project Approval no:
SSE/26/11/019-(1)

Name of the Student : A. Srinivas.
Reg No : 192211019
Department : computer science engineering.
Guide Name : Dr. V. Jaganathan.
Title : Experimental Analysis of SVM & RF for user Identification using keystroke dynamics.

1. Introduction:

— Approved —
12/03/25
(Dr. V. Jaganathan)

paragraph - 1:

* Research about: Comparision of support vector machine and randomforest for user identification using keystroke dynamics.
* Importance: Evaluating classification performance, accuracy and feature importance for keystroke-based authentication.
* Applications: Cybersecurity, user authentication, fraud prevention.

paragraph - 2:

* no. of related articles in last 5 years:
  - GoogleScholar - 52 articles.
  - ScienceDirect - 37 articles.
* most cited findings:

a. Killourhy & Maxion, IEEE, 2009 - SVM effective for biometric authentication. [Hal-science - 29]

b. Teh et al./ patteen recognition, 2013 - Random forest provides better feature selection. [Springer - 50]

c. Ahmed & Traore, IEEE Access, 2017 - SVM struggles with high-dimensional keystroke data. [Elsevier - 42]

- Best study: Teh et al./ patteen recognition, 2013 - Found RF outperforming SVM in user authentication tasks. [Springer - 50]

paragraph-3:

* Unanswered solution: No direct statistical comparision of SVM and RF on keystroke authentication.
* Aim: Compare classification accuracy and statistical significance of SVM vs RF for keystroke dynamics.

2. Materials and Methods:

paragraph-1:

* Study setting: SSE, SIMATS
* Ethical approval: N/A (No human samples directly involved).
* No. of groups: 2 (SVM, RF)
* Sample size/group: 10 each
* Total sample size: 20
* G-power: 80% (Kilovathy & makion, IEEE, 2009).

paragraph-2:

* Dataset source: Collected from CMU keystroke dynamics benchmark database.
* Feature extraction: Hold time, flight time, dwell time computed from keystroke events.

paragraph-3:

* preprocessing steps:
  - outlier detetion and removal using Z-score method.
  - Feature normalization using Min-Max Scaling.
  - Splitting data set into 80% training and 20% testing sets.

paragraph-4:

* Testing Setup:
  - Feature Extraction: python & Scikit-learn.
  - Model training: SVM vs Random Forest.
  - Evaluation metrics: Accuracy, precision, Recall, F1-score, ROC-AUC curve.

paragraph-5:

- Data collection method: Accuracy values recorded from multiple test runs and averaged for reliability.

paragraph - 6:

* Statistical software : spss
* Independent variables: model type (SVM vs RF)
* Dependent variables: classification accuracy.
* Analysis: Independent sample t-test, Descriptive statistics, Graphical representation (confusion matrix, Barchart).

3. Results and Discussion:

paragraph-1:

* Summary of results : RF has higher mean accuracy than SVM.

paragraph-2:

* Descriptive table: SVM (88.14 %), RF (93.38 %).
* Killourhy & Maxion, IEEE, 2009 - similar results observed in keystroke authentication.
* Teh et al. pattern recognition, 2013 - found RF reducing false positives in biometrics.
* Consensus: Random Forest performs significantly better.

paragraph-3:

- Statistically significance (Independent-sample t-test
  - $p < 0.05$, indicating significant difference
  - RF significantly outperforms SVM in keystroke classification.

paragraph-4:

* Graphical Representation of Results:

- Confusion matrix for both models to visualize misclassifications.
- Barchart comparing accuracy score of SVM & RF.
- ROC-AUC curve to show performance of both classifiers.

paragraph-5:

* limitations:

- small dataset, which may limit generalization.
- Randomforest might overfit, requiring further tuning.
- Keystroke timing variables due to user behaviour differences

paragraph-6:

* Future Scope:

- Expand dataset to test robustness.
- Experiment with deep learning models like CNN-LSTM.
- Improve real-time keystroke authentication using ensemble models.

4. Conclusion:

- within the limits of this study, Randomforest significantly outperforms SVM in user identification using keystroke dynamics ($p < 0.05$).

Title1:

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	SVM	10	88.7430	.65296	.20648
	RF	10	93.3870	.47237	.14938

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
ACCURACY	Equal variances assumed	1.342	.262	-18.222	18	<.001	<.001	-4.64400	.25485	-5.17942	-4.10858
	Equal variances not assumed			-18.222	16.395	<.001	<.001	-4.64400	.25485	-5.18320	-4.10480