CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Right for the right reason in malware classification**
Author's name:	**Zlochevskyi Vladyslav**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Ing. Václav Mácha, Ph.D.
Reviewer's department:	Gen Digital Inc.

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **ordinarily challenging**

How demanding was the assigned project?

The assignment required the student to understand advanced methods at a sufficient depth to comprehend the differences between them and successfully implement them.

Fulfilment of assignment **fulfilled**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The author completed all tasks specified in the assignment.

Methodology **correct**

Comment on the correctness of the approach and/or the solution methods.

The methodological approach employed in the thesis is correct. The author uses appropriate methods for defining experiments, evaluating them, and interpreting the results, demonstrating sound scientific methodology throughout the work.

Technical level **B - very good.**

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

The author successfully implemented multiple non-trivial Explanatory Interactive Learning (XIL) methods. All experiments were properly described with comprehensive detail, including neural network architecture specifications, training configuration parameters, evaluation criteria, and more. The author provided thorough discussion of the results for each method separately and conducted detailed comparisons between the methods. The experimental results are discussed from multiple different viewpoints, including performance, computational efficiency, robustness, and explanation alignment. Additionally, the author provided a GitHub repository with all code used for the experiments, ensuring reproducibility and transparency of the work.

Formal and language level, scope of thesis **D - satisfactory.**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

The thesis exhibits several issues with mathematical formalism and notation that affect its overall presentation quality. Key mathematical concepts such as the ReLU function, convolution layers, and dense layers are mentioned repeatedly throughout the work but lack formal mathematical definitions. The author tends to describe mathematical concepts using words rather than providing precise equations followed by explanations of their components. Additionally, equation numbering is inconsistent: for instance, the cross-entropy loss function defined on page 5 lacks an equation number despite being referenced multiple times later in the work (e.g., on page 18, where it is referenced by subsection number). Mathematical notation usage shows inconsistencies: for example, the loss function is denoted as **L** on page 4 but switches to $\mathcal{L}$ on page 9. More problematically, the same symbols are reused for different concepts: for example, **L** denotes both the loss function (page 4) and the number of convolutional layers of interest (equation 3.6 on page 18).

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

Selection of sources, citation correctness	**D - satisfactory.**
Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?	
The thesis demonstrates satisfactory citation practices with some notable gaps. The list of sources adequately covers the content of the work, and bibliographic citations are used in a standard way. The student's original work is properly distinguished from earlier work in the field. However, citations are missing in several instances where they would be appropriate: for example, the cross-entropy loss function is defined on page 5 without citation, and while this is a well-known definition, proper academic practice requires citing the source. More generally, the author should cite sources more frequently throughout the text to properly acknowledge the foundations upon which the work is built.	

Additional commentary and evaluation (optional)
Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

The thesis successfully addresses the assigned task with correct methodology and good technical implementation. While the theoretical part lacks consistency in notation and formal presentation, the experimental work is substantially stronger. The experiments are well-defined and thoroughly described with all necessary details. The results are properly discussed from multiple viewpoints, and differences between methods are clearly highlighted. Despite weaknesses in the theoretical part, the successful implementation of multiple non-trivial XIL methods and strong experimental analysis demonstrate the author's solid technical competence.

- On page 36 (paragraph "Equal weighting ensure counterexample influence"), the author claims that the Balanced CAIPI method is beneficial when the number of counterexamples is much smaller than the original data. However, in the following paragraph, the author argues that equal weighting can over-emphasize counterexamples when their number is larger than the original data. Could you clarify this apparent contradiction? Since Balanced CAIPI computes the mean loss separately for original data and counterexamples (equation 3.9), shouldn't the method be robust to the relative sizes of these datasets?
- In the "08 MNIST" experiment, the RRR method successfully suppressed the spurious dot signal, while CAIPI methods maintained attention on both cross and dot regions. The CAIPI methods achieved better performance when evaluated on data containing only 0s and 8s (without dots). Did the author consider evaluating the methods on data containing only the dot pattern?

The grade that I award for the thesis is **C - good.**

Date: 20.1.2026 Signature: