2024-05-03

LUND UNIVERSITY

Nämnden för utredning av
avvikelser från god forskningssed,
Uppsala universitet

Dept. of Electrical and Information Technology
Prof. Fredrik Tufvesson

Sakkunnigutlåtande, ärende UFV 2023/2045

Nedan är mitt utlåtande kring artikeln ”Deep Learning for Security in Digital Twins of Cooperative Intelligent Transportation Systems” [1] publicerad i *IEEE Transactions on Intelligent Transportation Systems* september 2022. Nämnden har bett om ett utlåtande kring följande tre frågor, vilka jag behandlar nedan.

1. Kan det som anförs i anmälan rörande artikeln Deep Learning for Security in Digital Twins of Cooperative Intelligent Transportation Systems (https://ieeexplore.ieee.org/document/9560727) utgöra en grund för oredlighet (https://npof.se/oredlighet-i-forskning/) eller avvikelse från god forskningssed?
2. Är Zhihan Lyus yttrande en acceptabel förklaring till det som presenterats som felaktigheter i artikeln?
3. Förekommer något i övrigt i artikeln vilket kan utgöra en avvikelse från god forskningssed/oredlighet i forskning?

Jag har efter samråd med nämndens sekreterare valt att besvara frågorna på engelska nedan så att den anmälde, Zhihan Lv (vilket är stavningen i [1], men som i Uppsala universitets personalkatalog stavas Zhihan Lyu), också kan läsa texten direkt.

1 Can what is stated in the report regarding the paper [1] constitute a basis for misconduct or deviation from good research practice?

In the report, two issues were brought up. The first one is about inconsistency. It was claimed in [1] that the loss is zero only when f(x) = y . The loss function in (7) does not satisfy this claim.
I agree with this comment in the report, the paper is inconsistent here. It is a minor but obvious inconsistency. I would not classify this alone as misconduct or deviation from good research practice. It could be just that the authors were not careful enough in their writing and internal review. The authors are responsible for the content of the paper and the inconsistency should not be there, but it could be a minor mistake. It could, however, also be interpreted as a sign that something is strange, in case the authors have implemented their analysis according to the description in the paper.

Dept. of Electrical and Information Technology, Box 118, SE-221 00 Lund, Sweden *Visiting address* Ole Römers väg 3, SE-223 63 Lund, Sweden *Telephone* Int. +46 46 222 46 29 *Telefax* Int. +46 46 12 99 48 *E-mail* fredrik.tufvesson@eit.lth.se *Internet* http://www.eit.lth.se

The second comment, that the curves in Fig. 6 seem not consistent, is more serious. The report gives an example that for the CNN with 20 epochs (the red curve), the precision is about 70% (fig 6b), the recall is about about 67% (fig 6c), but the F1 is below 67% (Fig 6d).

The F1 score (Fig. 6d) is calculated as $2*P*R/(P+R)$, where P is the precision (Fig. 6b) and R is the recall (Fig. 6c). This means that the F1 score is always larger or equal to the smallest of precision and recall, which is not the case in Fig. 6. The given example is about the red curve (CNN), but also for, e.g., the green curve (MLP) the F1 score is smaller than the smallest of precision and recall at an Epoc of 20; The F1 score for MLP at an Epoc of 20 is around 52.5%, whereas the recall is above 56% and precision is well above 60%. Hence one can conclude that Figures 6b, 6c, and 6d are not consistent and that it is not just a single mistake for one curve. The report is correct in its statement, and it can in fact be due to falsification or fabrication of data in the paper, hence, a case of scientific misconduct.

2 Is Zhihan Lyu's statement an acceptable explanation for what has been presented as inaccuracies in the article?

Zhihan Lv's reply to the comments is that this can happen since it is a multi-category classification task and that the number of samples in different categories varies greatly. I do not agree with this explanation since the F1 score is directly calculated from precision and recall from the same simulations set, it is just a combined metric of the other two parameters. Hence, this inconsistency cannot happen as explained by Zhihan Lv.

3. Is there anything else in the article that may constitute a deviation from good research practice/misconduct in research?

In fact, there are other strange things in the paper. The paper has 33 references. In the text where the references are given, there is, as common practice, a statement or a subject that you then expect to find in the reference. However, when reading many of the references this is not the case, the statement or subject is not dealt with in the given reference. Examples include:

"In the field of intelligence, the intelligent algorithms for autonomous extraction of DL data features will result in notable consequences when applied to path planning and road condition prediction, reducing various security accidents on the road [6]."

[6] is about Artificial intelligence and COVID-19: Deep learning approaches for diagnosis and treatment, and does not deal with path planning or security accidents on the road.

"When passing the toll station, the reading device on the lane establishes real-time communication with the card on the vehicles, and the toll is deducted from the user's toll card [17]."

[17] does not deal with toll stations, toll cards, in fact the term "toll" is not mentioned in [17] at all.

"In the proposed CNN-SVR-based online module of the CITS DTs coordination strategy, the traffic flow distribution scheme is based on an adaptive signal control scheme [22]"

[22] is about a dual-loop Control Scheme for a Single-Phase uninterruptable power supplies, i.e. unrelated to intelligent transportation systems. While it is possible that this control scheme is used, it is not likely and there are no details supporting that it actually is used.

"Support Vector Machine (SVM) is a linear classifier that applies to binary classification operations and regression problems. In the latter case, it is called SVR [20]."

