# STATEMENT OF PURPOSE
## PRACTICAL MACHINE INTELLIGENCE

**Divyam Madaan**
School of Computing
KAIST
dmadaan@kaist.ac.kr

## 1 INTRODUCTION

Machine learning has achieved remarkable performance in diverse application domains; however, it is well-established that many deep neural network architectures are over parameterized and extremely susceptible to natural and synthetic distribution shifts. Consequently, this limits the deployment and scalability of these networks to real-life applications. To tackle these challenges, I am primarily interested in working towards improving the generalization and reliability of these networks while making them efficient for deployment in resource-constrained devices.

Currently, I am a Masters student in the School of Computing at Korea Advanced Institute of Science and Technology (KAIST), under the supervision of Professor Sung Ju Hwang. I also closely collaborate with Professor Jinwoo Shin and researchers at FOR.ai to attain my objectives of scalable and reliable machine intelligence. My career goal is to pursue my research interests after PhD as a faculty in academia so that I can contribute to the fields of deep learning and neural networks while training others to solve relevant problems in these rapidly progressing fields.

The commitment of International Max Planck Research School for Intelligent Systems (IMPRS-IS) towards machine learning and deep learning, as well as its exemplary faculty, makes it an ideal environment for me to conduct research during the doctoral programme. My research interests align closely with the faculty and students at IMPRS-IS, and it would be an honour to collaborate with such distinguished researchers. Professors Matthias Hein, Mathias Bethge, and Weiland Brendell have contributed immensely to the field and in particular, Professor Hein's work on the security of models has inspired me immensely; his recent work AutoAttack (Croce & Hein, 2020) opens many research areas on evaluating and training adversarial defense methods, which I would love to explore during my PhD. Professors Mathias Bethge and Weiland Brendel's recent work on robustness against diverse image corruptions (Rusak et al., 2020) motivated my recent work on meta-learning an input-dependent noise to enhance the robustness against multiple perturbations (Madaan et al., 2020a).

## 2 RESEARCH EXPERIENCE

While I always have had a passion for computer science, my interest towards machine learning thrived in my third year of the undergraduate programme when I joined FOR.ai — a student-organized machine learning research group with members from all over the globe. My first research project in FOR.ai was improving Grave's Adaptive Computation Time (Graves, 2016) — instead of a learned ponder cost that requires additional processing steps, I proposed an alternative ponder function that scaled the ponder cost term with the inverse of task loss. Moreover, during the end of my third year of undergraduate studies, I worked at IIT Delhi to tackle the critical problem of air pollution in different areas of Delhi. In particular, I introduced VayuAnukulani framework (Madaan et al., 2019) to predict air pollution in different locations of Delhi. This project was a part of the Celestini Program by Marconi Society mentored by Dr Aakanksha Chowdhery (Google Brain) and Professor Brejesh Lal (IIT Delhi). This work was later published at GlobalSIP 2019. Although the proposed framework predicted the air pollution precisely, its scalability was constrained due to the large model size and lack of reliability measures.

Motivated by the problem of network reliability, I worked on a project that analyzed the Shake-Shake regularization (Gastaldi, 2017) to improve the performance of neural networks against adversarial perturbations. This project was led by Aidan Gomez (a PhD student at Oxford, advised by Professor Yarin Gal), in collaboration with Dr Ian Goodfellow (Google Brain) and Dr Lukasz Kaiser (Google Brain). Since ensemble networks are comparatively challenging to circumvent by an adversary, we looked into Shake Shake regularization to emulate multiple networks. Motivated by Adversarial Logit Pairing (ALP) (Kannan et al., 2018), we formulated logit packing and Refusal of Response (RoR) to refuse the classification of examples where the classifier is not confident. However, ALP was circumvented (Engstrom et al., 2018) due to the problem of obfuscation of gradients. Consequently, we deceived our defense by more vigorous evaluation techniques.

Even though we failed in our attempt to achieve adversarial robustness, I became better acquainted with the essentials of adversarial defenses evaluation. More importantly, I learnt how to develop research questions and methods to answer them. Further, to tackle the problem of scalability of neural networks, I contributed to developing Targeted Dropout (Gomez et al., 2019), which is a simple and effective regularisation scheme that can drastically increase the achieved parametric sparsity of neural networks. Notably, it does not require fine-tuning and its configurability, simplicity, and effectiveness across network architectures make it a promising approach for production model generation.

These research experiences helped me to initiate an independent research project under Professors Sung Ju Hwang and Jinwoo Shin at KAIST. The research focus of the project was to amalgamate the compression and robustness of neural networks. Specifically, we investigated the robustness of neural networks from the perspective of robustness in the latent-feature space.

We minimized the difference between the latent-features for clean and adversarial examples and pruned the latent-features with high distortion by utilizing Bayesian pruning. We conjectured that reducing the distortion in latent-features will improve the overall robustness of the network. As a result, we achieved state-of-the-art robustness and scalability in memory and resource-constrained devices. Finally, our paper, 'Adversarial Neural Pruning with Latent Vulnerability Suppression' (Madaan et al., 2020b) was published at ICML 2020.

