1 score, with a **0.9791** and by achieving **96.5%** accuracy. TFCL significantly outperforms both classical machine learning models and deep learning models, particularly in terms of generalization and personalization, making it the leading model for wearable sensor-based HAR tasks.

Model	F1 Score	Accuracy (%)
Logistic Regression	0.5213	48.1
SVM	0.6157	48.2
Decision Tree	0.6625	48.9
Random Forest	0.7384	51.2
CNN	0.7512	87.9
LSTM	0.8265	92.5
TFCL	0.9791	96.5