Password Inference via Wearable Devices Data Analysis

Ziqiang Ren

zichiang.ren@gmail.com

https://ziqiangren.github.io/

Outline

1 Backgrounds

2 System Design & Data collecting

3 Passwords Inference Based on DNN

4 Experimental Results & Analysis

5 Conclusion

Outline

1. Backgrounds
2. System Design & Data collecting
3. Passwords Inference Based on DNN
4. Experimental Results & Analysis
5. Conclusion

The popularity of wearable devices

- Smart wearable devices：

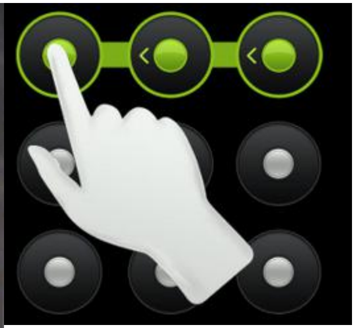
A portable device that is worn directly by the user or integrated into the user's clothing or accessories. Wearable devices' sensor can record a variety of user data, such as running speed, heart rate, body temperature and so on.

Data security problems

High-performance wearable devices show a new potential insecurity: if users input private information wearing high-performance devices, their hand movement may be caught by the sensor embed in the devices. Using the sensor data attacker could build an inference model to obtain the password or text sequence.

3 main scenarios

- Security system based on numeric keypad, eg. ATM
- Text input system based on QWERTY keyboard
- Security system based on smart phone lock screen pattern



Outline

1 Backgrounds

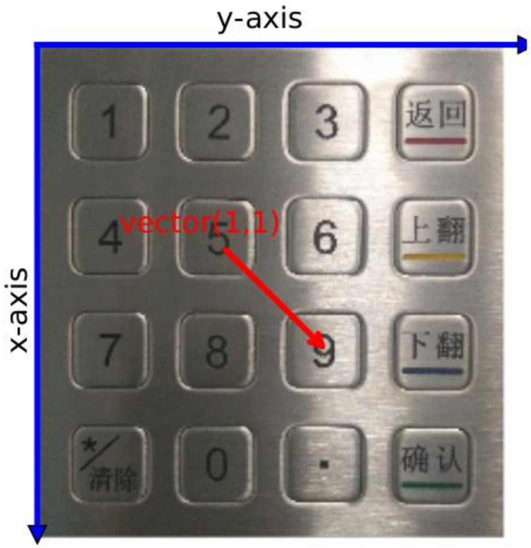
2 System Design & Data collecting

3 Passwords Inference Based on DNN

4 Experimental Results & Analysis

5 Conclusion

Motion vectors and corresponding labels



Vector	Label	R-label	Vector	Label	R-label
(0,0)	1	1	(0,1)	2	3
(0,2)	4	5	(1,0)	6	7
(1,1)	8	9	(1,-1)	10	11
(1,2)	12	13	(1,-2)	14	15
(2,0)	16	17	(2,1)	18	19
(2,2)	20	21	(2,-1)	22	23
(2,-2)	24	25	(3,0)	26	27
(3,1)	28	29	(3,-1)	30	31

Table 1: The motion vectors corresponding to their label and reverse label (R-label)

Acceleration data processing

1.Gravity filtering

2.Continuous data segmentation

3.Removal of fixed bias based on fragments

4.Smooth filtering

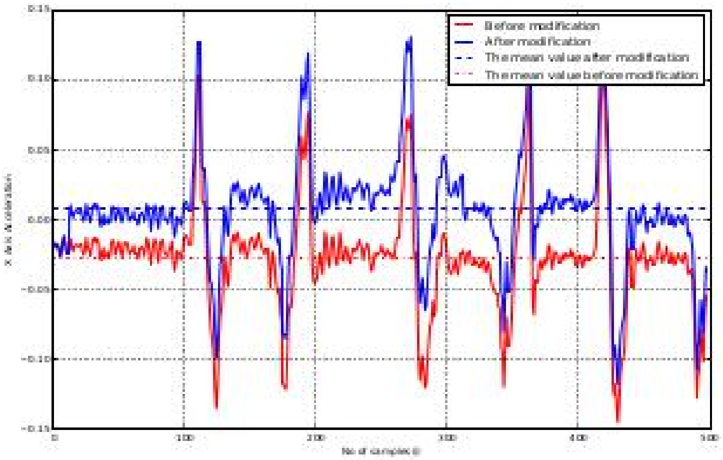


Fig. 5: A contrast before and after adaptive modification of fragment data.

