J. Madhava Prasath

Residence: 135b, T.H.Road Minjur 601203

E-mail: madhavprasath088@gmail.com   *Phone number:* 6379319226

Date of birth: 23-06-2004

Github (Games): Madhavaprasath

Website: madhavaprasath23.github.io

Objective

To be successful in a culture of excellence and learn industry-ready abilities that I can use in my future profession, as well as to know theoretical concepts and make a use full contribution to the feild am interested in.

Education

B.Tech in Artificial Intelligence and Machine learning	*R.M.D. Engineering college*

Bachelor' degree
Currently pursuing Honors
Related Topics: AI, Neural networks, Machine learning, Deep learning, Reinforcement learning, Natural language processing, Computer vision
CGPA: 8.03 (up to 6th semester)
(2025)

Class 12th from CBSE Board	*Velammal International School*

Computer science
Percentage: 92.4%
(2020)

Class 10th from CBSE Board	*Velammal International School*

Percentage:83%
(2018)

Skills

Machine Learning, Deeplearning , Reinforcment learning , Multimodal distribution alignment, Self Supervised learning/Representation Learning, Semi supervised Learning, Distributed training (FSDP,DDP,TP), Distributed Computing (Apache spark) ,Statistics and Probability, Game development using Godot Engine

Technical skills

Programming Languages/Tools	C, Java, Python, R, Julia (beginner level), GD-Script(Godot Game Engine)
Machine Learning Frameworks Known	TensorFlow, Pytorch, Pytorch-lighting, Scikit-learn, matplotlib, Seaborn, Plotly XG boost, lightGBM, OpenCV, open slide (medical images), NumPy, JAX, Flask, DASK,RAPIDS (cuML,cuDF).

Achievements

- Won second place in the Inter-Departmental Paper Presentation for our paper on "Extreme Gradient Approach to Classify Histopathology Images," where we used t-SNE to reduce the higher dimension vectors from Convolution Neural Network output to a lower dimension and used extreme gradient boosting methods to classify examples of histopathology images.

- Won the best paper award on the topic mentioned above in the Thesis Precized event organised by the computer science and business systems department at Chennai Institute of Technology.

- Won Third place in ML boot camp organized by Anna university, Chennai.

- Won Innovative Game award in Global Game Jam 2024 Chennai Edition

- Won 1st on interdepartmental AI Odyssey contest where making recommendation system in a short duration using suprise-scikit library

Participation

- Participated in a 3-day hackathon named "I-Hackfest" organised by PSG Institute of Technology and Applied Research, Coimbatore. Where we made a game on inclusion and diversity.

- Participated in 24-hour hackathon conducted by Vellore Institute of Technology, Chennai. We created a Chatbot to assist naive users with college admissions.

- Participated in a 48-hour AIML hackathon conducted by Open Weaver in collaboration with Vellore Institute of Technology. We made cyclone detection based on convolutional neural networks.

- Participated in a 72-hour game jam conducted by Weloadin Studios and Next Wave Multimedia. We made a game that is based on evolution and survival.

Projects

1) AUTOMATIC ATTENDENCE SYSTEM USING MACHINE LEARNING

- In order to detect faces, we trained the model using the LBHF face classifier and the Viola-Jones algorithm. We then created a prototype website using Flask that allowed teachers or attendees to manually check and correct any machine-generated errors in attendance.

- Future improvements will involve training the YOLO model, which is trained on Wider Face data set, and using the faces that are detected to identify students even if they have a beard or moustache.

2) CLASSIFYING HISTOPATHOLOGY IMAGES USING NOVEL PROCESS CALLED ATTENTION EXPLOITATION

- Originally We thought this to be taking the parts of PVT-V2 to and inject query and value part of it similar to LoRA and pre train (use SSL) techniques in discarded patches from extraction from YALE cohort dataset but failed to do that due to resource on training on large data so we have downgraded our project which will be discussed in below paragraph

- In this study, we aligned the representations from the Vision transformer and Resnet 18 with minimal parameters, using a novel approach called Attention Exploitation.

- Overall, our model was able to generalize across all of BreakHis Dataset's zoom levels. We trained a single model by combining the data from all zoom levels, and it outperformed all the models that were trained on individual zoom levels.

3) MIXTURE OF GRADIENT BASED VOTERS AND ATTENTION BASED FEED FORWARD NETWORKS IN SMALL VISION TRANSFORMER(ON GOING)

- The idea for this project came from Mixtral, a combination of expert-based feed forward networks, where we experimented with using Lambda-infused output that had a gradient linked to it, allowing the feed forward networks to vote on it for a specific class.

