ВІДГУК

офіційного опонента

доцента кафедри кібербезпеки та програмного забезпечення

Центральноукраїнського національного технічного університету,

доктора технічних наук, професора Мелешко Єлизавети Владиславівни

на дисертаційну роботу Горносталя Олексія Андрійовича

«Ансамблевий метод ідентифікації стану комп'ютерних систем»,

представлену на здобуття наукового ступеня доктора філософії

за спеціальністю 123 – «Комп'ютерна інженерія»

1. Ступінь актуальності теми дисертаційної роботи.

У сучасному світі комп'ютерні системи відіграють ключову роль у всіх сферах життя. При цьому постійно зростає кількість інформаційних загроз, які спричинені зловмисним вторгненням. Серйозність їх наслідків варіюється від втрати конфіденційності інформації до порушення функціонування критично-важливої інфраструктури. Враховуючи велику кількість інформаційних атак, підвищення їх складності, а також недосконаліть наявних методів, задача ідентифікації стану комп'ютерних систем залишається вкрай важливим напрямком для забезпечення безпеки та стабільності сучасного цифрового середовища.

З метою вирішення цієї наукової задачі необхідно удосконалювати існуючі та розробляти нові інноваційні методи виявлення вторгнень в комп'ютерні системи. При цьому, слід враховувати велику кількість різних підходів та приділяти особливу увагу тим, які передбачають використання технологій штучного інтелекту та машинного навчання. Саме тому, дисертаційна робота Горносталя Олексія Андрійовича, яка спрямована на розробку та удосконалення методів, що базуються на ідеї ансамблевих класифікаторів є актуальним завданням.

Основна ідея роботи полягає у дослідженні різних підходів та алгоритмів об'єднання базових моделей класифікаторів в ансамбль для підвищення якості ідентифікації стану комп'ютерної системи.

2. Зв'язок роботи з науковими програмами, планами, темами.

Дисертація виконувалась відповідно до наукової програми 123 – «Комп'ютерна інженерія» та була впроваджена на кафедрі комп'ютерної

інженерії та програмування, навчально-наукового інституту комп'ютерних наук та інформаційних технологій, НТУ «ХПІ».

Здобувач брав участь у науково-дослідній роботі «Моделі і методи обробки та захисту інформації в комп'ютерних системах» (ДР №0122U200526), де замовником виступало ТОВ «Передові цифрові рішення», у якості відповідального виконавця.

3. Оцінка змісту дисертації, її завершеності й оформлення.

Структура та оформлення дисертації відповідає прийнятим для наукового дослідження нормам. Усі положення, які винесені на захист, висвітлені в тексті дисертації. Зміст дисертаційної роботи відповідає її назві.

Дисертаційна робота Горносталя Олексія Андрійовича складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та 3 додатків.

У вступі обґрунтовано актуальність теми дисертації, розглянуто зв'язок роботи з науковими програмами, планами, темами, сформульовано мету, задачі, об'єкт, предмет та методи дослідження, розглянуто наукову новизну та практичне значення отриманих результатів. Відзначено особистий внесок здобувача, апробацію результатів дисертаційної роботи на конференціях та симпозіумах, наведено відомості про публікації за темою роботи.

У першому розділі досліджено проблеми систем виявлення вторгнень, проаналізовано їх основні складові. Розглянуто методи ідентифікації стану комп'ютерних систем, сформульовано їх переваги та недоліки. Обґрунтовано вибір ансамблевих методів для подальшого дослідження. Виконано постановку наукової задачі.

У другому розділі наведено формальну постановку задачі класифікації. Розглянуто особливості використання беггінг-ансамблів у задачах ідентифікації стану комп'ютерних систем. Виконано вибір етапів та методів попередньої обробки даних. Досліджено процес формування вхідних послідовностей при побудові беггінг-ансамбля, а також процедуру налаштування мета-алгоритму та базових моделей. За результатами дослідження удосконалено метод ідентифікації стану комп'ютерної системи, що включає процедуру попередньої обробки даних, вибір алгоритму формування вхідних даних та побудову беггінг-класифікатора з налаштуванням параметрів базових моделей та ансамблю, що дозволило підвищити його ефективність.

У третьому розділі сформульовано основні недоліки класичної процедури ансамблювання у беггінг-класифікаторах та розглянуто основні підходи її вдосконалення. Проаналізовано ефективність використання багатошарового перцептрону у якості базової моделі. Проведено експериментальне дослідження ефективності використання різних варіантів ансамблевої обрізки, зваженого голосування, калібрування впевненості та адаптації за рахунок мета-навчання та мета-ознак.