As a consequence of the research independence, I learned to make my own decisions, to develop my methods, to communicate my results often and effectively, and to ask for feedback and assistance when it was necessary. I also realized that most of the existing defenses are brittle — they overfit to a single type of perturbations and can be circumvented by attacks unseen during training. Consequently, I analyzed the generalization across multiple $\ell_p$ norms. The primary concern for multi-perturbation training was the training time, which was ten folds compared to the standard training. I first resolved this challenge by stochastically sampling the perturbation and enforcing the label consistency across multiple perturbations. This was motivated by the fact that we don't need all the perturbations subsequently, and the model should output the same predictions for different perturbations of the same image. Furthermore, I utilized a meta-learning framework to generate input-dependent noise to minimize adversarial loss and promote label consistency across multiple perturbations. This work (Madaan et al., 2020a) was recently accepted at the NeurIPS workshop on Meta-Learning, 2020 and is currently under review for publication. I believe that our method can be a firm guideline when other researchers pursue similar tasks in the future.

To further improve the performance of compression techniques, I am currently working with the team at FOR.ai to develop sparse ensembles to produce ensembles with memory equivalent to a single model with significantly better performance. It involves increasing the diversity of members during a single training scheme to improve the generalization and uncertainty across the wide-variety of datasets and architectures. In addition, I am investigating the problems of fairness and bias in the state-of-the-art compression methods as a part of my research at KAIST. In other words, I am formulating a technique to resolve the algorithmic bias towards natural distribution shifts in existing sparsification techniques.

## 3 Academic Background

To broaden my knowledge in deep learning, I credited various courses offered at KAIST — 'Advanced deep learning', 'Machine Learning for AI', 'Artificial Intelligence and Machine Learning'. One other course that I found extremely useful was 'Advanced Information Security' — it helped me get a better understanding of information security and the threats to the machine learning systems. Periodically, I also attend the seminars of School of Computing and Graduate School of AI at KAIST, Institute for Advanced Study (IAS), and Deep Learning: Classics and Trends (DLCT) for keeping myself updated about the research advancements in the area of machine learning and deep learning. During my third year in the undergraduate program, I founded the Programming Club at my college, where I often taught the concepts of machine learning and organized programming contests. It holds the record for the highest number of members (more than 700) and sponsors in the history of the college.

## 4 Industrial Experience

I spent close to three years of my undergraduate degree contributing and mentoring in various open-source organizations, including KDE, FossAsia, and Coala. With my strong affection for teaching, I made an educational module in GCompris, a high-quality educational software suite, including a large number of activities for children aged 2 to 10. To this end, my developed module teaches kids to categorize 30 categories of images and words. I also designed a module to help kids play the piano and identify its notes. During my undergraduate sophomore year, I worked at StemLending as a Full Stack Developer for their mortgage services. Some of my key contributions include enhancement of the security and efficiency of their system using orchestration, design of the REpresentational State Transfer for their Application Programming Interface, and implementation of the unit tests.

## 5 Conclusion

In conclusion, these rich research and industry experiences make me a perfect candidate for being a doctoral student at IMPRS-IS. At IMPRS-IS, I hope to get an opportunity to explore the edges of machine learning research and make the existing methods sustainable for practical deployment. I also believe that my prior coursework and research experience will allow me to get the most out of a graduate career at IMPRS-IS. The resources and breadth of the machine learning research performed at IMPRS-IS will provide me with the experience that I need to pursue my passion for learning and contributing to this incredible and relevant area of science. I have been extremely fortunate to have been introduced to research in the field, and I am grateful to all my diverse and exceptional group of mentors for their guidance and for fostering deep appreciation and fascination within me.

My most sincere gratitude and appreciation for both your time and consideration.

## REFERENCES

Francesco Croce and Matthias Hein. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *ICML*, 2020.

Logan Engstrom, Andrew Ilyas, and Anish Athalye. Evaluating and understanding the robustness of adversarial logit pairing. *arXiv preprint arXiv:1807.10272*, 2018.

Xavier Gastaldi. Shake-shake regularization. *arXiv preprint arXiv:1705.07485*, 2017.

Aidan N. Gomez, Ivan Zhang, Kevin Swersky, Yarin Gal, and Geoffrey E. Hinton. Learning sparse networks using targeted dropout. *arXiv preprint arXiv:1905.13678*, 2019.

Alex Graves. Adaptive computation time for recurrent neural networks. *arXiv preprint arXiv:1603.08983*, 2016.

Harini Kannan, Alexey Kurakin, and Ian Goodfellow. Adversarial logit pairing. *arXiv preprint arXiv:1803.06373*, 2018.

D. Madaan, R. Dua, P. Mukherjee, and B. Lall. Vayuanukulani: Adaptive memory networks for air pollution forecasting. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019.

Divyam Madaan, Jinwoo Shin, and Sung Ju Hwang. Learning to generate noise for robustness against multiple perturbations. In *NeurIPS MetaLearning Workshop*, 2020a.

Divyam Madaan, Jinwoo Shin, and Sung Ju Hwang. Adversarial neural pruning with latent vulnerability suppression. In *ICML*, 2020b.

Evgenia Rusak, Lukas Schott, Roland Zimmermann, Julian Bitterwolf, Oliver Bringmann, Matthias Bethge, and Wieland Brendel. A simple way to make neural networks robust against diverse image corruptions. In *ECCV*, 2020.