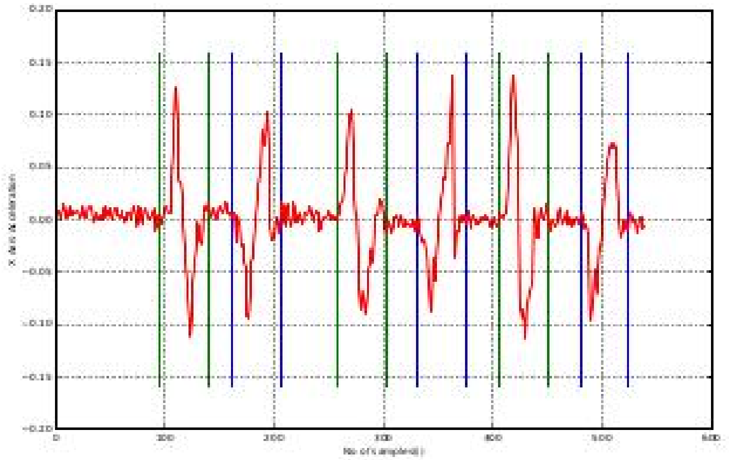
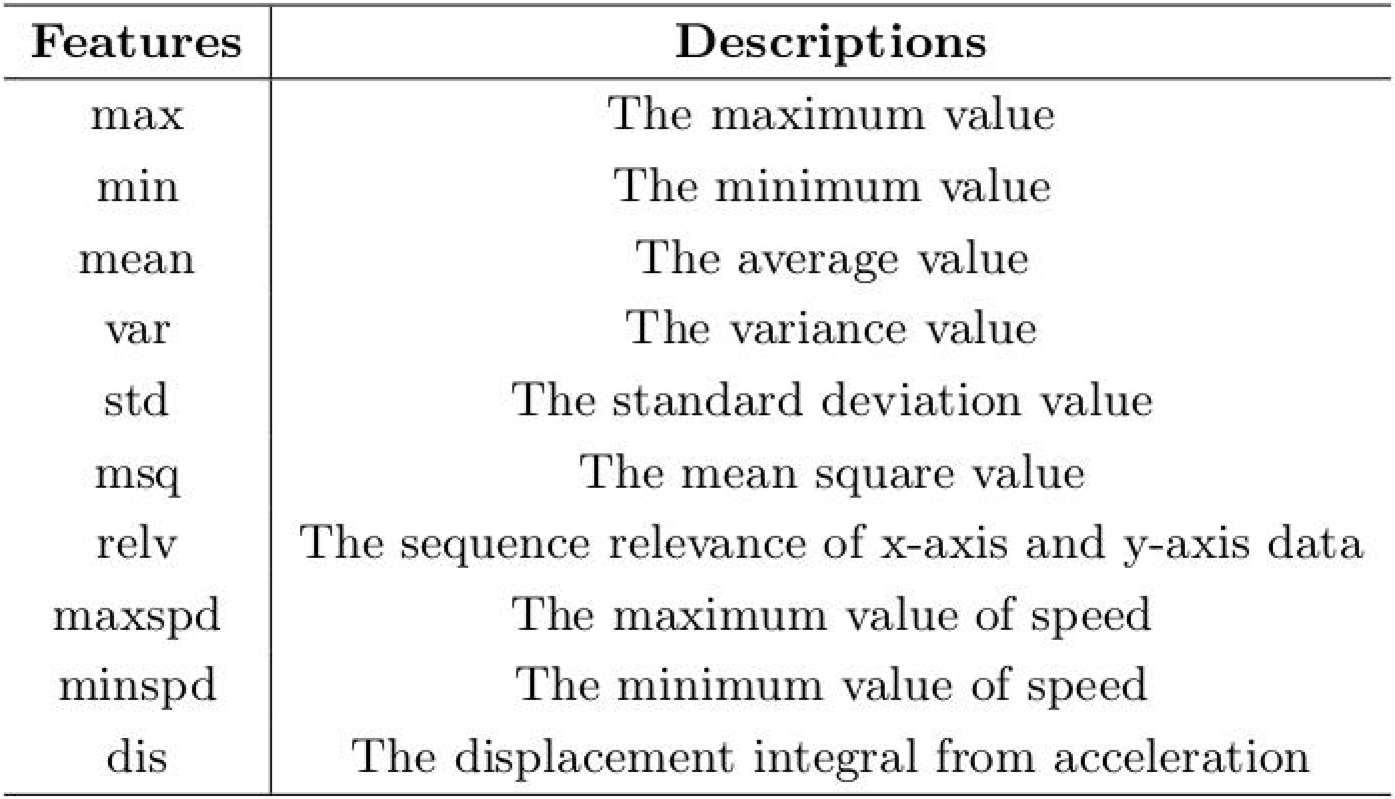


Fig. 6: Illustration: Acceleration data is divided into 6 segments.

Feature selection



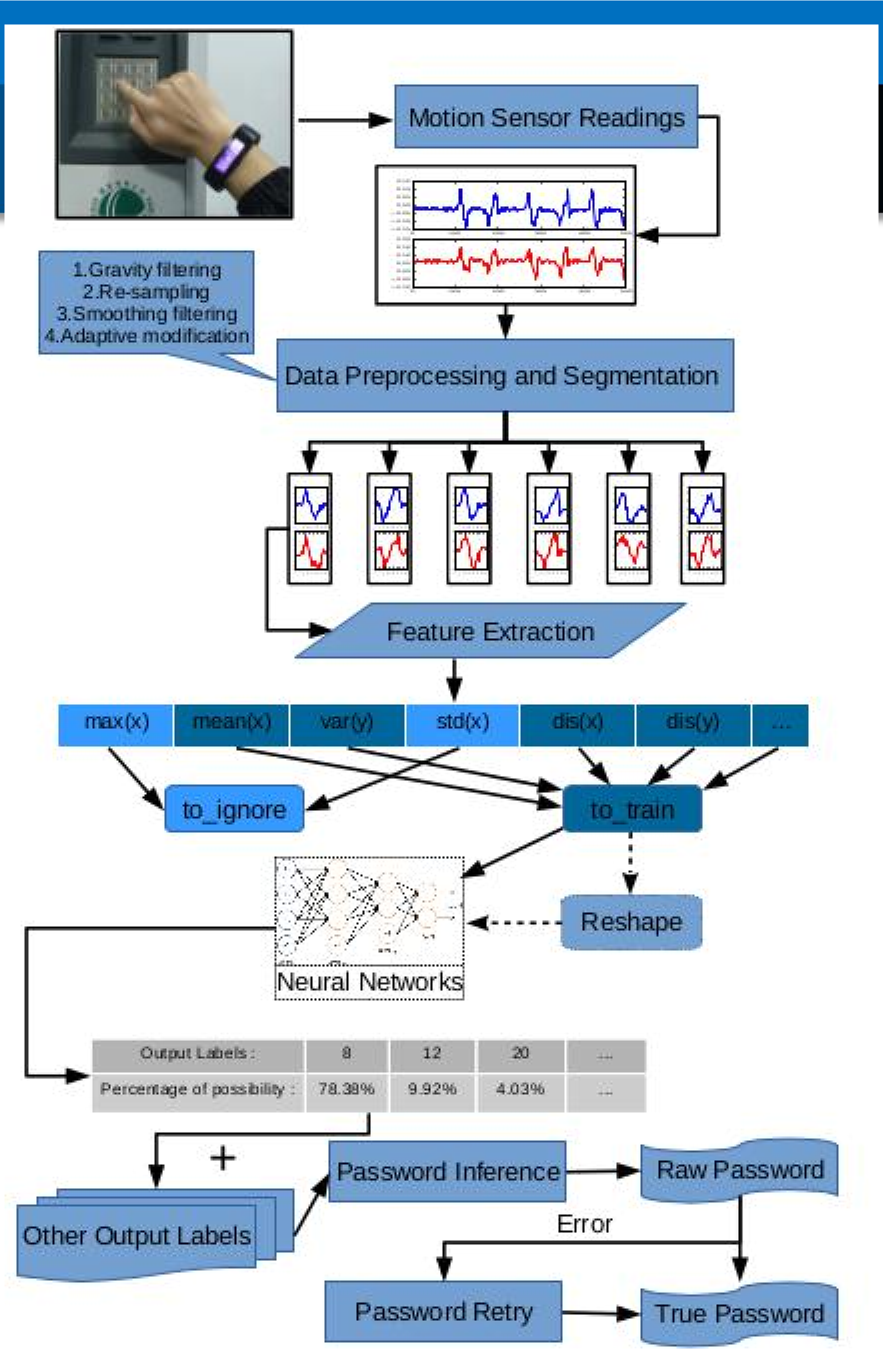
Features	Descriptions
max	The maximum value
min	The minimum value
mean	The average value
var	The variance value
std	The standard deviation value
msq	The mean square value
relv	The sequence relevance of x-axis and y-axis data
maxspd	The maximum value of speed
minspd	The minimum value of speed
dis	The displacement integral from acceleration

Table 2: Examples of 10 kinds of time domain features.

System Design

The model is DNN;

In the feature selection period, it is not necessary to select all the features; we set some features as "ignored features".



Output Labels :	8	12	20	...
Percentage of possibility :	78.38%	9.92%	4.03%	...

