Lam – Annotated Bibliography

Idea 1: Using Machine Learning Algorithms to predict air pollution

Paper 1:

Dynamic Forecasting Of Air Pollution In Delhi Zone Using Machine Learning Algorithm

Anurag Sinha, and Shubham Singh. 2021. "Dynamic Forecasting Of Air Pollution In Delhi Zone Using Machine Learning Algorithm". In *Quantum Journal of Engineering, Science and Technology,* 40-53. Https://qjoest.com/index.php/qjoest/article/view/28/24

- Refer back to Djalalova, Guo, and others as the inspirations for this paper
- Mentions Box, Li about the model
- Lists 8 methods to estimate outdoor air pollution levels
- Dataset used is a little bit small.

This paper has a deep background and state a focus on particulate matter 2.5 as one of the factors leading to air pollution levels, which is similar to what I will investigate on my research. Also, machine learning models are done on Python libraries, which enables me to implement my proposal should I choose this topic. Dataset is clearly displayed and documented; however, as it is relatively small, I would love to find other bigger data in Kaggle that also measure air pollution detecting air pollution components in other places besides India - the place subject of this paper. Lastly, the result details of different techniques will be a great source as I focus on comparing the efficiency between neural network and random forest.

Paper 2:

Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models

Doreswamy, Harishkumar K S1, Yogesh KM, and Ibrahim Gad. 2020. "Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models". In *Procedia Computer Science,* 2057-2066.
https://reader.elsevier.com/reader/sd/pii/S1877050920312060?token=02989FA2E4BF2766207

240EA2C6B63ECBD7D7981ADF7CFB138D118373128DECF2119719D479905306DCB1A41EA6FB3A2&originRegion=us-east-1&originCreation=20210824001014

- Every machine learning algorithm is used to predict air pollution level
- The paper chose a portion of Taiwan to collect data.
- Statistical models: RMSE, MAE, MSE, R-square

This paper brings up a lot of statistics and results in numbers from different machine learning algorithms (including my proposed random forest), which can serve as useful information to compare with not only other methods from the paper but also the neural network, which is another subject to be compared on my research. That the research looks at PM 2.5 helps with my research as well because the subject to dive in is the same. Also, the detail image description of how researchers set up the model structure diagram could be a suggestion for me to make my capstone project better.

Paper 3:

Evolving Differential evolution method with random forest for prediction of Air Pollution

Rubal, and Dinesh Kumar. 2018. "Evolving Differential evolution method with random forest for prediction of Air Pollution". In *Procedia Computer Science,* 824-833. https://reader.elsevier.com/reader/sd/pii/S1877050918308263?token=C957332DEE57D11BE7FCC8C31A2D80D4E99446831057B44EA5A4701F360CC6295541B322895330F3BB7A71D2E9357BA3&originRegion=us-east-1&originCreation=20210824003139

- Mentions other existing techniques to predict level of air pollution (independent, multi-label)
- Look up: Differential Evolution
- Check the source of the dataset for accessibility

This paper is also in the range of my research topic as it mentions random forest as a technique to predict air pollution. The data visualization is more beneficial to me to compare between random forest and neural network, which is the main part of my research for capstone project. The informative flow chart diagram helps me understand the structure of the models. This

paper also sketches a pseudocode of the algorithms that will be a great source to build a software that supports my capstone project.

Paper 4:

Air Pollution Prediction with Multi-Modal Data and Deep Neural Networks

Jovan Kalajdjieski, Eftim Zdravevski, Roberto Corizzo, Petre Lameski, Slobodan Kalajdziski, Ivan Miguel Pires, Nuno M. Garcia, and Vladimir Trajkovik. 2020. “Air Pollution Prediction with Multi-Modal Data and Deep Neural Networks”. In *Multidisciplinary Digital Publishing Institute*, 1-19. https://www.mdpi.com/2072-4292/12/24/4142

- List some related works to predict air pollution level
- The research takes place in Macedonia
- Check: World Weather Online website for dataset validity

This paper focuses on convolutional neural network (CNN) as an approach to predict air pollution level. Forecasting the level of pollutants based on the image is an interesting field and that this method uses Keras library will be an inspiration for me to do the same with other cities on my project, especially in Vietnam. The diagram of instruction and procedures are well-designed and the results are carefully demonstrated as well. The paper also lists PM 2.5 as one of the pollutant factors, which is helpful for my project because it is my intended subject to do research for my project.

