Požymių konvertavimo į vaizdus metodų palyginimas kenkėjiškų programų aptikimo efektyvumui gerinti

Matas Lukšys, Viktor Medvedev

Vilniaus universitetas, Matematikos informatikos fakultetas,
Duomenų mokslo ir skaitmeninių technologijų institutas,
Akademijos g. 4, LT-08412 Vilnius, Lietuva
matas.luksys@mif.stud.vu.lt

Santrauka. Straipsnyje palyginami vykdomųjų PE failų požymių konvertavimo į vaizdus metodai. Vykdomųjų failų požymiai gaunami iš kenkėjiškų ir saugių vykdomųjų PE failų, pateiktų *PE Malware Machine Learning Dataset* duomenų aibėje. Požymių atrankai buvo naudojami pagrindiniai vykdomųjų failų struktūriniai ir elgsenos požymiai, tokie kaip antraštės, sekcijų statistika, dydžio ir entropijos parametrai. Eksperimentiniai tyrimai atlikti naudojant tiesioginius požymių konvertavimo metodus – BIE ir HSV bei netiesioginius konvertavimo metodus, tokius kaip cBIE, IGTD ir LMIGTD. Gauti vaizdai buvo klasifikuojami naudojant konvoliucinį neuroninį tinklą. Eksperimentų rezultatai parodė, kad netiesioginiai metodai leidžia pasiekti aukštesnį kenkėjiškų programų aptikimo tikslumą.

Raktiniai žodžiai: vykdomasis failas, požymių konvertavimas, vaizdų klasifikavimas, mašininis mokymasis, kenkėjiškų programų aptikimas.

1 Įvadas

Kenkėjiškos programos išlieka viena didžiausių grėsmių šiuolaikinėms informacinėms sistemoms, nepaisant nuolat tobulėjančių kibernetinio saugumo priemonių. 2023 metais pasaulyje užfiksuota daugiau nei 1 milijardas kenkėjiškų programų pavyzdžių [1]. Remiantis ankstesnėmis prognozėmis, 2025 metais kibernetinių nusikaltimų žala pasaulio ekonomikai gali pasiekti iki 10,5 trilijono JAV dolerių [2]. Šie skaičiai atspindi ne tik augantį kenkėjiškų programų kiekį, bet ir didėjančią jų įvairovę bei sudėtingumą.

Kenkėjiškos programos nuolat evoliucionuoja, pasitelkia pažangias maskavimo, kodavimo ir aplinkos analizės technikas, todėl tradiciniai aptikimo metodai tampa vis mažiau veiksmingi. Norint efektyviai aptikti naujas ir modifikuotas grėsmes, vis plačiau taikomi mašininio mokymosi metodai. Viena iš inovatyviausių krypčių – vykdomųjų failų požymių konvertavimas į

DOI: https://doi.org/10.15388/LMITT.2025.17

į vaizdus [3], leidžiantis išnaudoti konvoliucinių neuroninių tinklų gebėjimą atpažinti sudėtingus struktūrinius ir elgsenos modelius.

Šiame tyrime lyginami skirtingi vykdomųjų failų požymių konvertavimo į vaizdus metodai, vertinant jų efektyvumą kenkėjiškų programų aptikimui. Analizuojami tiek tiesioginiai, tiek netiesioginiai konvertavimo būdai, siekiant nustatyti, kurie metodai yra tinkamiausi vykdomųjų failų požymių paruošimui mašininio mokymosi algoritmams. Tyrimo rezultatai gali prisidėti prie pažangesnių kenkėjiškų programų aptikimo sistemų kūrimo ir padėti geriau suprasti šiuolaikinių grėsmių klasifikavimo principus.

2 Duomenų rinkiniai

Atliekant tyrimą, buvo išanalizuoti ir palyginti du duomenų rinkiniai, tinkami kenkėjiškų programų aptikimo uždaviniui spręsti: *PE Malware Machine Learning Dataset*[1] ir *MABEL 2.0 Dataset*[2]. Jų pagrindinės charakteristikos, privalumai bei ribojimai pateikti 1 lentelėje. Palyginimo rezultatai parodė, kad *PE Malware Machine Learning Dataset* rinkinys atitinka darbo tikslus – jis skirtas dvejetainiam klasifikavimui ir suteikia galimybę dirbti su neapdorotais vykdomaisiais PE failais (angl. *portable executable file*), todėl galima savarankiškai išgauti bei analizuoti įvairius struktūrinius požymius, būtinus efektyviam kenkėjiškų programų aptikimui [4]. Priešingai, *MABEL 2.0 Dataset* orientuotas į daugialypę kenkėjiškų programų šeimų klasifikaciją, neturi saugių failų pavyzdžių ir nepateikia neapdorotų vykdomųjų failų [5]. Šis rinkinys nėra tinkamas dvejetainiam klasifikavimui ar požymių inžinerijai, paremtai failų struktūra.

