The Effect of Noisy and Blurry Data on Deep Learning: Application in Brain Image Classification

Muhammad Fajar Azka Fadillah, Dewinda Julianensi Rumala, Mauridhi Hery Purnomo, I Ketut Eddy Purnama

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

INTRODUCTION

Convolutional Neural Networks (CNN) is one of the best Deep Learning algorithms **commonly used** for **computer vision tasks**, including **medical image analysis**.

CNN can learn the **representational features** from **images** starting from the **lower to complex features**. However, **noisy data** can **affect** the **generalization of the networks**, which is **often found** in **medical images**, such as **Magnetic Resonance Imaging (MRI)**.

In this research, we want to see the **relation** between **noisy and blurry data** and the **performance of CNN models**

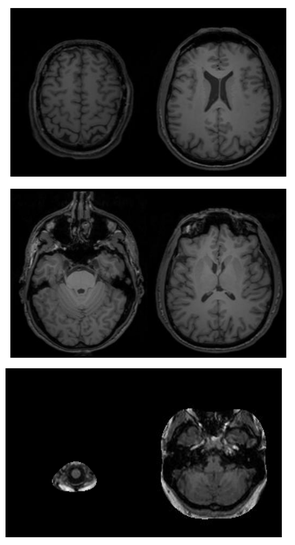

Figure 1. Anatomical categorization of the brain

METHODS

We investigate a **clinical task** of brain image classification, specifically for **anatomical classification of the brain** using **MRI.** For this classification task, we classify the brain into **Class A:** Upper part of the brain, **Class B :** Middle part of the brain, and **Class C:** Lower part of the brain.

We build **three CNN models** to evaluate **three different scenarios**: **original data, blurry data, noisy data.**

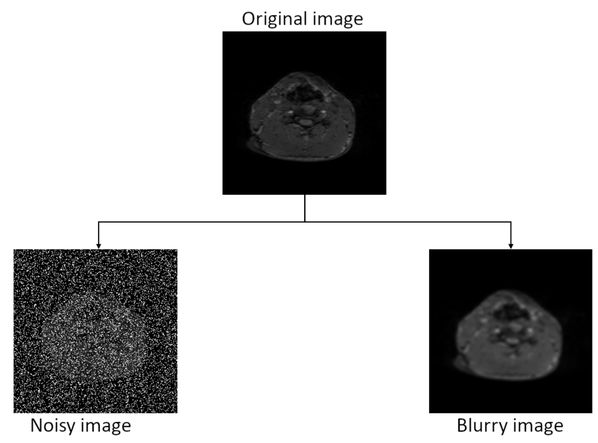

Figure 2 . Three scenarios on data input

RESULTS

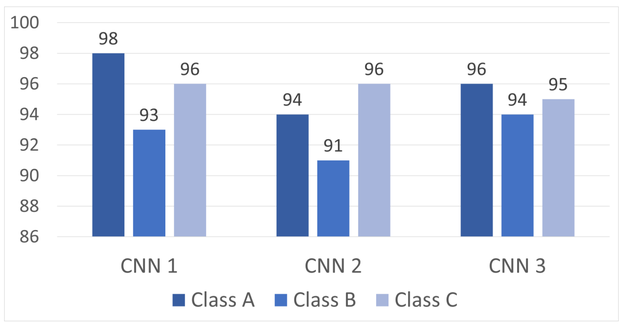

Figure 3. Accuracy of the CNN models on original data

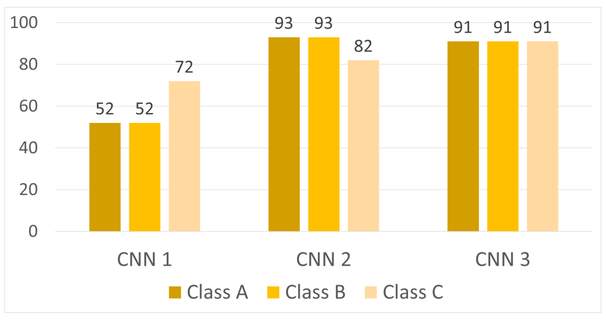

Figure 4. Accuracy of the CNN models on blurry data

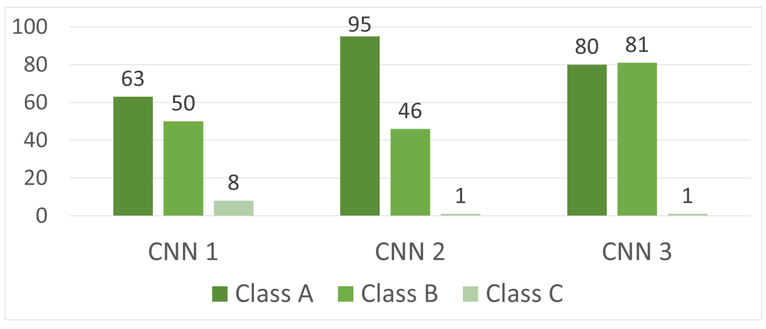

Figure 5. Accuracy of the CNN models on noisy data

- Figure 3 shows that the developed CNN 1 is more powerful in identifying original data, however the accuracy difference between each CNN model is not significant.
- Figure 4 and Figure 5 shows that the accuracy of CNN 1 falls drastically when evaluating blurry and noisy data. Meanwhile CNN 2 shows better performance in classifying blurry data, but performs not good on noisy data.
- On average, CNN 3 shows superior performance than the other models in the classification task using noisy and blurry data.

CONCLUSION

- **Noisy** and **blurry** data can **hurt** the **CNN performance** by **16.00%** and **47.67%,** respectively, on average.
- **CNN Models** with **deeper layers** and **smaller convolutional kernels** that are **trained on an ideal epoch** can deliver **better outcome** when **dealing with blurry** and **noisy data.**

ACKNOWLEDGMENT

This paper is partially funded by UCE AIHes of Sepuluh Nopember Institute of Technology and Indonesia Endowment Fund for Education under the scheme of Riset Inovatif Produktif (RISPRO) - Invitasi 2019 Grant.

REFERENCE

1. S. Gupta, A. Gupta, "**Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review**", Procedia Computer Science, Volume 161, 2019.
2. Jose A. Saez, Julian Luengo, Francisco Herrera**"Evaluating the classifier behavior with noisy data considering performance and robustness: the Equalized Loss of Accuracy measure."** Neurocomputing, 176 (2016), pp. 26-35
3. Luis PF Garcia, Andre CPLF de Carvalho, Ana C. Lorena, "**Effect of label noise in the complexity of classification problems**" Neurocomputing, 160 (2015), pp. 108-119
4. Pelletier, Charlotte et al. (2017), "**A Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series.** ***Remote Sensing"***, 9: 173.