Title: Poster: NATICUSdroid: A malware detection framework for Android using native and custom permissions

Authors: Akshay Mathur[a], Laxmi Mounika Podila[a], Keyur Kulkarni[a], Quamar Niyaz[b], Ahmad Y. Javaid[a]

[a]The University of Toledo, 2801 W Bancroft St, Toledo, OH 43606, USA
[b]Purdue University Northwest, 2200 169th St, Hammond, IN 46323, USA

Date: Available online 13 January 2021

Venue: Journal of Information Security and Applications, Volume 58, May 2021.

Abstract: The rapid growth of Android apps and its worldwide popularity in the smartphone market has made it an easy and accessible target for malware. In the past few years, the Android operating system (AOS) has been updated several times to fix various vulnerabilities. Unfortunately, malware apps have also upgraded and adapted to this evolution. The ever-increasing number of native AOS permissions and developers' ability to create custom permissions provide plenty of options to gain control over devices and private data. Therefore, newly created permissions could be of great importance in detecting current malware. Previous popular works on malware detection used apps collected during 2010–2012 to propose malware detection and classification methods. A majority of permissions used in those apps are not as widely used or do not exist anymore. In this work, we present a novel malware detection framework for Android called NATICUSdroid, which investigates and classifies benign and malware using statistically selected native and custom Android permissions as features for various machine learning (ML) classifiers. We analyze declared permissions in more than 29,000 benign and malware collected during 2010–2019 to identify the most significant permissions based on the trend. Subsequently, we collect these identified permissions that include both the native and custom permissions. Finally, we use feature selection techniques and evaluate eight ML algorithms for NATICUSdroid to distinguish benign apps from malware. Experimental results show that the Random Forest classifier-based model performed best with an accuracy of 97%, a false-positive rate of 3.32%, and an f-measure of 0.96.

Link to PDF (available until March 04, 2021): https://authors.elsevier.com/c/1cPOP7tT2CiC-L

DOI: https://doi.org/10.1016/j.jisa.2020.102696

Poster: NATICUSdroid: A malware detection framework for Android using native and custom permissions

Akshay Mathur [a], Laxmi Mounika Podila [a], Keyur Kulkarni [a], Quamar Niyaz [b], Ahmad Y. Javaid [a]

a - The University of Toledo, 2801 W Bancroft St, Toledo, OH 43606, USA

b - Purdue University Northwest, 2200 169th St, Hammond, IN 46323, USA

1. INTRODUCTION

- Malware infections increasing with popularity of Android devices
- Use of dated datasets and obsolete features in recent malware detection frameworks is alarming
- Need for a scalable system based on robust and significant features
- Permissions used as features before, but only "native" permissions are insufficient to classify good vs. bad

2. CONTRIBUTIONS

- Proposed and built android malware detection framework, NATICUSdroid (NATIve and CUStom permissions analysis for Android)
- Utilized native and custom (created by third-party app vendors) permissions of 29k+ apps
- Built additional baseline malware detection framework using native permissions
- Achieved better results compared to the state-of-the-art techniques
- Explained achieved results leveraging XAI (eXplainable Artificial Intelligence) [7].

3b. METHODOLOGY

Application Database:

- *Benign Apps:*
  - -- 14630 API level 23+ apps from Androzoo,
  - -- rated benign by VirusTotal
- *Malware Apps:*
  - -- 14700 apps from Arguslab Android Malware Dataset (AMD)

Feature extraction and dataset generation:

- Extracted permissions using Androguard
- Generated two datasets:
  - -- only native permissions (*Native*)
  - -- native + custom permissions (*Naticus*)

Feature Selection:

- *Frequency Counting*
  - -- Permission occurences in apps
- *Backward Elimination*
  - -- Insignificant permissions removed
- *Multicollinearity Removal*
  - -- Only one of highly correlated permissions kept