Remiantis šiuo palyginimu, tolimesniems eksperimentams pasirinktas *PE Malware Machine Learning Dataset*. Šis rinkinys leidžia išgauti vykdomųjų PE failų požymius, tokius kaip antraštės, sekcijų statistika, importų ir eksporto lentelės, failo ilgis ir entropija. Tyrime naudoti 54 skirtingi požymiai, apimantys pagrindinius struktūrinius ir elgsenos parametrus, kurie yra reikšmingi klasifikacijai. Nebuvo naudojami detalūs sekcijų dydžiai, importų sąrašai, resursų detalės ir kiti mažiau svarbūs atributai, nes jie nėra kritiški bendram klasifikavimui ir gali apsunkinti analizę.

---

1 https://practicalsecurityanalytics.com/
2 https://github.com/action-ai-institute/MABEL-dataset

1 lentelė. Duomenų rinkinių palyginimas.

Charakteristika	PE Malware Machine Learning Dataset	MABEL 2.0 Dataset
Neapdoroti failai	Pateikiami	Nepateikiami
Pavyzdžių skaičius	201 549 unikalūs failai (86 812 saugūs, 114 737 kenkėjiški)	90 414 unikalūs kenkėjiški failai
Klasių struktūra	2 klasės: kenkėjiški failai ir nekenkėjiški failai	400+ kenkėjiškų šeimų, nėra saugių pavyzdžių
Metaduomenys	Apima failo identifikavimo reikšmes, entropijos reikšmes, failo tipą, antivirusinių programų aptikimus ir pateikimo datą	Išsamūs duomenys apie failo architektūrą, sekcijas, išskaidytą kodą ir importuojamas bibliotekas
Tinkamumas klasifikavimui	Pritaikytas binariniam klasifikavimui	Pritaikytas viruso šeimų klasifikavimui
Apribojimai	Negalima nustatyti kenksmingo failo šeimos	Nepateikia vykdomųjų failų tiesioginei analizei; nėra saugių failų binariniam klasifikavimui; pateikiami išgauti failo požymiai

Analizuojamame duomenų rinkinyje yra daugiau nei 200 tūkstančių failų, suskirstytų į dvi klases: kenkėjiški ir saugūs (nekenksmingi). Dėl techninių resursų apribojimų nebuvo galimybės apdoroti visų įrašų, todėl siekiant užtikrinti rezultatų patikimumą ir pakankamą eksperimentų apimtį, buvo sudarytos subalansuotos imtys. Modelio mokymui naudota 4259 kenkėjiškų ir 4254 saugių failų imtis, o validavimui – atitinkamai 1064 ir 1063 failai. Tokia atranka leidžia objektyviai įvertinti skirtingų požymių konvertavimo metodų efektyvumą sprendžiant binarinį klasifikavimo uždavinį.

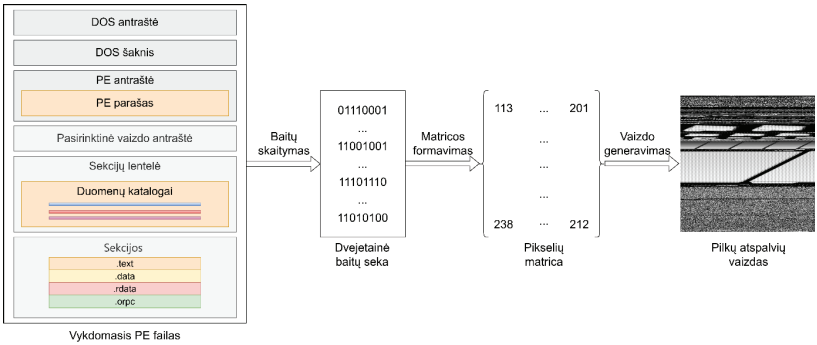
3 Metodika

Tyrime buvo lyginami požymių konvertavimo į vaizdus metodai, kurie skirstomi į tiesioginius ir netiesioginius. Tiesioginiai metodai leidžia atkurti pradinius baitus iš vaizdo, nes kiekvienas pikselis tiesiogiai atitinka konkretų baitą. Netiesioginiai metodai taiko optimizacijas ar klasterizavimą, todėl atstatyti pradinius duomenis iš vaizdo nebeįmanoma.

