CLAIMS

We Claim:

1. A computer-implemented method for improving security of a biometrics-based authentication system, comprising:

receiving, by one or more servers, enrolled biometric samples of an enrolled user during an enrollment stage of the biometrics-based authentication system;

applying, using a projection deep neural network, a first function $f(\cdot)$ to the enrolled biometric sample, x_i, to generate a first enrolled biometric template $f(x_i)$ and to project the enrolled biometric samples, x_i, to a common embedding subspace.

creating, by the one or more servers, augmented biometric samples by adding learned perturbations to the enrolled biometric samples of the enrolled user;

receiving, by the one or more servers, during a request for authentication, submitted biometric samples from a second user purporting to be the enrolled user;

applying, using the projection deep neural network, the first function $f(\cdot)$ to the submitted biometric sample, x', to generate a first submitted biometric template f(x') and to project the submitted biometric samples, x', to the common embedding subspace;

comparing, by the one or more servers, the submitted biometric samples of the second user to the enrolled biometric samples and to the augmented biometric samples of the enrolled user based on predefined metrics, wherein the comparing comprises computing a distance in the common embedding subspace between the first enrolled biometric template $f(x_i)$ and the first submitted biometric template f(x'); and

based on the comparison, determining, by the one or more servers, that the submitted biometric samples of the second user have been modified to impersonate the enrolled user.

2. The method of claim 1, further comprising: storing the enrolled biometric samples x_i, the first enrolled biometric template $f(x_i)$ and the augmented biometric samples as an enrolled biometric dataset.

3. The method of claim 1 or claim 2, comprising:

comparing the distance to a threshold; and

if the distance is greater than the threshold, rejecting the request for authentication.

4. The method of any one of the preceding claims, wherein comparing the submitted biometric samples of the second user to the enrolled biometric samples and to the augmented biometric samples of the enrolled user based on predefined metrics, further comprises, responsive to determining that the distance is less than a first threshold, temporarily authorizing the request for authentication.

5. The method of claim 4, wherein determining that the submitted biometric samples of the second user have been modified is activated only in response to the request for authentication being temporarily authorized.

6. The method of claim 5, further comprising:

responsive to the request for authentication being temporarily authorized, receiving, by an adversarial perturbation detector, the enrolled biometric samples x_i of the enrolled user and the submitted biometric samples of the second user, where the submitted biometric samples comprise perturbed adversarial samples $x' + \Delta x'$; and

responsive to detecting any perturbed adversarial samples, rejecting the request for authentication of the second user, and otherwise, granting the request for authentication and authorizing the second user.

7. The method of claim 6, wherein determining that any perturbed adversarial samples are detected further comprises:

applying a second function $k(\cdot)$ to both the enrolled biometric samples x_i and the perturbed adversarial samples $x' + \Delta x'$ to generate a second biometric template $k(x_i)$ and a second submitted biometric template $k(x' + \Delta x')$, which are in a transformed subspace;

computing a distance F between the second biometric template $k(x_i)$ and the second submitted biometric template $k(x' + \Delta x')$ in the transformed subspace; and

determining that the second submitted biometric template $k(x' + \sum \Delta x')$ is adversarial when the distance F is greater than a second threshold t' indicating that one or more perturbations have been detected.

8. The method of claim 7, wherein the enrolled biometric samples x_i include augmented biometric samples xi' such that the enrolled biometric samples xi = [xi, xi'], and wherein the x_i and x_i' of the enrolled biometric samples are input to the second function $k(\cdot)$ using one of a parallel model and a sequential model.

9. The method of claim 7 or claim 8, further comprising:

generating training enrolled biometric samples, training public biometric samples, and training perturbed adversarial samples;

inputting the training enrolled biometric samples, the training public biometric samples, and the training perturbed adversarial samples to the second function k(·), where k(·) has learnable parameters θ to generate a training enrolled biometric template, a training public biometric template, and a training perturbed adversarial template;

classifying, by a classifier with learnable parameters σ, the training perturbed adversarial template as a 1 and the training public biometric template as 0 based on the training enrolled biometric template; and

based on a result of the classification, updating the learnable parameters θ and σ.

10. The method of any one of the preceding claims, wherein the enrolled biometric samples comprise images of the enrolled user's face or images of the enrolled user's head.

11. A non-transitory computer readable medium having stored thereon software instructions that, when executed by a processor, cause the processor to perform the method of any one of the preceding claims.

12. A system, comprising:

a memory;

a processor coupled to the memory; and

a software component executed by the processor that is configured to perform the method of any one of claim 1 to claim 10.