Outline

1. Backgrounds
2. System Design & Data collecting
3. Passwords Inference Based on DNN
4. Experimental Results & Analysis
5. Conclusion

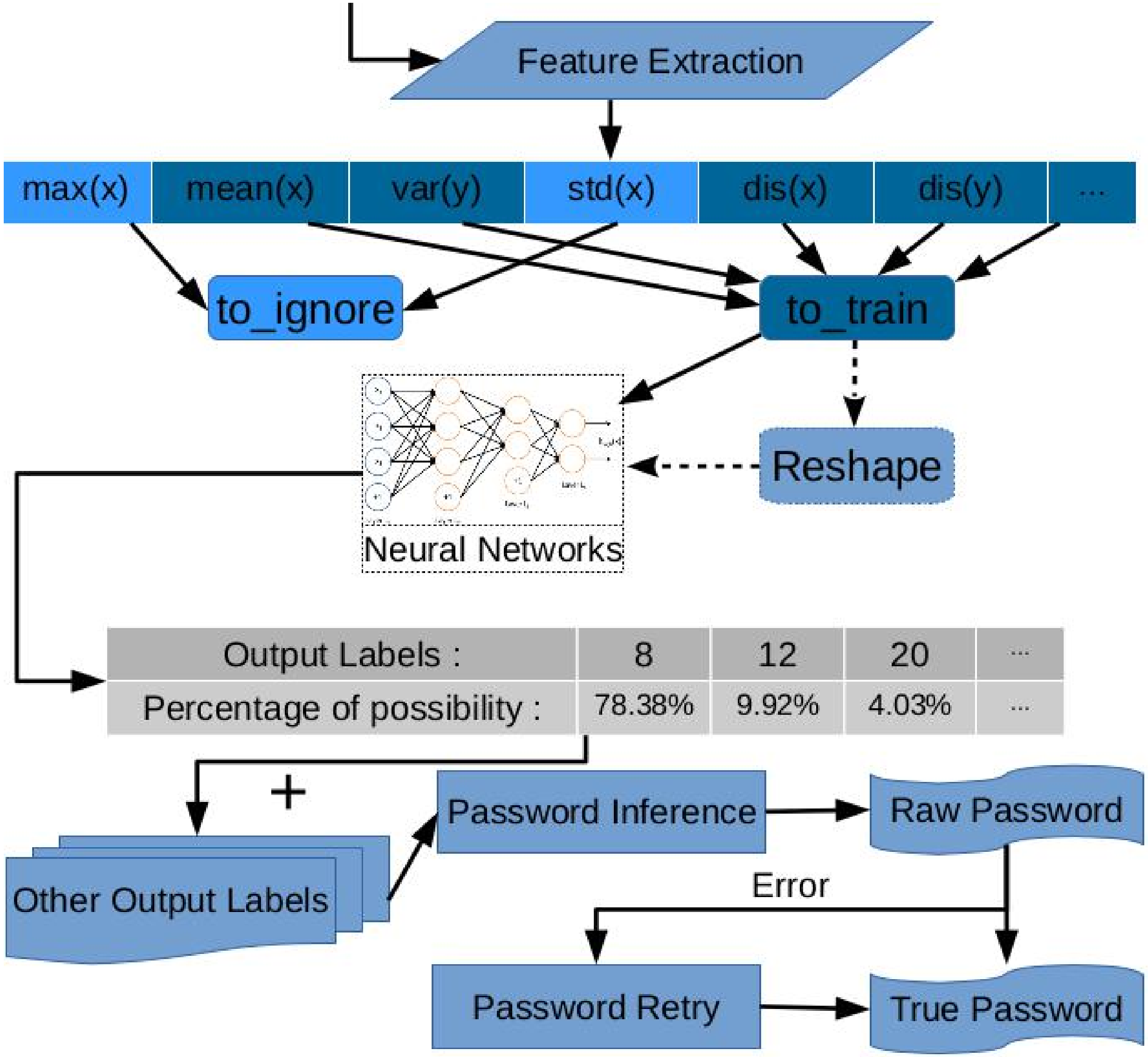
Building DNN

Based on TensorFlow1.0；

Simplified 4-layer neural network:

1. Input layer (feature layer), the final model identified as 7 nodes (corresponding to 7-dimensional features)
2. Full connection layer (32 nodes)
3. Full connection layer (32 nodes)
4. Output layer (31 nodes), corresponding to all 31 kinds of motion vector labels

DNN diagram & PIN backward inference



Feature Extraction

max(x)

mean(x)

var(y)

std(x)

dis(x)

dis(y)

...

to_ignore

to_train

Reshape

Neural Networks

Output Labels :	8	12	20	...
Percentage of possibility :	78.38%	9.92%	4.03%	...