Tiesioginio dvejetainio kodavimo metodas (angl. *binary image encoding, BIE*) yra vienas paprasčiausių – kiekvienas failo baitas paverčiamas pilkos spalvos pikseliu ir išdėstomas kvadratinėje matricoje (žr. 1 pav.). Tokiu būdu išlaikoma bazinė failo struktūra, tačiau prarandama dalis semantinės infor-

macijos, nes neišryškinami sudėtingesni požymių tarpusavio ryšiai [6]. BIE metodas yra universalus ir dažnai naudojamas kaip atskaitos taškas vertinant sudėtingesnius metodus [6]. BIE metodo rezultato pavyzdys pateikiamas 2 paveiksle viduryje.

Kitas tiesioginio konvertavimo metodas – HSV (angl. *hue, saturation, value*) metodas. Jame požymiai normalizuojami ir priskiriami HSV spalvų kanalams [7], taip suteikiant papildomos informacijos apie požymių pasiskirstymą ir leidžiant vizualiai išskirti skirtingus požymių tipus. Literatūroje pabrėžiama [7], kad spalvų erdvės pasirinkimas gali reikšmingai paveikti klasifikavimo tikslumą. HSV metodo rezultato pavyzdys pateikiamas 2 paveiksle kairėje.



1 pav. Dvejetainio failo konvertavimo į vaizdą proceso schema.

Netiesioginio konvertavimo metodas IGTD *(*angl*. image generation for tabular data)* optimizuoja požymių išdėstymą vaizde pagal jų tarpusavio panašumus, apskaičiuojant porinius atstumus tarp požymių [6]. Tokiu būdu vaizde susiformuoja požymių klasteriai, išryškinantys jų tarpusavio ryšius. IGTD nereikalauja srities žinių ir leidžia CNN modeliams geriau išnaudoti erdvinius duomenų požymius. IGTD metodo rezultato pavyzdys pateikiamas 2 paveiksle dešinėje.

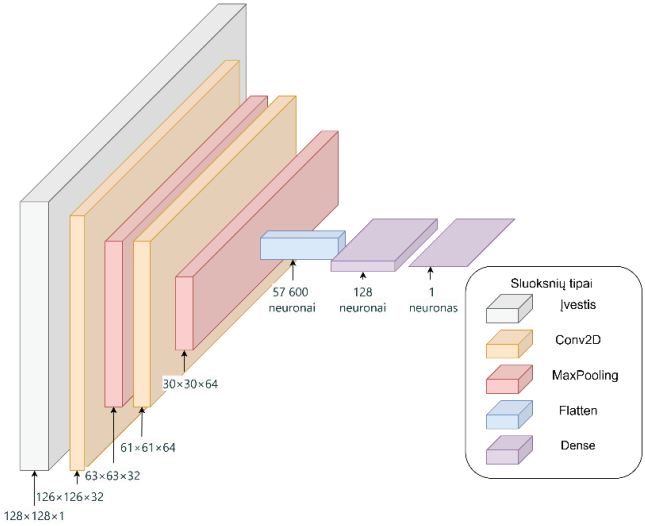
LMIGTD *(*angl*. localized and modified IGTD)* metodas yra IGTD metodikos patobulinimas, kuriame požymiai automatiškai grupuojami į funkcinius regionus naudojant klasterizavimo algoritmus. Šis metodas sumažina triukšmą ir išryškina susijusius failo elementus, nes kiekvienas regionas atspindi funkciniu požiūriu susijusių požymių grupę. Tokia vizualizacija padeda CNN modeliams lengviau atpažinti sudėtingas struktūras ir pagerina klasifikavimo tikslumą.

cBIE *(*angl*. cluster-based binary image encoding)* metodas yra BIE plėtinys [6], kuriame požymiai papildomai pertvarkomi pagal jų tarpusavio koreliacijas ar panašumus, dažniausiai taikant hierarchinį klasterizavimą. Tokiu būdu vaizde išryškėja funkcinės zonos – panašūs požymiai atsiduria greta, o skirtingi – atskiruose regionuose. Tai leidžia CNN efektyviau išnaudoti požymių tarpusavio ryšius ir dažnai pagerina klasifikavimo rezultatus sudėtinguose duomenų rinkiniuose.

