+91-9953421233
neelabh.madan@gmail.com
neelabh-madan
neelabh17
UDJZ6bYAAAAJ

Neelabh Madan

Education

2018–2022 **Indian Institute of Technology Delhi**, *Hauz Khas*, Delhi, India
B.Tech (Bachelors) in Mechanical Engineering – CGPA: 9.3/10 (Rank: 8/75)
with Minor Degree in Computer Science – CGPA 9.75/10

Publications

*C = Conference, P = Preprint/Submitted — * implies equal/core contribution*

C.1 **A Stitch in Time Saves Nine: A Train-Time Regularizing Loss for Improved Neural Network Calibration**
Ramya Hebbalaguppe*, **Neelabh Madan***, Jatin Prakash*, Chetan Arora
CVPR 2022 (Oral presentation) [Paper]

P.1 **Enhancing Tail Performance in Extreme Classifiers by Label Variance Reduction**
Anirudh Buvanesh*, Rahul Chand*, Jatin Prakash*, Bhawna Paliwal, Mudit Dhawan, **Neelabh Madan**, Manik Varma et al.
Under review at ICLR 2024 [Review Scores (Out of 10): 8,6,6,5]

P.2 **A Teacher Affects Eternity: A New State-of-the-art for Calibration using Knowledge Distillation**
Ramya Hebbalaguppe*, Mayank Baranwal*, Jatin Prakash*, **Neelabh Madan**, Kartik Anand, Chetan Arora
Under review at CVPR 2024

Scholastic Achievements and Awards

- **JEE Advanced 2018:** Secured All India Percentile of 99.7 in JEE Advanced 2018 among the 12 Lakh
- **INOI 2017:** Cleared ZIO and qualified for INOI 2017 (Indian National Olympiad in Informatics)
- Received **Prof. S.K. Saha's Award for Best Report for a Robotics Project** and cash prize of $150
- Received $20 000 from Micron Technology for winning the International Micron UV Design Challenge
- Received **Engineering Excellence Award** (2023-2024) from Microsoft for outstanding product impact
- Received Top 7% merit certificate for excellent performance in 4^{th} semester with a cash reward of 2500

Research Experience

July'22-Present **Microsoft Research**, *eXtreme Classification **(XC)** group*, Bengaluru, India
Pre-Doctoral Research Fellow
Advisors: Manik Varma, Amit Sharma
- Worked on Personalized Recommendation by introducing eXtreme classifiers (XC)
- Enhance tail performance of XC models [Under review ICLR 2024]
- Distilling GPT4 to highly-efficient XC models
- Deploying large-scale XC models on Microsoft Audience Network (MSAN) platform

May'22-July'22 **Tata Consultancy Services (TCS) Research**, Noida, India
Research Intern
Advisor: Ramya Hebbalaguppe
- Knowledge Distillation and Calibration [Under Review CVPR 2024]

May'21-Aug'21 **Adobe Research (Media and Data Science Research Lab - MDSR)**, Noida, India
Research Intern
Advisor: K Balaji
- Worked on Document Visual Question Answering (DocVQA); Augmented LayoutLMV2 architecture with visual, textual, and layout modalities for a high-level understanding of documents.

Jul'21-July'22 **Indian Institute of Technology**, Delhi, India
Undergraduate Student
Advisors: Chetan Arora, Arnob Ghosh
- Worked with Prof. Chetan on "A Stitch in Time Saves Nine" as a project work [Accepted CVPR 2022 (ORAL)]
- Investigated, with Prof Arnob, the consequences of partial and full information on Contextual Multi-Armed Bandit as part of B.Tech Thesis; Studied the effect of various non linear loss functions on NeuralUCB algorithm

Real World Deployments

Jul'22-Present **Improved Personalized Ad Recommendation using XC models on MSAN**, *(Microsoft Research)*

- Modeled Personalized Ad recommendation as an XC task; trained large eXtreme Classifier models to improve performance over previously deployed Siamese style models on Microsoft Audience Network (MSAN) platform.
- Achieve **200-400% gains** in offline recall metrics, which resulted in **20+ mainstreamings**, generating approx. 200$ revenue for Microsoft and higher click-through-rates (CTR)
- Deployed models were able to predict ads, extending beyond just lexical matches; **Extended Ads coverage to approx 1 Billion Ads** by training classifiers on meta-clusters of Ads
- Received the **Engineering Excellence Award** for the Financial Year of 2023-2024