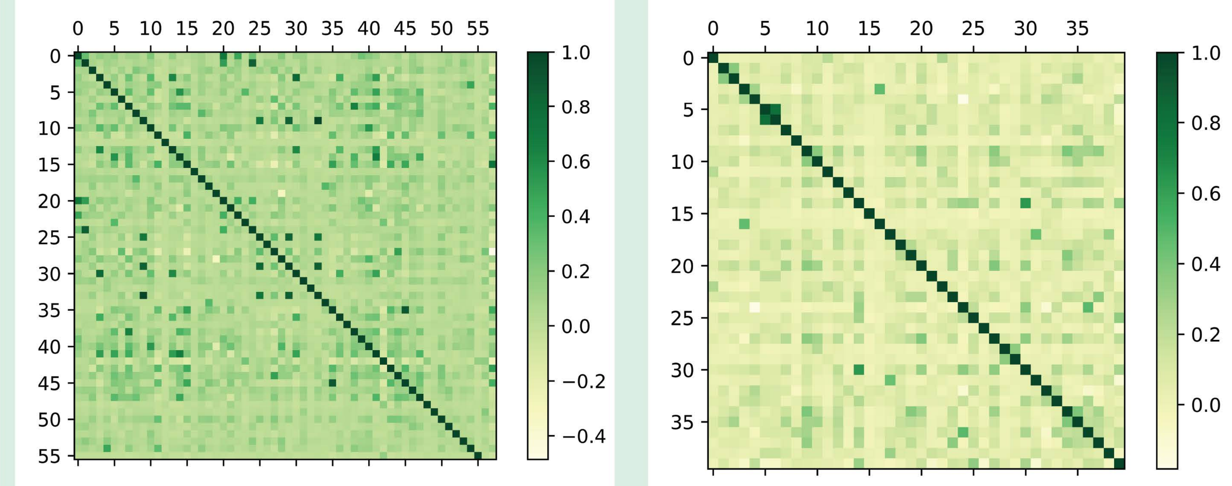
Step	Naticus permissions	Native permissions
Feature extraction	6761	325
Permission frequency counting	86	52
Backward elimination	58	39
Collinearity check	55	39

Permissions remaining after each selection step in the datasets

3c. CLASSIFICATION