2 pav. Skirtingų požymių konvertavimo į vaizdus metodų sugeneruoti kenkėjiškų failų vaizdai: HSV (kairėje), BIE (viduryje), IGTD (dešinėje).

Konvoliucinis neuroninis tinklas (angl. *convolutional neural network*) yra giliojo mokymosi modelis, skirtas vaizdų analizei ir klasifikavimui. Šio modelio pagrindinė savybė – gebėjimas išmokti ir atpažinti svarbius vaizdo požymius [8]. Konvoliucinio neuroninio tinklo architektūra sudaryta iš kelių specializuotų sluoksnių [8], kurie leidžia automatiškai išmokti ir atpažinti vaizdo požymius. Toks sluoksnių išdėstymas (žr. 3 pav.) leidžia efektyviai apdoroti vaizdus ir automatiškai išskirti požymius, reikalingus klasifikavimu [8].

Modelio įvestis – 128×128 pikselių pilkos spalvos vaizdas. Pirmasis konvoliucinis sluoksnis (angl. *Conv2D*) su 32 filtrais aptinka bazinius vaizdo požymius. Po jo taikomas maksimalios sutelkties sluoksnis (angl. *max pooling*), kuris sumažina duomenų apimtį ir išlaiko naudingą informaciją. Antrasis konvoliucinis sluoksnis su 64 filtrais leidžia išskirti sudėtingesnius požymius, o antrasis maksimalios sutelkties sluoksnis dar labiau sumažina matmenis. Išlyginimo sluoksnis (angl. *flatten*) paverčia duomenis į vienmatį vektorių, kuris perduodamas tankiajam sluoksniui (angl. *dense*) su 128 neuronais. Galiausiai, išvesties sluoksnis (angl. *output*) su vienu neuronu pateikia galutinę prognozę – ar failas yra kenkėjiškas, ar saugus.



3 pav. Konvoliucinio neuroninio tinklo architektūra kenksmingų programų klasifikavimui.

4 Rezultatai

Tyrimo metu buvo palyginti penki požymių konvertavimo į vaizdus metodai, siekiant įvertinti jų tinkamumą kenkėjiškų programų aptikimui taikant konvoliucinį neuroninį tinklą. Kiekvieno metodo efektyvumas vertintas pagal pagrindines klasifikavimo metrikas: bendrą tikslumą (angl. *accuracy*), kenkėjiškų failų atkūrimą (angl. *recall*), preciziškumą (angl. *precision*) ir F1 rodiklį. Praktikoje vienas svarbiausių rodiklių yra klaidingai neigiamų rezultatų (angl. *false negatives*) skaičius, t. y. kiek kenkėjiškų failų sistema neaptiko ir priskyrė prie saugių. Šis skaičius pateikiamas 2 lentelėje. Sistemos gebėjimą aptikti kenkėjiškus failus apibūdina ir atkūrimo rodiklis, kuris parodo, kokia dalis visų tikrų kenkėjiškų failų buvo teisingai aptikta.

2 lentelėje pateikiami kiekvieno požymių konvertavimo į vaizdus metodo rezultatai, nurodant praleistų kenkėjiškų failų skaičių ir procentą. BIE metodas praleido 171 kenkėjišką failą, kas sudaro 16,1 % visų kenkėjiškų pavyzdžių. HSV metodas pasižymėjo šiek tiek geresniu rezultatu – praleisti 149 kenkėjiški failai, tai yra 14 %. IGTD metodas dar labiau sumažino praleistų failų skaičių – jų buvo 117, o tai sudaro 10,9 %. LMIGTD metodas pasiekė

dar geresnį rezultatą – praleisti 106 kenkėjiški failai, arba 9,9 %. Geriausiai pasirodė cBIE metodas, kuris praleido tik 85 kenkėjiškus failus, o tai sudaro 7,9 % visų kenkėjiškų pavyzdžių. Šie rezultatai rodo, kad netiesioginio atvaizdavimo metodai, ypač cBIE ir LMIGTD, leidžia reikšmingai sumažinti klaidingai neaptiktų kenkėjiškų failų skaičių, palyginti su tiesioginiais metodais.

2 lentelė. Praleistų kenkėjiškų programų skaičius ir procentas.