+

Other Output Labels

Password Inference

Raw Password

Error

Password Retry

True Password

Error of Inference & Error Types

1.**non-out-of-range errors:**

no starting point or end point of any vector falls outside the keyboard range.

2.**out-of-range errors:**

A more common case of error. one or more of the inference results falls outside the keyboard range.

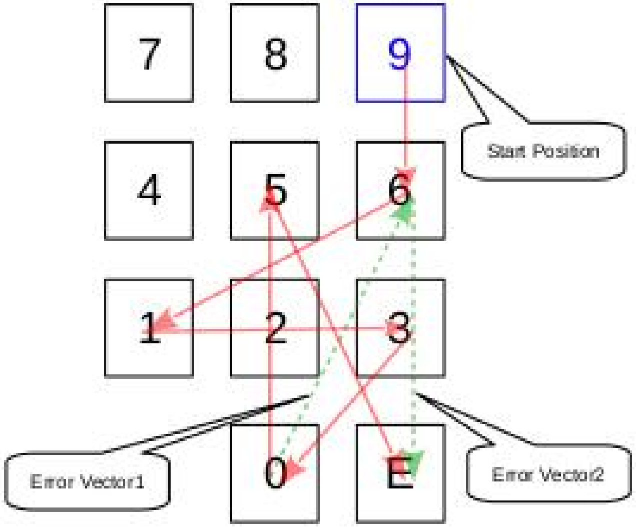


Fig. 7: A example of non-out-of-range error of inferred password as '961306' while the correct password is '961305'.

Automatic Password Retrying

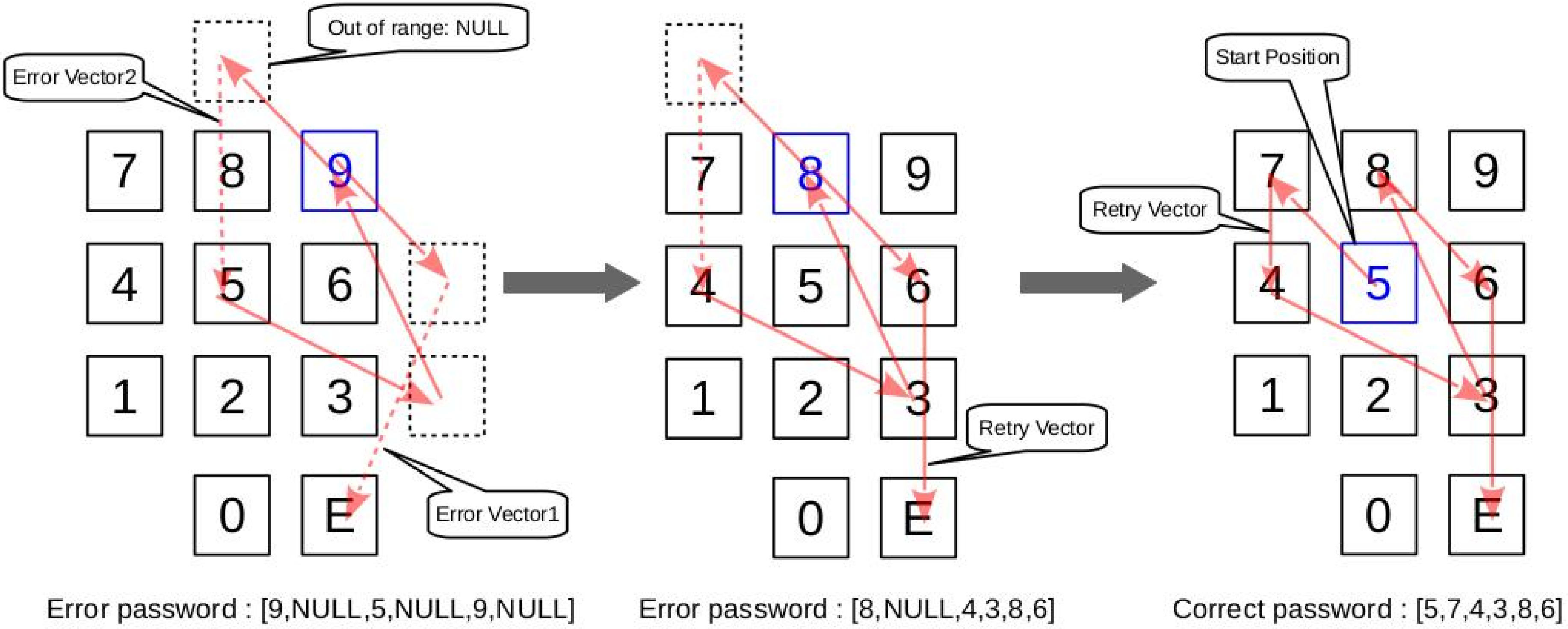


Fig. 8: Retry process diagram of a correct password of ‘574386’.