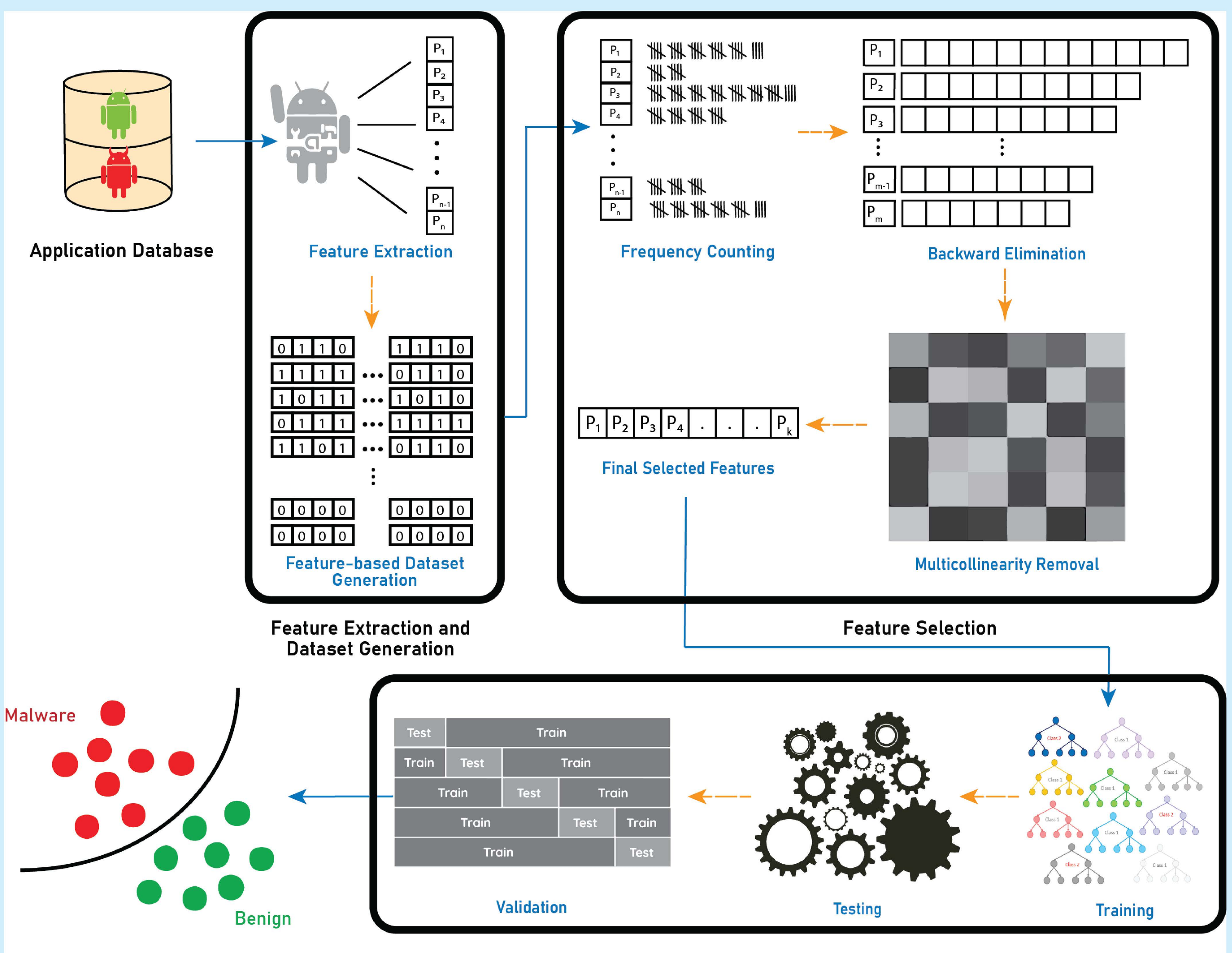
- Single Learners: Logistic Regression (LR), k- Nearest Neighbtor (KN), Support Vector Machines (SVM)
- Ensemble Learners: Random Forests (RF), Extra Trees (ET), XGBoost (XG), AdaBoosting (AB), Bagging (BG)
- Metrics: Accuracy + F-Score, Training + Detection Time, ROC Curves