Paper 5:

PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data

Mehdi Zamani Joharestani, Chunxiang Cao, Xiliang Ni, Barjeece Bashir, and Somayeh Talebiesfandarani. 2019. “PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data”. In *Multidisciplinary Digital Publishing Institute*, 1-19. https://www.mdpi.com/2073-4433/10/7/373

- Compare the performance of different machine and deep learning techniques

- Check the data sources
- Future work: high spatial resolution

This paper includes different machine learning and deep learning techniques so that it will strengthen my arguments about comparing the performance between 2 intended methods: random forest or neural network. The data visualization and result demonstration are well-crafted. It also mentions PM 2.5 as one of factors that creates air pollutant, which also supports my project as I am doing research on this subject as well.

Paper 6:

RAQ–A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems

Ruiyun Yu, Yu Yang, Leyou Yang, Guangjie Han, and Oguti Ann Move. 2016. "RAQ–A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems". In *Multidisciplinary Digital Publishing Institute,* 1-18. https://www.mdpi.com/1424-8220/16/1/86

- Random Forest classification is part of RAQ Algorithm
- RAQ is compared with other machine learning algorithms
- Sensing system is a subject to be experimented

This paper compares random forest (a classification method of RAQ algorithm) with other machine learning techniques, which will be helpful for me to evaluate random forest and neural network on each performance with other techniques in different papers. The result visualization is elaborately manifested. Although the clear name for the dataset is not mentioned, images demonstrating the database and its description could be a great source for me to find similar datasets.

Paper 7:

Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China

Rui Feng, Hui-jun Zheng, Han Gao, An-ran Zhang, and Chong Huang. 2019. "Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case

study in Hangzhou, China". In *Journal of Cleaner Production*, 1005-1015. https://www.sciencedirect.com/science/article/pii/S0959652619318554

- Refers to Feng, Kaminska, etc on their previous works
- Look up: WRF
- 5 parameters for both random forest and neural network

This is another paper that involves comparing the efficiency in performance between random forest and neural network (in this case, recurrent), which will support my setting of arguments for my project. Data visualization are displayed in wide variety of styles, from graphs to geographic maps. The paper also focuses on listing the pollutant factors that make up the air pollutions, and PM 2.5 – my main subject of focus – is mentioned.

Idea 2: Machine Learning in music genre classification

Paper 1:

Music Genre Classification using Neural Networks

Kinjal Mhatre, Viraj Almeida, Samuel Lopes, and Garima Tripathi. 2020. "Music Genre Classification using Neural Networks". In *Mukt Shabd Journal,* 2167-2172. http://shabdbooks.com/gallery/240-may2020.pdf

- Mentions concurrent neural network as a subject of comparison
- NumPy is built to generate RNN model
- GTZAN dataset.
- Look up: LSTM
- Forecasting to classify music based on mood, lyrics, chord progression, etc.

This paper contains a detailed setup of a model using recurrent neural network (RNN). This model is built on NumPy, which inspires me to learn more on how to build a neural network model. With the RNN being the subject of this research, this paper supports my plan to do a

research on the idea of music genre classification, which is to compare CNN and RNN. Also, the clear dataset description makes it easier for me to find other datasets with the similar type.

Paper 2:

Comparative Analysis of Three Improved Deep Learning Architectures for Music Genre Classification

Quazi Ghulam Rafi, Mohammed Noman, Sadia Zahin Prodhan, and Sabrina Alam. 2020. "Comparative Analysis of Three Improved Deep Learning Architectures for Music Genre Classification". In *International Journal of Information Technology and Computer Science*, 1-14. https://www.researchgate.net/profile/Quazi_Ghulam_Rafi/publication/350605124_Comparative_Analysis_of_Three_Improved_Deep_Learning_Architectures_for_Music_Genre_Classification/links/60685750299bf1252e24e2e8/Comparative-Analysis-of-Three-Improved-Deep-Learning-Architectures-for-Music-Genre-Classification.pdf

- Introduces improved versions of CNN and RNN
- Ballroom and Extended Ballroom: check the source information
- Look further: CRNN

This paper is an interesting one to look because it will strengthen my claim on either RNN or CNN is efficient by introducing their improved versions. The dataset information is available and I will just need to check the source, look at their components in order to find myself a new dataset of similar type. The flow chart for each method is carefully displayed and so is the data visualization for the diagram.