Outline

1. Backgrounds
2. System Design & Data collecting
3. Passwords Inference Based on DNN
4. Experimental Results & Analysis
5. Conclusion

Experimental description

- Datasets:

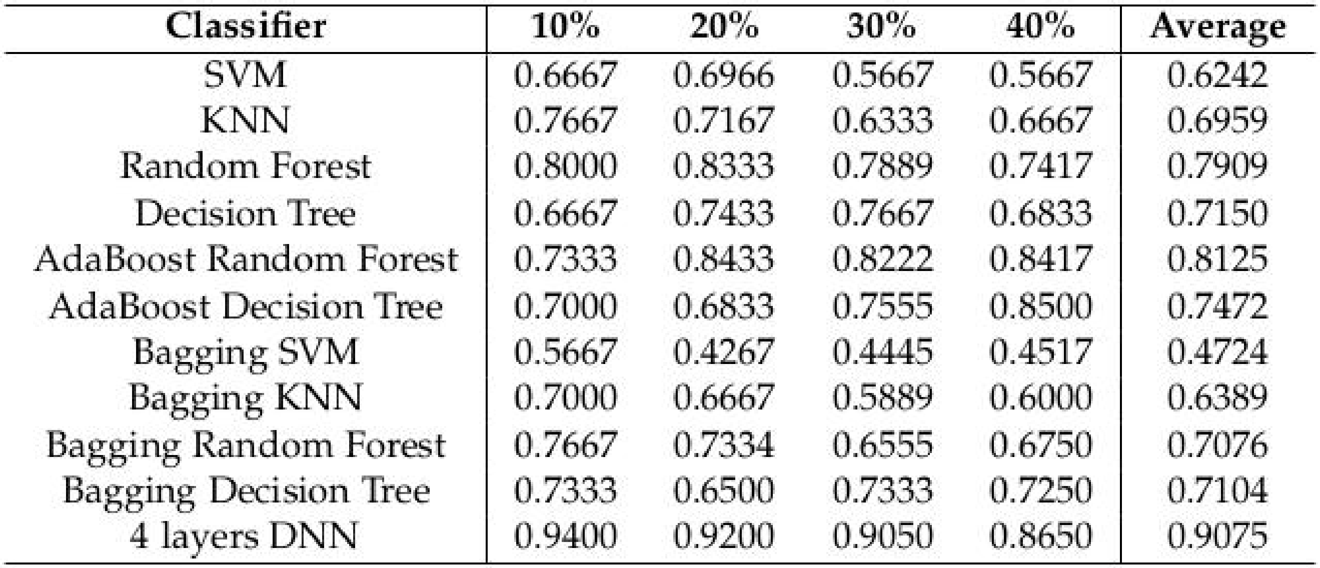
Wearable Device: Microsoft Band I
Each motion vector is sampled 20 times.
We add random Gaussian noise to double the number of samples.
Finally we get a total of 31*40 training data.
Randomly generate 100 groups of 6-digits password and input them, which means 100 sets of test data.

- Evaluation criteria:

•Cross validation accuracy (for single motion vector classification)
•Top-k accuracy (in the case of a complete 6-digits password inference case, the probability that the first k candidates have the correct password)