The only place in [20] where SVR is mentioned is "Non-parametric models are then proposed to address the problems of the nonlinear systems [10]. Hong et al. [11] proposed a support vector regression (SVR) model optimized by continuous ant colony algorithms to forecast inter-urban traffic flow." Hence it is not supporting the claim.

"Meanwhile, the following three variables, Market Penetration Rate (MPR) [27], Congestion Level (CL) [28], Average Velocity (AV) [29], Average Traveling Time (ATT) [30], Average Delay (AD) [31], Average Pollutant Emission (APE) [32], and Following Rate (FR) [33], are employed as influencing factors to analyze the security impact of the system in the transportation field."

[27] does not mention market penetration rate

[28] does not mention congestion level

[29] does not mention average velocity

[30] does not mention average travel time

[32] is now retracted by the publisher

[33] does not mention following rate.

Such referencing is clearly a deviation from good research practice. As it is so frequent and seems to be systematic it can probably be considered as a serious deviation from good research practice misleading the reader.

In addition, for the simulations in [1] the authors claim that they use a tool Veins "The proposed CITS system DTs model is predicted and analyzed on the Veins online simulation platform to verify its effectiveness."

To be precise there is no Veins online simulation platform, but one can *guess* that the authors mean Instant Veins, which is "a virtual machine you can use to quickly try out Veins or to run Veins on systems you do not have administrative access to". No reference is given to the origin of the Veins simulation platform, though it is an essential tool for the paper, and the Veins creators suggest that you cite it if you use it:

"If you use Veins or one of its component models, we would appreciate a citation of our work:

- IEEE Christoph Sommer, Reinhard German and Falko Dressler, "**Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis**," *IEEE Transactions on Mobile Computing (TMC)*, vol. 10 (1), pp. 3-15, January 2011. "

Furthermore, in [1] there are no detailed simulation settings described:

"The proposed CITS system DTs model is predicted and analyzed on the Veins online simulation platform to verify its effectiveness. The IEEE 802.11 MAC protocol is adopted. During experiments, the vehicles are indeed covered in the transmission range of wireless signals. However, the signal transmission may fail due to various factors. Relevant experimental data are obtained by simulating the constructed system model…. During simulation experiments, the hyperparameters are set as follows: 20 iterations, 1,000 seconds of simulation, 0.002 learning rate, and 128 batch size."

One would expect such detailed simulation parameters to be as the simulation results are heavily dependent on the scenario and parameters used. Hence it is impossible to validate or reproduce the work. One can even question if simulations are made and simulation data has been used as there is no indication of simulation results or scenario description given.

Finally, Zhihan Lv has authored another paper "Semi-Supervised Support Vector Machine for Digital Twins Based Brain Image Fusion", published in Frontiers in Neuroscience 2021 [2]. Zhihan Lv is corresponding author of this paper and "was responsible for the design of the experiment and the inspection of the results". This paper deals with Brain Image Fusion using "Magnetic Resonance Imaging (MRI) data from the Brain Tumor Department of a Hospital".

"Brain image data in experiments come from MRI records of the Brain Tumor Department of a Hospital. It contains 20 groups of clinical data, and each group of data includes four kinds of multi sequence MRI images: T1, T2, Tlc, and FLAIR. The selected image data are processed by registration, skull peeling, contrast enhancement and other operations."

It is strange that no proper reference is given to the dataset in this paper, especially since the results are heavily dependent on the MRI dataset, it is not even specified which hospital the dataset comes from. The strangest thing though is that the curves in Figure 9 in [2] "Brain image assessment indicators of different models with iterations (A) The Jaccard coefficient; (B) DSC; (C) PPV; (D) Sensitivity." Are exactly the same as the curves of Figure 6 in [1] "Comparison of the accuracy curves of different learning algorithms." The two figures are coming from different datasets: vehicular simulations, and brain images, respectively. The lines in the two figures mean different things but they have the exactly the same shape. This cannot happen by accident and one can conclude that at least one of the papers [1] and [2], maybe both, are based on fabricated or falsified data. Given the inconsistencies under point 1, it is likely that at least [1] contains fabricated or falsified data.

Summary

The paper [1] contains inconsistent results that cannot occur naturally. Paper [1] and paper [2] show result figures that are exactly the same, though they should show different things and come from totally different datasets. Such a similarity of the curves cannot be explained to occur naturally but indicates that [1], or [2], or both papers contain falsified or fabricated data. Given the inconsistencies in Fig. 6 of [1], and the lack of simulation parameters, simulation scenario and simulation results, it is likely that [1] is based on falsified or fabricated data. Hence, this can clearly be a case of research misconduct.

Furthermore, [1] contains many strange references and statements that are not supported by the references given, and is lacking a reference to a key tool for the investigation. It is a deviation from good research practice and due to its systematic misuse of referencing it could possibly be seen as a serios deviation from good research practice.

Referenser:

[1] Z. Lv, Y. Li, H. Feng and H. Lv, "Deep Learning for Security in Digital Twins of Cooperative Intelligent Transportation Systems," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16666-16675, Sept. 2022, doi: 10.1109/TITS.2021.3113779.

[2] Wan Z, Dong Y, Yu Z, Lv H and Lv Z (2021) Semi-Supervised Support Vector Machine for Digital Twins Based Brain Image Fusion. Front. Neurosci. 15:705323. doi: 10.3389/fnins.2021.705323

Lund, 2024-05-03

Prof. Fredrik Tufvesson