Четвертий розділ містить огляд основних переваг та недоліків гомогенних та гетерогенних ансамблів. Обґрунтовано вибір базових моделей машинного навчання, які будуть брати участь у процесі формування класифікатора, а також розглянуто основні показники, які можуть використовуватися в процесі їх відбору. Запропоновано метод побудови гетерогенного ансамблю, який включає триетапний процес підбору базових моделей класифікатора на основі технології Pasting, що дозволило підвищити якість класифікації.

Кожен з розділів містить висновки та рекомендації щодо використання розглянутих підходів та методів. У висновках до роботи зазначено основні результати дисертаційного дослідження та їх впровадження.

Список використаних джерел містить 137 посилань та широко охоплює область дослідження .

В додатку А представлено список наукових праць здобувача за темою дисертації. В додатку Б представлені фрагменти текстів програм, які використовувалися для експериментальної перевірки розглянутих в основних розділах методів ідентифікації стану комп'ютерних систем з використанням беггінг-ансамблів. В додатку В зазначені акти впровадження наукових результатів.

4. Наукова новизна одержаних результатів.

До основних нових наукових результатів дисертації слід віднести наступне:

1. *Отримав подальший розвиток* метод ідентифікації стану комп'ютерної системи на основі беггінг-ансамблю з деревами рішень у якості базових моделей, розробленою процедурою попередньої обробки даних та за рахунок вибору оптимальних гіперпараметрів класифікатора.

2. *Отримав подальший розвиток* ансамблевий метод ідентифікації стану комп'ютерної системи, який використовує багатошаровий перцептрон у якості

базової моделі та процедуру вибору оптимальних гіперпараметрів налаштування класифікатора.

3. *Удосконалено* ансамблевий метод ідентифікації стану комп'ютерної системи завдяки комплексному використанню процедури обрізки ансамблю та зваженого голосування з вагами на основі функції логарифмічних втрат разом з гомогенним класифікатором.

4. *Вперше* запропоновано метод ідентифікації стану комп'ютерної системи з використанням розробленої процедури побудови гетерогенного ансамблю шляхом поетапного вибору різнорідних базових моделей та їх відбору за допомогою технології Pasting.

5. Достовірність отриманих результатів та висновків.

Достовірність отриманих наукових положень, висновків та рекомендацій забезпечується аргументованими результатами досліджень і підтверджується співставленням з результатами проведених експериментів.

6. Практична цінність одержаних результатів та рекомендації щодо їх подальшого використання.

Практична цінність роботи підтверджується успішним впровадженням результатів дослідження у діяльність компанії ТОВ «Передові цифрові рішення», а також у навчальний процес Національного технічного університету «ХПІ».

Серед практичних здобутків можна виділити наступні:

1. Реалізовано програмну модель попередньої обробки даних, що дозволило підвищити швидкість розпізнавання до 1,62 разів, а також зменшити час навчання моделей до 24,76 разів.

2. Використання розробленого методу ідентифікації стану комп'ютерної системи, який включає сформовану процедуру попередньої обробки даних, процес вибору алгоритму формування вхідних даних та побудову беггінг-класифікатора з налаштуванням його гіперпараметрів, дозволило підвищити значення *AUC-ROC* класифікатора на навчальній вибірці на 11%, а на тестовій вибірці – на 3%.

3. Розроблено програмне забезпечення на основі багатошарового перцептрону з процедурою вибору оптимальних налаштувань, використання якого дозволило підвищити точність класифікації на 4,67%.

4. Розроблено програмне забезпечення, яке виконує обрізку ансамблю на основі максимізації абсолютної точності базових класифікаторів та класифікує за допомогою зваженого голосування з використанням вагових коефіцієнтів на основі функції логарифмічних втрат, що дозволило підвищити показники якості класифікації беггінг-ансамблю, а саме, значення метрики F_1- *Score* – на 2,4%.

5. Розроблено програмне забезпечення на основі запропонованого методу формування гетерогенного ансамблю, що дозволило підвищити якість класифікації, а саме, збільшити показник F_1-*Score* моделі при роботі на тестових даних на 9,5% у порівнянні зі стандартним однорідним беггінг-ансамблем на основі дерев рішень та на 2% у порівнянні з максимальним значенням серед однорідних ансамблів.

7. Оформлення дисертації, дотримання вимог академічної доброчесності та повнота викладення наукових положень і результатів в опублікованих працях.

Дисертаційне дослідження виконано відповідно до наукових стандартів та академічної доброчесності. Отримані результати підтверджують оригінальність наукового дослідження. У тексті присутні авторські ідеї та не виявлено використання концепцій інших вчених без належних посилань. Результати досліджень опубліковані у 20 роботах, серед яких: 1 стаття у науковому фаховому виданні України, що індексується у науково-метричній базі Scopus, 3 статті у співавторстві з науковим керівником та 1 стаття у співавторстві з двома чи більше особами в наукових фахових виданнях України категорії «Б», а також 15 матеріалів міжнародних конференцій.