Cross validation accuracy under different testset ratios

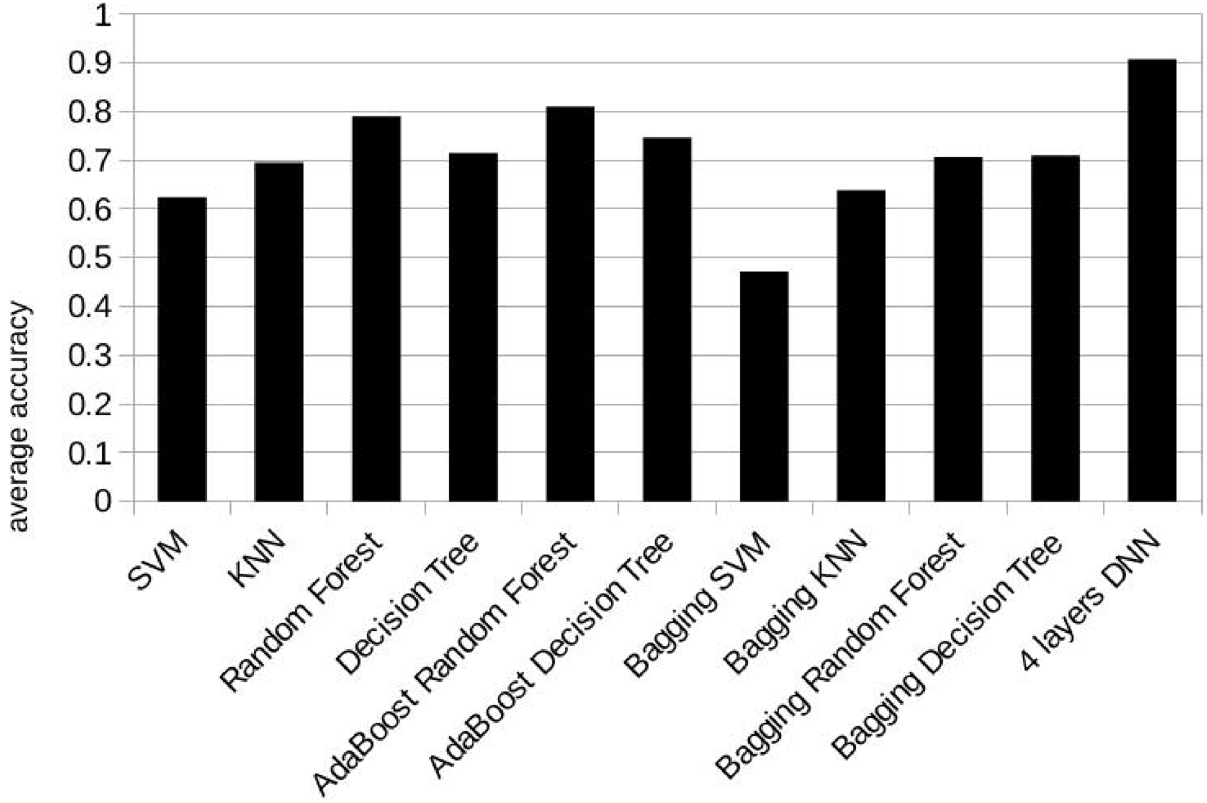
In addition to DNN, we use other classic classifiers to show performance comparisons.



Classifier	10%	20%	30%	40%	Average
SVM	0.6667	0.6966	0.5667	0.5667	0.6242
KNN	0.7667	0.7167	0.6333	0.6667	0.6959
Random Forest	0.8000	0.8333	0.7889	0.7417	0.7909
Decision Tree	0.6667	0.7433	0.7667	0.6833	0.7150
AdaBoost Random Forest	0.7333	0.8433	0.8222	0.8417	0.8125
AdaBoost Decision Tree	0.7000	0.6833	0.7555	0.8500	0.7472
Bagging SVM	0.5667	0.4267	0.4445	0.4517	0.4724
Bagging KNN	0.7000	0.6667	0.5889	0.6000	0.6389
Bagging Random Forest	0.7667	0.7334	0.6555	0.6750	0.7076
Bagging Decision Tree	0.7333	0.6500	0.7333	0.7250	0.7104
4 layers DNN	0.9400	0.9200	0.9050	0.8650	0.9075

TABLE 3: Comparison of Cross - Validation Accuracy of 11 Kinds of Different Classifiers..