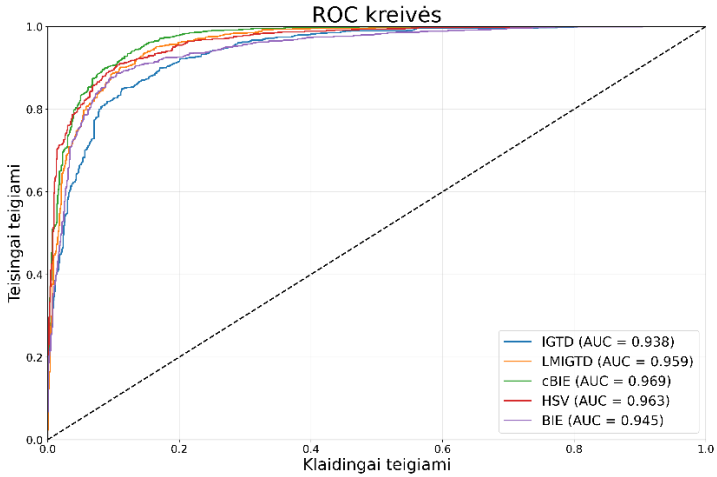
Metodas	Praleistų kenkėjiškų failų skaičius	Praleistų kenkėjiškų failų procentas (%)
BIE	171	16,1
HSV	149	14,0
IGTD	117	10,9
LMIGTD	106	9,9
cBIE	85	7,9

3 lentelė. Klasifikavimo metrikos pagal požymių konvertavimo metodą.

Metodas	Tikslumas	Atkūrimas	Preciziškumas	F1
BIE	0,88	0,84	0,92	0,88
HSV	0,89	0,86	0,92	0,89
IGTD	0,86	0,89	0,84	0,87
LMIGTD	0,89	0,90	0,89	0,90
cBIE	0,91	0,92	0,89	0,91

Analizuojant eksperimento rezultatus, pateiktus 2 lentelėje pateikiami kiekvieno požymių konvertavimo į vaizdus metodo rezultatai, nurodant praleistų kenkėjiškų failų skaičių ir procentą. BIE metodas praleido 171 kenkėjišką failą, kas sudaro 16,1 % visų kenkėjiškų pavyzdžių. HSV metodas pasižymėjo šiek tiek geresniu rezultatu – praleisti 149 kenkėjiški failai, tai yra 14 %. IGTD metodas dar labiau sumažino praleistų failų skaičių – jų buvo 117, o tai sudaro 10,9 %. LMIGTD metodas pasiekė dar geresnį rezultatą – praleisti 106 kenkėjiški failai, arba 9,9 %. Geriausiai pasirodė cBIE metodas, kuris praleido tik 85 kenkėjiškus failus, o tai sudaro 7,9 % visų kenkėjiškų pavyzdžių. Šie rezultatai rodo, kad netiesioginio atvaizdavimo metodai, ypač cBIE ir LMIGTD, leidžia reikšmingai sumažinti klaidingai neaptiktų kenkėjiškų failų skaičių, palyginti su tiesioginiais metodais.

3 lentelėje, nustatyta, kad pagal kenkėjiškų programų atkūrimo ir F1 rodiklius geriausiai pasirodė cBIE metodas: atkūrimo rodiklis siekė 0,92, o F1 – 0,91. Tai rodo, kad taikant šį metodą praleistų kenkėjiškų programų dalis buvo mažiausia. LMIGTD ir HSV metodai taip pat užtikrino aukštą aptikimo efektyvumą, tačiau jų rezultatai buvo kiek žemesni. Tuo tarpu BIE ir IGTD metodai pasižymėjo mažesniu bendru tikslumu ir prastesniu kenkėjiškų programų aptikimu.



4 pav. Konvoliucinio neuroninio tinklo architektūra skirtingų požymių išskyrimo metodų analizei.

ROC (angl. *receiver operating characteristic*) kreivių (žr. 4 pav.) analizė papildomai patvirtina metodų efektyvumo skirtumus. Aukščiausias AUC pasiektas taikant cBIE metodą – šio metodo AUC reikšmė yra 0,969. LMIGTD metodas taip pat pasižymėjo pajėgumu, jo AUC siekia 0,959. HSV metodas pasiekė 0,963 AUC reikšmę. BIE metodo AUC buvo 0,945, o IGTD metodo – 0,938. Šie rezultatai rodo, kad netiesioginio atvaizdavimo metodai pasižymi didesniu jautrumu ir patikimumu. ROC kreivės rezultatai patvirtina, kad cBIE ir LMIGTD metodai yra efektyvūs, o BIE, HSV ir IGTD metodų jautrumas yra mažesnis.