4a. EXPERIMENTAL RESULTS



Correlation Heatmaps of the two datasets after Feature Selection

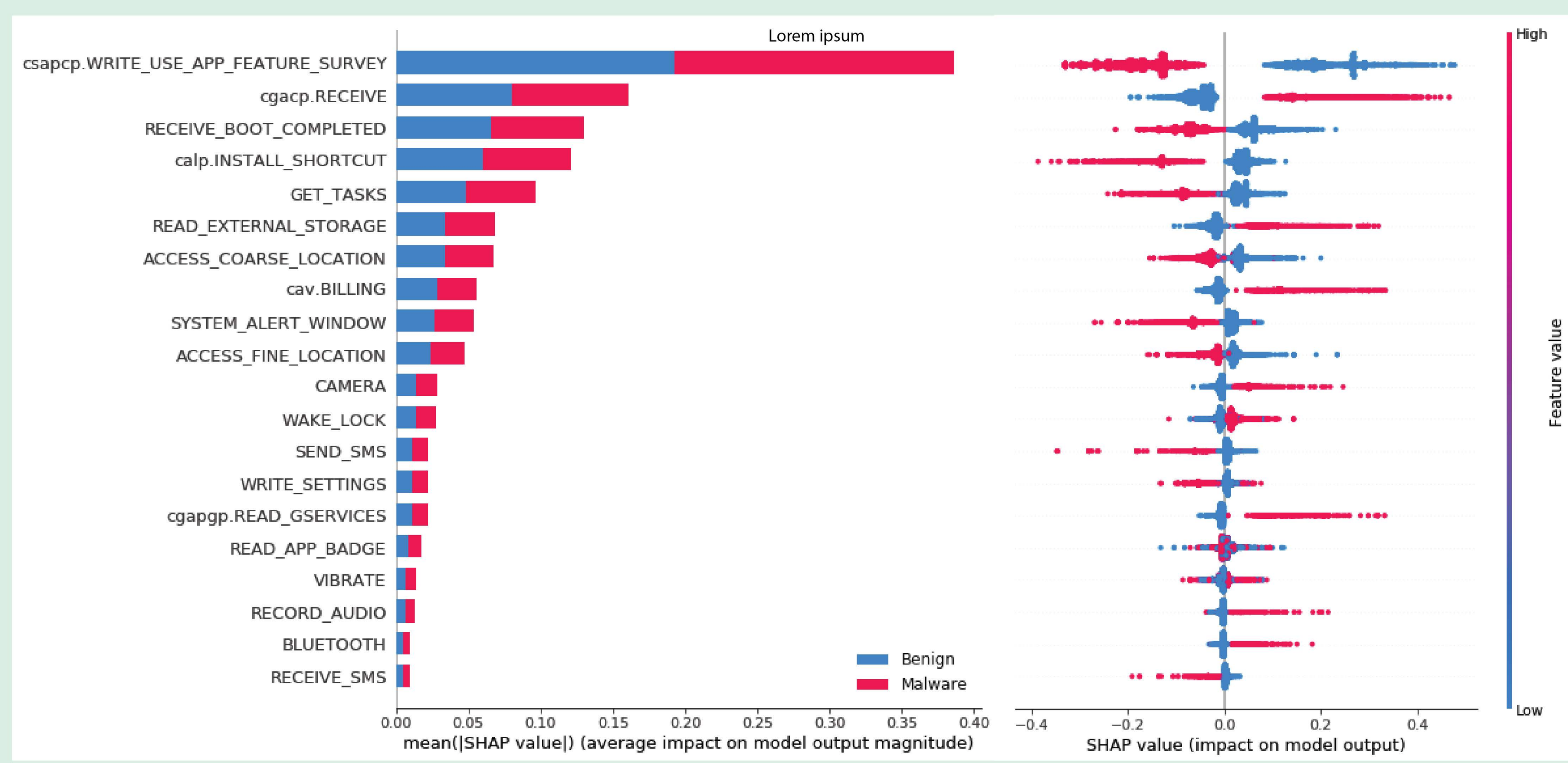
3a. NATICUSdroid WORKFLOW



4b. EXPERIMENTAL RESULTS

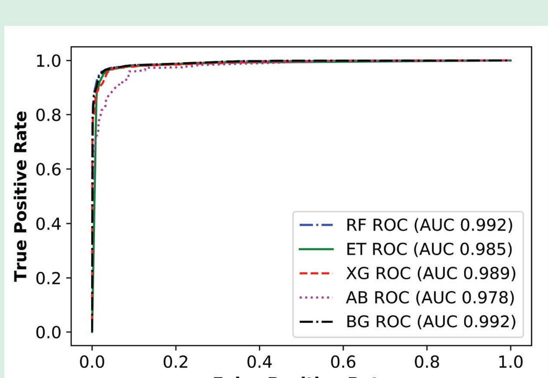
Classifier	Validation accuracy (%)		Detection accuracy (%)		F-Score		Training time (s)		Detection time (s)	
	Naticus	*Native*	*Naticus*	*Native*	*Naticus*	*Native*	*Naticus*	*Native*	*Naticus*	*Native*
KN	96.13	84.85	96.13	84.65	0.9617	0.8742	0.75	0.98	10.91	5.04
SVM	95.31	80.79	95.32	81.06	0.9537	0.8518	34.31	165.96	1.46	6.30
LR	95.93	77.75	95.95	77.9	0.9598	0.8158	0.09	0.08	0.01	0.001
RF	**97.10**	**86.03**	**96.95**	**85.98**	**0.9662**	**0.8835**	**0.17**	**0.12**	**0.11**	**0.11**
ET	96.45	85.06	96.49	84.67	0.9650	0.8704	0.13	0.12	0.11	0.11
XG	96.02	82.85	96.17	82.95	0.9620	0.8635	0.68	0.69	0.02	0.01
AB	92.87	77.34	92.18	77.05	0.9225	0.8378	1.27	1.38	0.15	0.15
BG	96.49	85.81	96.58	85.84	0.9659	0.8817	24.35	14.09	2.34	1.68

