**Accuracy Metrics**

* True Positive Rate (TPR): Measures the proportion of actual legitimate attempts that are correctly identified. High TPR is crucial for ensuring legitimate users can access their devices without hindrance.
* True Negative Rate (TNR) / Specificity: Reflects the proportion of actual unauthorized attempts that are correctly identified. A high TNR indicates the system's effectiveness in keeping intruders out.
* False Positive Rate (FPR): The rate at which legitimate users are incorrectly identified as unauthorized. Lower FPRs are desirable to minimize inconvenience to legitimate users.
* False Negative Rate (FNR): The rate at which unauthorized attempts are mistakenly allowed. Minimizing FNR is critical for security.
* Precision: The proportion of positive identifications that were actually correct. Precision complements recall by showing the accuracy of positive predictions.

**Security Metrics**

* Resistance to Attack Vectors: Evaluate the system's robustness against common attack vectors like shoulder surfing, smudge attacks, and brute force attempts. This can be assessed through simulated attacks and penetration testing.
* Entropy: Measures the unpredictability or randomness of the jumbled number pad patterns. Higher entropy indicates a greater level of security as it makes predicting the keypad layout more difficult for attackers.

**Usability Metrics**

* Authentication Time: The total time taken for a user to successfully authenticate. Shorter authentication times indicate better usability but must be balanced against security needs.
* Error Rate: The rate at which users make mistakes during authentication (e.g., pressing the wrong key). A lower error rate suggests a more user-friendly interface.

**Performance Metrics**

* Model Training Time: The time required to train the machine learning model, which impacts the feasibility of updates and retraining.
* Inference Time: The time taken for the model to make an authentication decision during actual use. Faster inference times lead to a smoother user experience.
* Resource Utilization: Measures the computational resources (CPU, memory) consumed by the system, which is especially important for mobile devices with limited resources.
* Scalability: The system's ability to handle a growing number of users without a significant drop in performance.

**Adaptability Metrics**

* Re-Training Frequency: The frequency with which the model needs retraining to adapt to changes in user behavior or to incorporate new data. Less frequent retraining with sustained accuracy is ideal.
* Sensitivity to User Behavior Variability: Assesses how well the system adapts to natural variations in user behavior over time, maintaining high accuracy without frequent manual recalibrations.