- We developed the theory and training plan for the model building. To learn the representations from the images, we will train on Imagenet 64. Afterwards, we will fine-tune the model using smaller datasets like CIFAR-100 and CIFAR-10.

4) DIFFUSION MODELS FOR SEMI SUPERVISED LEARNING

- We have created Hourglass diffusion Transformer like network using Mamba models where we employ Channel based scan and Dimension based scan on each Mamba model.

- DDP Training strategy of using GPUs available in the server with multi step training process 1) Pre-training on the annotated which are labeled 2) Downstreaming on particular data with hate or non hate speech

- Making a Mamba based discriminator to classify the pseudo labels generated by the diffusion process

Internships

Internship at Revature (Skills: NLP, Transformer workflow, Parametric efficient fine tuning)

- Worked as Machine Learning Engineer at Revature India Consultancy PVT ltd, Where we used transformers to evaluate the persons answer with Baseline answer provided by the company. We have used Bert and Family of Bert models. We have taken NLI (Natural Language Inference) weights which are publicly available and pretrained on STS-B data set to score the similarity between sentences and fine tuned on particular Baseline answer and User given answer. We have integrated with existing tech stack from the company with newer model providing the aggregate score. Further more we have also made Lighting Models for future training with multiple GPUs (Graphics processing unit) and TPUs (Tensor Processing Unit)

Internship at IIT-Indore (Skills: Semi Supervised Learning, Survey, Paper Reading, Distributed Training)

- Working Under a Dr Nagendra Kumar in IIT-Indore working on Making a Mamba based Diffusion model for Semi Supervised Learning .

- Published a Survey Paper on Discussing the Impact of Mamba v1 (s6 - SSMs) on Medical Image Analysis, and Beyond (arXiV)

Things which am Working on in this Internship mentioned below

- Semi supervised diffusion model which is in turn inspired from Hour Glass diffusion Transformers we wrote the entire diffusion model from scratch to generate fakes from 1D data which also includes on sampling procedures such as DDIM for noise corruption on 1D data

- Multi GPU training strategy such as DDP (since the model can be fit in one GPU) is used for training the model and one dedicated GPU for just sampling Those 1D images with the help of BERT model

- Another strategy is also used as GPUs are needed for other people who are working there in that we just use one GPU for training the Model and other 2 GPUs for sampling 1D images as mentioned above

Extracurricular Activities

- Organized an Event named Analyticon where a participant has to provide ML solution for the particular data set
- Overall Technical coordinator and Co Organized a Event Named Neuramatrix where participants need to build ML model on suprise dataset on Avinyaa 24 (National Level Technical Symposium) Organized by the department of AIML RMDEC

Volunteering

Part of Paper reading and Paper organizing in AI4Bharat Community discord server. where we conduct session which we both moderate as well as partake on paper reading session some of highlights of the session which I partook mentioned below

- Multi token prediction Making small scale Training the LLM from scratch with Indic BART embedding layer and with silapathigaram Data which eventually fitted in one GPU of collab instance training on bfloat 16 precision from scratch
- Llama -3.1 3 part Series where I discussed about How Llama models data are collected, trained and aligned with human preference in detail
- KV-Cacheing on Nano GPT - Pratical implementation of KV cacheing (still on works fixing bug) to make the inference of Transformers much faster

Certifications

1) **Machine Learning Specialization(Coursera):**

   a) Supervised Machine Learning: Regression and Classification.

   b) Advanced Learning Algorithms.

   c) Unsupervised Learning, Recommenders, Reinforcement Learning.

2) **Deep Learning(Coursera):**

   a) Neural Networks and Deep Learning.

   b) Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization.

   c) Sequence Models.

   d) Convolutional Neural Network.

   e) Structuring Machine Learning Projects.

3) **Deep Learning (NPTEL)**

4) **Data Science (NPTEL)**

   a) Python for Data Science.

   b) Data science for Engineers.

5) **AI for Medical Diagnosis (Coursera)**

6) **GANS specialization (Coursera)**

7) **Applied Accelerated Artificial Intelligence (NPTEL)**

8) **Generative AI with Large Language Models (Coursera)**

9) Sample-based Learning Methods in Reinforcement learning (Coursera)

10) Fundamentals of Reinforcement Learning (Coursera)

11) Policy and Control Function approximation (Coursera)