Paper 3:

Classifying Audio Music Genres Using CNN and RNN

Ahmed Khamees, Hani Hejazi, Muhammad Alshurideh, and Said Salloum. 2021. "Classifying Audio Music Genres Using CNN and RNN". In *Advanced Machine Learning Technologies and Applications: Proceedings of ALMTA 2021,* 315-323.

https://books.google.com/books?hl=en&lr=&id=kqkhEAAAQBAJ&oi=fnd&pg=PA315&dq=music&ots=yh0SW0b5Vf&sig=43ZQv5UadnGi-q_WJ4TWFT_GKw4#v=onepage&q=music&f=false

- Look up: LSTM and Max-pooling
- Training/Test ratio is 80/20
- Audio are displayed in a 2D array

This paper has a clear structure of only introducing the neural networks and displaying the results. Therefore, it is easier for me to catch up. Dataset are from a fixed source, so I will look at other types of datasets that are inspired from the original source. There is also a machine learning library mentioned in the process of building the models to which I can refer. The data visualization and results are great sources to compare CNN with RNN as my proposed work.

Paper 4:

Deep Learning Based Music Genre Classification Using Spectrogram

Athulya K M, and Sindhu S. 2021. "Deep Learning Based Music Genre Classification Using Spectrogram". In *SSRN Electronic Journal,* 1-7. https://www.ssrn.com/abstract=3883911

- CNN model uses a 2-D array
- The model is built using Keras from TensorFlow
- Future work: Uses RNN as a model

This paper lists some dynamic libraries such as Keras, TensorFlow, which enables me to build my own project technically in Python based on the proposed idea. The instruction diagram is simple but understandable, and I will consider adding it in my project. The data for the results is detailed enough for me to do research on the performance between CNN and RNN in general.

Paper 5:

Deep attention based music genre classification

Yang Yu, Sen Luo, Shenglan Liu, Hong Qiao, Yang Liu, and Lin Feng. 2020. "Deep attention based music genre classification". In *Neurocomputing*, 84-91. https://linkinghub.elsevier.com/retrieve/pii/S0925231219313220

- Refer to Zhang, Costa, and others' existing approaches
- Look up: GRU
- Uses Bidirectional RNN

This paper is focuses on the performance of recurrent neural network (RNN) on genre classification with its specified bidirectional method. It is also helpful to strengthen my arguments about whether CNN or RNN is better in the approach of classifying music. The results are displayed in full detail, and the visualization is elaborately manifested. The instruction in diagram is also well-documented, which will be one of the sources for my project.

Paper 6:

Music Genre Classification Using A Convolutional Neural Network

Khalil Miri, Erick Enriquez, and Aidan Donohue. "Music Genre Classification Using A Convolutional Neural Network". In *CS230*, 1-5. http://cs230.stanford.edu/projects_spring_2020/reports/38954030.pdf

- CNN method includes setting up a 2D array of audio
- Future work focuses on lack of data in the dataset.

This paper focuses on the performance of CNN, which also helps with my topic of comparing this method with RNN (with LSTM model). The information in this paper will support some of my arguments about the performance between the structures within both CNN and RNN. There are not so many visualizations; however, the results are stated on point and concisely. The dataset information and description are detailed also.

Paper 7:

Music Genre Classification: A Comparative Study Between Deep-Learning and Traditional Machine Learning Approaches

Dhevan Lau. 2020. "Music Genre Classification: A Comparative Study Between Deep-Learning and Traditional Machine Learning Approaches". In *Riteshajoodha*. 1-8. https://riteshajoodha.co.za/sitepad-data/uploads/2021/02/2020-Dhiven.pdf

- Look up: Automatic Music Classification, Automatic Music Retrieval
- Compare the neural network with traditional machine learning techniques

This paper compares the CNN with traditional machine learning techniques such as random forest, regressions, etc. This will be beneficial for me to compare to RNN by evaluating each method's performances with other traditional machine learning methods. Also, the dataset's description will help me determine the type of dataset for my project. The diagram and visualizations are also clear as well.