Classification Results for *Naticus* and *Native* datasets

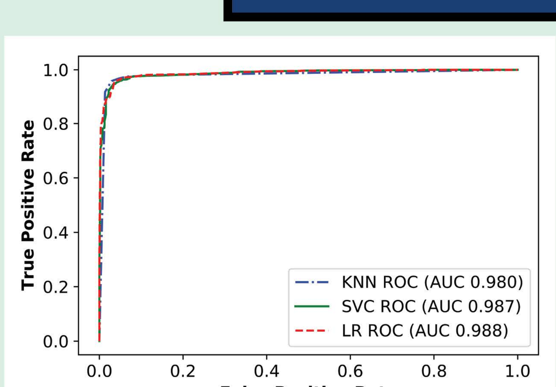


Feature Importance in Train and Test sets of *Naticus* Dataset

4c. EXPERIMENTAL RESULTS

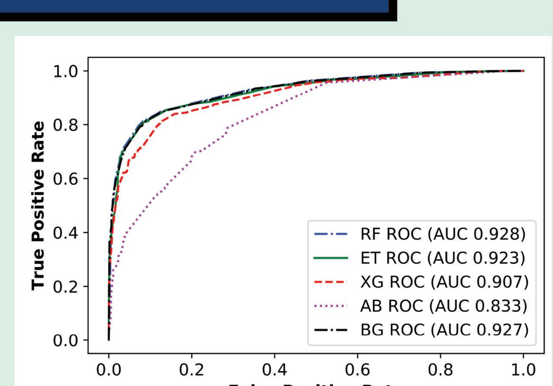


Single Learners

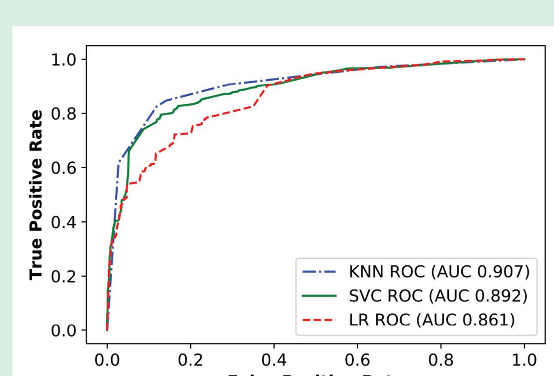


Ensemble Learners

ROC Curves for *Naticus* Dataset



Single Learners



Ensemble Learners

ROC Curves for *Native* Dataset

Work	Dataset		CP	Accuracy	FPR	DT
	Year	*Apps*				
PMDS [1]	2010–12	2950	✗	92 - 94	1.52–3.93	✗
ApkAuditer [2]	2010–12	8762	✗	88	✗	✗
CFG based detection [3]	2017	20 693	✗	98.8	2.9–9.1	✗
System calls and LSTM based detection [4]	2010–17	7005	✗	93.4	9.3	1 s
Drebin [5]	2010–12	129 013	✗	93	1	0.75 s
Signature and Heuristic based detection [6]	2015	401	✗	85	6.45	85 s
NATICUSdroid	**2010-19**	**29 330**	✔	**96.95**	**3.32**	**0.11 s**

Comparison of *NATICUSdroid* with state-of-the-art

5. CONCLUSION

- NATICUSDroid exhibits accuracy of 96.9% and F-Score of 0.97 in 0.11 seconds in detecting test dataset malware
- Capable of regular updates with newer native and custom permissions from newer Android versions

REFERENCES

[1] Rovelli Paolo, Vigfússon Ýmir. Pmds: Permission-based malware detection system. In: International conference on information systems security. Springer; 2014, p. 338–57.

[2] Talha Kabakus Abdullah, Alper Dogru Ibrahim, Aydin Cetin. Apk auditor: Permission-based android malware detection system. Digit Investigation 2015;13:1–14.

[3] Ma Zhuo, Ge Haoran, Liu Yang, Zhao Meng, Ma Jianfeng. A combination method for android malware detection based on control flow graphs and machine learning algorithms. IEEE Access 2019;7:21235–45.

[4] Xiao Xi, Zhang Shaofeng, Mercaldo Francesco, Hu Guangwu, Sangaiah Arun Kumar. Android malware detection based on system call sequences and LSTM. Multimedia Tools Appl 2019;78(4):3979–99.

[5] Arp Daniel, Spreitzenbarth Michael, Hubner Malte, Gascon Hugo, Rieck Konrad, Siemens CERT. Drebin: Effective and explainable detection of android malware in your pocket. In: Ndss, Vol. 14. 2014, p. 23–6.

[6] Rehman Zahoor-Ur, Khan Sidra Nasim, Muhammad Khan, Lee Jong Weon, Lv Zhihan, Baik Sung Wook, Shah Peer Azmat, Awan Khalid, Mehmood Irfan. Machine learning-assisted signature and heuristic-based detection of malwares in android devices. Comput Electr Eng 2018;69:828–41.

[7] Lundberg Scott M, Lee Su-In. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in neural information processing systems, Vol. 30. Curran Associates, Inc.; 2017, p. 4765–74