5 Išvados

Šiame tyrime buvo atliktas penkių skirtingų vykdomųjų PE failų požymių konvertavimo į vaizdus metodų palyginimas, siekiant įvertinti jų efektyvumą kenkėjiškų programų aptikimui taikant konvoliucinius neuroninius tinklus. Tyrimui naudotas *PE Malware Machine Learning Dataset* duomenų rinkinys, leidžiantis dirbti su neapdorotais failais ir sudaryti subalansuotas mokymo bei validavimo imtis. Analizuoti tiesioginiai ir netiesioginiai konvertavimo metodai, kurie skiriasi požymių išdėstymo vaizde principais ir informacijos išryškinimo galimybėmis.

Rezultatai parodė, kad netiesioginio atvaizdavimo metodai, ypač cBIE ir LMIGTD, leidžia pasiekti aukštesnį kenkėjiškų programų aptikimo tikslumą ir geresnį bendrą modelio našumą. cBIE metodas išsiskyrė mažiausiu praleistų kenkėjiškų failų kiekiu ir aukščiausiu bendru tikslumu, o BIE metodas pasižymėjo didesniu praleistų kenkėjiškų failų skaičiumi ir ryškiu persimokymu. HSV ir LMIGTD metodai taip pat pasiekė aukštus rezultatus, tačiau cBIE lenkė juos pagal pagrindines klasifikavimo metrikas. IGTD ir BIE metodų rezultatai buvo žemesni tiek pagal bendrą tikslumą, tiek pagal kenkėjiškų programų atpažinimą

Atlikus tyrimą, kurio metu buvo lyginami skirtingi požymių konvertavimo į vaizdus metodai kenkėjiškų programų aptikimui, gautos šios išvados:

- Požymių konvertavimo į vaizdus metodai yra priemonė, leidžianti efektyviai spręsti kenkėjiškų programų aptikimo uždavinius, nes suteikia galimybę išnaudoti konvoliucinių neuroninių tinklų gebėjimą atpažinti sudėtingus struktūrinius ir elgsenos modelius.
- Netiesioginio požymių konvertavimo į vaizdus metodai, ypač cBIE ir LMIGTD, yra efektyviausi kenkėjiškų programų aptikimui naudojant konvoliucinius neuroninius tinklus. cBIE metodas praleido tik 7,9 % visų kenkėjiškų failų ir pasiekė 91 % bendrą tikslumą, o LMIGTD metodas praleido 9,9 % kenkėjiškų failų.
- Tiesioginio dvejetainio kodavimo metodas BIE praleido 16,1 % kenkėjiškų failų ir pasižymėjo ryškiu persimokymu, todėl nėra optimalus sudėtingesnių grėsmių aptikimui.
- HSV metodas praleido 14 % kenkėjiškų failų, o IGTD metodas – 10,9 %, todėl šie metodai taip pat gali būti laikomi efektyviomis alternatyvomis, tačiau jų rezultatai kiek nusileidžia cBIE metodui.

Literatūra

[1] A. Institute, „AV-TEST," [Tinkle]. Available: https://www.av-test.org/en/statistics/malware/. [Kreiptasi 02 2025].

[2] J. Ferdous, R. Islam, A. Mahboubi ir M. Z. Islam, „A Review of State-of-the-Art Malware Attack Trends and Defense Mechanisms," *IEEe Access,* t. 11, pp. 121118-121141, 2023.

[3] M. Guven, „Leveraging deep learning and image conversion of executable files for effective malware detection: A static malware analysis approach," *AIMS Mathematics,* t. 9, nr. 6, p. 15223–15245, 2024.

[4] M. Lester, „PE Malware Machine Learning Dataset," Practical Security Analytics, 8 6 2021. [Tinkle]. Available: https://www.practicalsecurityanalytics.com. [Kreiptasi 02 2025].

[5] A. A. Institute, „MABEL: Malware Analysis Benchmark for Artificial Intelligence and Machine Learning," GitHub, 2023.

[6] J. Halladay, D. Cullen, N. Briner, D. Miller, R. Primeau, A. Avila, W. Watson, R. Basnet ir T. Doleck, „BIE: Binary Image Encoding for the Classification of Tabular Data," *Journal of Data Science,* t. 23, nr. 1, p. 109–129, 2025.

[7] Z. Xian, R. Huang, D. Towey ir C. Yue, „Convolutional Neural Network Image Classification Based on Different Color Spaces," *Tsinghua Science and Technology,* t. 30, nr. 1, pp. 402-417, 2025.

[8] J. Saxe ir H. Sanders, Malware Data Science: Attack Detection and Attribution, No Starch Press, 2018.