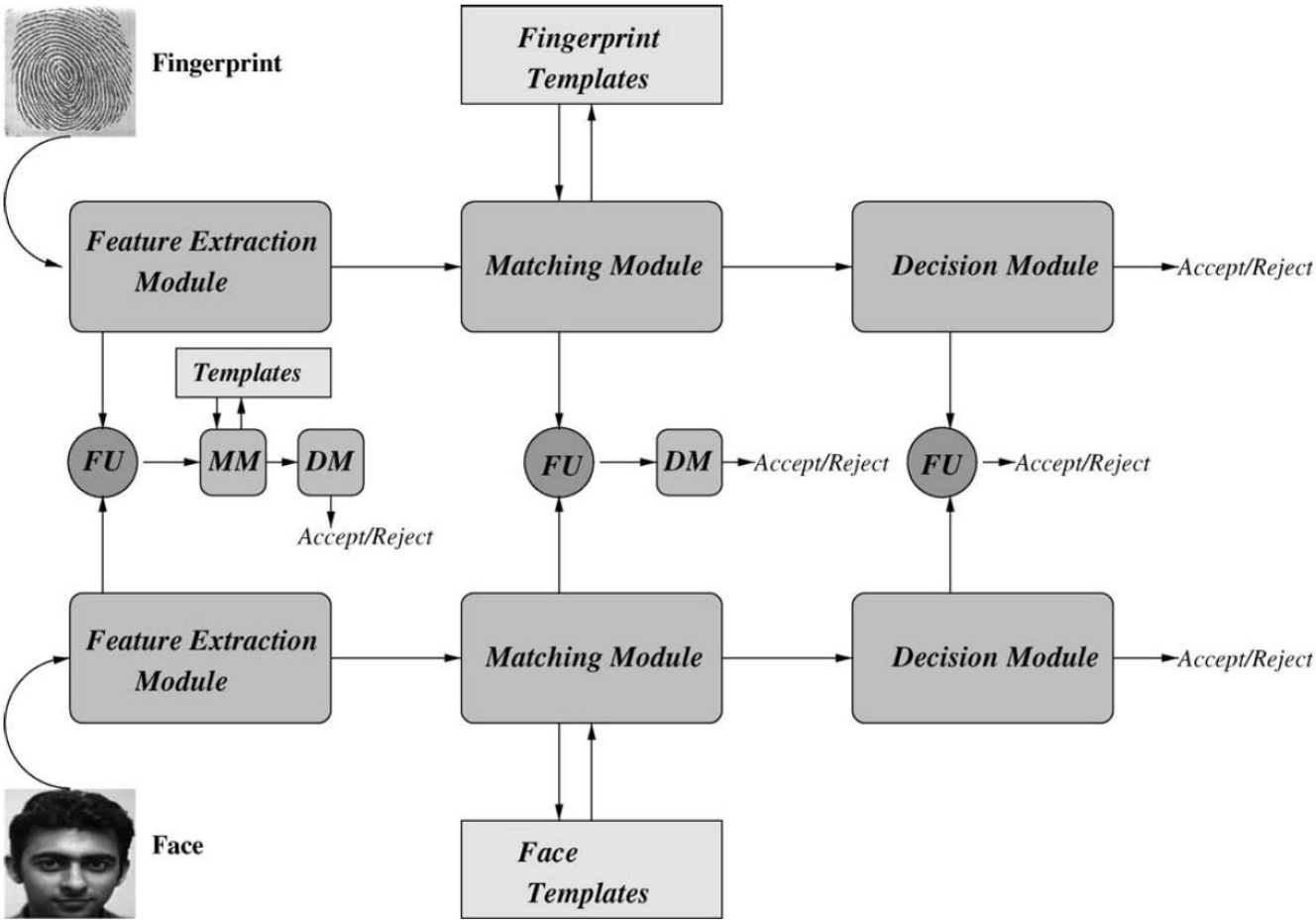
Worksheet 5 (StepScan)