У дисертаційній роботі зазначено особистий вклад здобувача у наукових друкованих працях, які були опубліковані у співавторстві.

За темою дисертації зараховано 4 публікації, які відповідають вимогам пункту 8 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого Постановою КМУ від 12.01.2022 р. №44: 1 стаття у періодичному науковому фаховому виданні України, яке проіндексовано у наукометричній базі даних Scopus, 3 статі з двома співавторами (разом із здобувачем) у фаховому виданні України категорії «Б» та 1 стаття з чотирма співавторами у фаховому виданні України категорії «Б»

(прирівнюється до 0,5 публікації). Таким чином, виходячи з вимог 8 пункту вищенаведеної постанови, наукові результати дисертації висвітлені у 4,5 наукових публікаціях здобувача, що є достатнім для дисертації рівня доктора філософії.

8. Дискусійні питання та зауваження до дисертаційної роботи.

1. У першому розділі при розгляді різних груп методів ідентифікації стану комп'ютерних систем основна увага сфокусована на евристичних методах, й менш детально розглянуті сигнатурні. Також відсутня порівняльна таблиця розглянутих методів.

2. У другому розділі виконано теоретичне обґрунтування вибору конкретних показників якості класифікації, які можна використовувати в процесі оцінки ефективності роботи досліджуваних ансамблевих класифікаторів, проте воно є досить стислим, адже не вистачає більш поглибленого опису окремих ситуацій, в яких було б краще використати той чи інший показник якості, наприклад, при балансуванні помилок першого та другого роду.

3. У третьому розділі теоретично обґрунтовано вибір використання саме багатошарового перцептрону у якості базової моделі у рамках ансамблів, а також процедуру вибору оптимальних налаштувань параметрів, проте не проведено експериментальне порівняння з іншими моделями та підходами.

4. У четвертому розділі запропоновано процедуру формування гетерогенного ансамблю та введено параметр k, який відповідає за кількість різних методів машинного навчання, моделі яких можуть бути об'єднані в рамках одного ансамблю. При цьому, розглянуто лише 3 з можливих значень цього параметру, а саме 2, 3 та 5. Вважаю, що доцільним було б розглянути більше значень даного параметру.

5. В дисертації у висвітленні практичного значення отриманих результатів присутні наступні формулювання: «розроблено метод…», «запропоновано метод…». Вважаю, що подібні формулювання відносяться до наукової новизни. В той же час для опису практичної значимості доречно використовувати такі формулювання: «розроблено програмне забезпечення на основі запропонованого методу…», «розроблені алгоритми для реалізації запропонованого методу…» тощо. Тим паче, що з тексту пунктів практичної значимості видно, що саме це, зокрема, розробку програмного забезпечення, й мав на увазі здобувач.

9. Висновки.

Зазначені недоліки до дисертаційної роботи не впливають на загальне позитивне враження від проведеного наукового дослідження, не зменшують якість та наукову цінність роботи.

Дисертаційна робота Горносталя Олексія Андрійовича «Ансамблевий метод ідентифікації стану комп'ютерних систем» за своїм змістом відповідає спеціальності 123 – «Комп'ютерна інженерія». Дисертація є завершеною науково-дослідною роботою, яка розв'язує важливу науково-практичну задачу, що полягає у підвищенні якості ідентифікації стану комп'ютерних систем шляхом використання ансамблевих класифікаторів у процесі вдосконалення та розробки різних методів.

Отже, враховуючи актуальність теми, отримані результати та певну практичну значущість, вважаю, що дисертаційна робота Горносталя Олексія Андрійовича «Ансамблевий метод ідентифікації стану комп'ютерних систем» відповідає вимогам до дисертацій на здобуття наукового ступеня доктора філософії, а саме вимогам пунктів 6, 7, 8 і 9 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого Постановою КМУ від 12.01.2022 р. №44, та вимогам до оформлення дисертації, затвердженим Наказом МОН України від 12.01.2017 № 40, а здобувач Горносталь Олексій Андрійович заслуговує присудження наукового ступеня доктора філософії за спеціальністю 123 – «Комп'ютерна інженерія».

Офіційний опонент:

доктор технічних наук, професор
доцент кафедри кібербезпеки та
програмного забезпечення
Центральноукраїнського національного
технічного університету **Єлизавета МЕЛЕШКО**

"29" травня 2024 р.

Підпис Мелешко Є.В. засвідчую:
Проректор з науково-педагогічної роботи ***Андрій КИРИЧЕНКО***
"29" травня 2024 р.