Selected Research Projects

Sept'23-Present **Data Distillation for Sequential Recomendation using LLMs**, *(Microsoft Research)*
Advisors: Amit Sharma, Manik Varma — [In Progress]

- Used LLM (GPT-4) to distill down a smaller LM to summarize long sequential user histories (made up of events such as bing searches, websites visited, etc) into shorter text representation for personalized recommendation.
- Shorter yet diverse textual representation of long user histories enables retention of multiple intent in the user embedding generated by BERT like encoders. This is an improvement over previously used methods that embed-then-aggregate individual user events.

May'23-Present **Infusing world knowledge in efficient retrieval systems**, *(Microsoft Research)*
Advisors: Amit Sharma, Manik Varma — [Planning to submit to ICML 2024]

- The LLM (GPT-4) was utilized to distill a smaller LM, focusing on the enriching the data with diversified intents (which were previously absent) using additional information, leading to better quality retrievals
- GPT-4 can generate this extra information, however it is not scalable. Therefore we distill a smaller LM to extract information in the form of psuedo-labels/queries and intermediates from sources of unstructured data like webpages. These psuedo-labels/queries and intermediates are used in various ways to train superior models

Dec'22-May'23 **Enhancing XC models on tail labels**, *(Microsoft Research)*
Advisors: Manik Varma — [under review ICLR 2024; Review Scores:(8,6,6,5)/10]

- XC models (consisting of one-vs-all linear classifiers) are notorious for sub-par performance on labels that do not have enough training points (tail labels). Siamese models, however, have superior performance on tail labels
- Exploiting the above, we proposed a framework that improves tail performance of XC classifiers by distilling tail-robust encoders in the form of soft-label targets, resulting in SOTA (upto 5% absolute increase in metrics)

July'21-May'23 **Uniting Knowledge Distillation (KD) with Calibration**, *(Indian Institute of Technology Delhi)*
Advisor: Prof. Chetan Arora — [Under review CVPR 2024]

- Proposed a simple framework to calibrate models by distilling calibrated teachers in order to get calibrated students. These calibrated students produced have SOTA calibration as compared to previous methods.
- Explored KD with the lens of calibration and verified that only certain teachers can distill calibrated students. Additionally, we showed that even distilling from smaller teachers can lead to calibration in larger students.

July'21-May'23 **Calibration of deep neural networks**, *(Indian Institute of Technology Delhi)*
Advisor: Prof. Chetan Arora — [CVPR 2022 ORAL]

- Proposed a novel trainable calibration method that calibrates all predicted classes (unlike other methods that focuses only on the maximum one). Introduced an auxiliary loss term that can be used in a modular manner.
- MDCA (proposed auxiliary loss) performs better or on par with the then SOTA methods (Focal Loss, Label Smoothing, etc.) by reducing both top-class (ECE) and multi-class (SCE) miscalibration metrics.

Teaching

Jan-Jul'2020 **Introduction to Computer Science**, *Academic Mentor*

Extra Curricular

Sports Secured **3rd Position (2019) and 1st Position (2020)** at **Inter Hostel Football Championship**
Won the overall General Championship, IITD (2020)

Music Played **Sitar** and performed in **School Orchestra** for 4 years

Coding Team Member at **Algorithms and Coding Club, IITD** and a regular competitive programmer

Robotics Participated in DD ABU ROBOCON **(2019)**, DD ABU ROBOCON **(2020)**, DD ABU ROBOCON **(2021)**
Won the **International Micron UV Design Challenge** against **MIT** and **Georgia Tech (2020)**

Leadership **Overall Coordinator (Robotics Club):** Supervised 150 members and revised Club structure | Lead the club through collaborations with Innovation Hub for Cobotics (IHFC), Lumos Labs and Tech Analogy and organised robotics workshops and talks.
Organised 6 freshmen summer projects on various fields of robotics. View Control Simulator Project