Average cross validation accuracy



Top-k accuracy

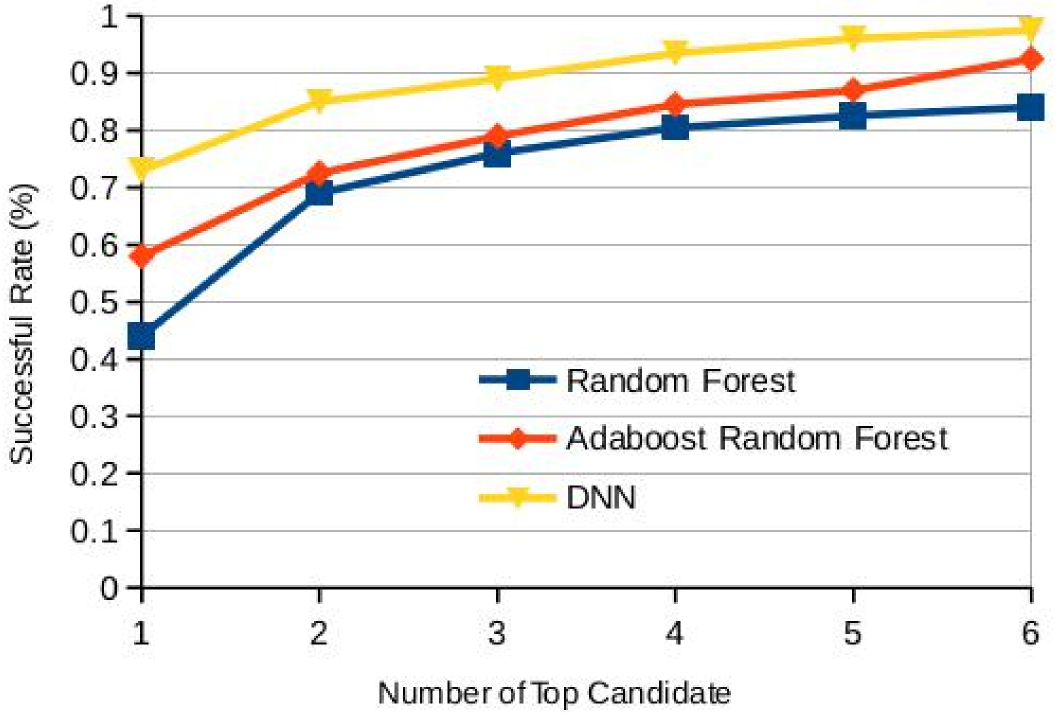
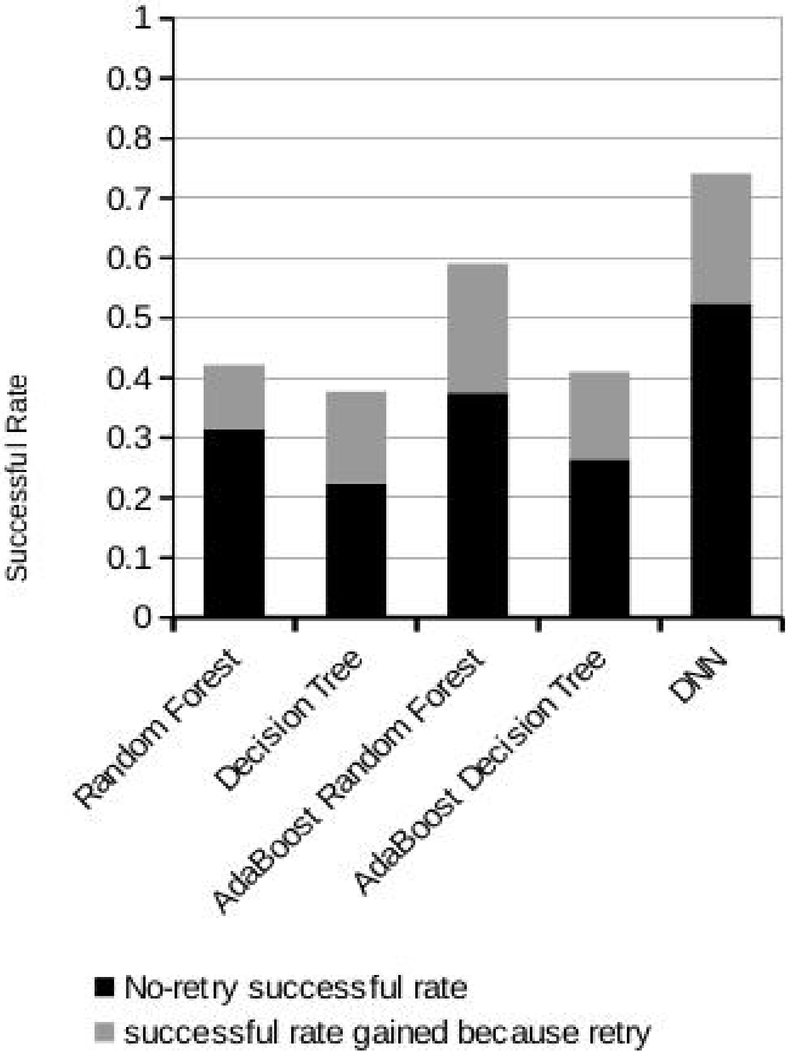


Fig. 11: Successful rate of recovering password sequence within top-k candidates.

Top-1 accuracy: the contribution of the automatic retrying

According to the figure, the elimination of out-of-range errors will bring a huge performance improvement to our password inference model.



Outline

1. Backgrounds
2. System Design & Data collecting
3. Passwords Inference Based on DNN
4. Experimental Results & Analysis
5. Conclusion

Conclusion

- **Password Inference System:**
- We present a novel user password inference system based on wearable device sensors, which includes data acquisition and processing, multi-feature extraction and fusion, multi-classifier performance comparison.

- **Automatic retrying algorithm:**
- We also present an automatic retrying algorithm that can correct one of the two types of errors. It increases the top-1 inference accuracy significantly .

Thanks！