Topic 1: Information Fusion in Biometrics

1) The paper entitled “Information Fusion in Biometrics” [1] is the main resource on this topic.
2) From Section 3 in the paper, the three possible levels of fusion are:
   a. **Fusion at the feature extraction level**: The new feature vector with a higher dimensionality can be created by combining many features together. To extract useful features from the larger set of features, feature reduction techniques may be employed: (1) feature projection, such as PCA (Worksheet 3 Topic 2) and (2) feature selection, such as mRMR (Worksheet 4 Topic 3).
   b. **Fusion at the matching score level**: Each system provides a matching score indicating the proximity of the feature vector with the template vector. These scores can be combined to assert the veracity of the claimed identity. In Worksheet 3 Topic 1, we applied the minimum, mean, or median distance to the scores, and we can consider these as the simple fusion techniques at the matching score level. In this worksheet, we will try other fusion approaches at the matching score level (more details in Topic 2, 3, and 4).
   c. **Fusion at the decision level**: The resulting feature vectors individually classified into the two classes—accept or reject. A majority vote scheme can be used to make the final decision.



Topic 2: Matching Score Level Fusion – Template Clustering

1) The paper entitled “The Smart Floor: A Mechanism for Natural User Identification and Tracking” [2] is the main resource on this topic.
2) The simplest way to choose the identity of the unknown footstep is to use the identity of the footstep that is closest in feature space (minimum distance). A more sophisticated method is based on **footstep clustering**. Each user has a number of known training footsteps; these footsteps form clusters of points in feature space. The distance is computed from the unknown identity footstep to every footstep in every cluster. The identity of *the cluster with the lowest average distance* is chosen as the identity of the unknown footstep.
3) In the paper, each subject provided separate data for left and right feet as well as data for as many shoe types as possible. The footsteps for a single condition (referring to one foot and one shoe type) is considered as a footstep cluster (*supervised clustering*).

Topic 3: Matching Score Level Fusion – Template Selection

1) The paper entitled “Biometric Template Selection and Update: A Case Study in Fingerprints” [3] is the main resource on this topic.
2) Given a set of *N* footsteps corresponding to a single user, instead of calculating the minimum, mean, or median distance from all *N* templates, we can select *K* templates that ‘best’ represent the variability as well as the typicality observed in the *N* footsteps, $K < N$.
3) Two methods for template selection are proposed in [2]:
    a. **DEND**: In this method, the *N* footsteps corresponding to a user are grouped into *K* clusters using hierarchical clustering (*unsupervised clustering*), such that footsteps within a cluster are more similar than footsteps from different clusters. Then for each cluster, a prototype (representative) footprint that typifies the members of that cluster is chosen, resulting in *K* templates. This technique, therefore, selects prototypes that represent the variability observed in the footprints.
    b. **MDIST**: This method sorts the footprints based on their average distance score with other footprints and selects those footprints that correspond to the *K* smallest average distance scores. The rationale is to select templates that exhibit maximum similarity with other footprints and, hence, represent typical data measurement.
4) Finally, we can calculate the minimum, mean, or median distance from the *K* templates, instead of *N* templates.

Note: We can combine supervised footstep clustering together with either DEND or MDIST for multi-stage fusion.

Topic 4: Matching Score Level Fusion – k-Nearest Neighbors and Other Machine Learning Algorithms

1) The paper entitled “Information Fusion in Biometrics” [1] is the main resource on this topic.
2) Instead of using summary statistics like minimum or mean, we can create a (second stage) feature vector from matching scores and then use as an input for machine learning methods like k-nearest neighbors (kNN) (an extended version of the minimum distance), decision trees (DT) and linear discriminant analysis (LDA), as proposed in [1], or any other like support vector machine (SVM).

Note: There are many advanced fusion methods [4], but we will work on those algorithms later.

[1] Ross & Jain, “Information Fusion in Biometrics”, Pattern Recognition Letters, 24, pp. 2115-2125, 2003.
[2] Orr & Abowd, “The Smart Floor: A Mechanism for Natural User Identification and Tracking,” in *Proc. CHI 2000*, pp. 275-276.
[3] Uludag et al., “Biometric Template Selection and Update: A Case Study in Fingerprints,” Pattern Recognition, 37, pp. 1533-1542, 2004.
[4] Singh et al., “A Comprehensive Overview of Biometric Fusion,” Information Fusion, 52, pp. 187-205, 2019.

Checkpoint

(1) To compare the performance of footstep clustering, DEND, MDIST, hybrid versions (footstep clustering + DEND, footstep clustering + MDIST), kNN, DT, LDA, and SVM for the fusion method using the matching scores.
(2) To apply majority voting on the nine systems in (1).