Mid Term Presentation of BTP project

Real Time Battery Monitoring System Using Machine Learning

Presented by

Tathya Bhatt – 191030011041

Gurpreet Singh - 191030012005

Under the guidance of

Dr. Jagat Rath

Department of Mechanical and Aerospace Engineering

Institute of Infrastructure, Technology, Research and Management

Table of contents

Recap

What we know so Far

- **Battery monitoring systems** currently rely on current and voltage measurement to estimate state of charge (SOC) but ignore the crucial parameter of temperature.
- **Traditional algorithms** like Coulomb Counting and Voltage Method **suffer from hysteresis** and do not account for battery aging, reducing accuracy.
- **Machine learning-based BMS** solutions are scarce, but could provide more efficient and **accurate results**.
- A **data-driven approach**, which considers the **effect of battery temperature**, can help overcome these challenges and improve accuracy.
- Our goal is to study different **machine learning approaches** and find the **best solution** for **accurate BMS**.

Recap

Previous Results

- LG 18650HG2 Li-ion Battery Data
- A brand new 3Ah LG HG2 cell was tested.
- A series of tests were performed at different temperatures (10℃, 25℃, 40℃) and the battery was charged after each test at 1C rate to 4.2V.
- All the three regression algorithms (Ordinal Least Square, Lasso Regression and Ridge Rigression) were giving the similar score of R-Squared of around 100%.

Challenges

- The amount of data.
- Limited Features
- Lack of comprehensive analysis

Objective

- To provide a more robust solution by incorporating more complex dataset so that the models performance is improved.

Literature Review

Sr.No	Paper Title	Authors	Summary
1.	XGBoost: A Scalable Tree Boosting System (2016) *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794	Chen, T. Guestrin, C.	• XGBoost is a powerful and flexible tool for building high-performance predictive models • One of the key advantages of XGBoost is its speed and scalability, making it suitable for large datasets and real-time applications.
2.	Prediction of soil water infiltration using multiple linear regression and random forest in a dry flood plain, eastern Iran. *CATENA*, *194*, 104715.(2020)	Pahlavan-Rad M. R. Dahmardeh K. Hadizadeh M., Keykha G. Mohammadnia N. Gangali M.Keikha M. Davatgar N.Brungard	• Random Forest found to be the best solution for predictive algorithms compared to multiple linear regression. • Can be used in working with high dimensional data
3.	A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222.(2004).	*Smola, A. J.* *Schölkopf, B.*	• SVM Linear kernel and SVM RBF are the two types of methods used if data is linear and non linear. • Support Vectors decide the hyperplane in the dataset

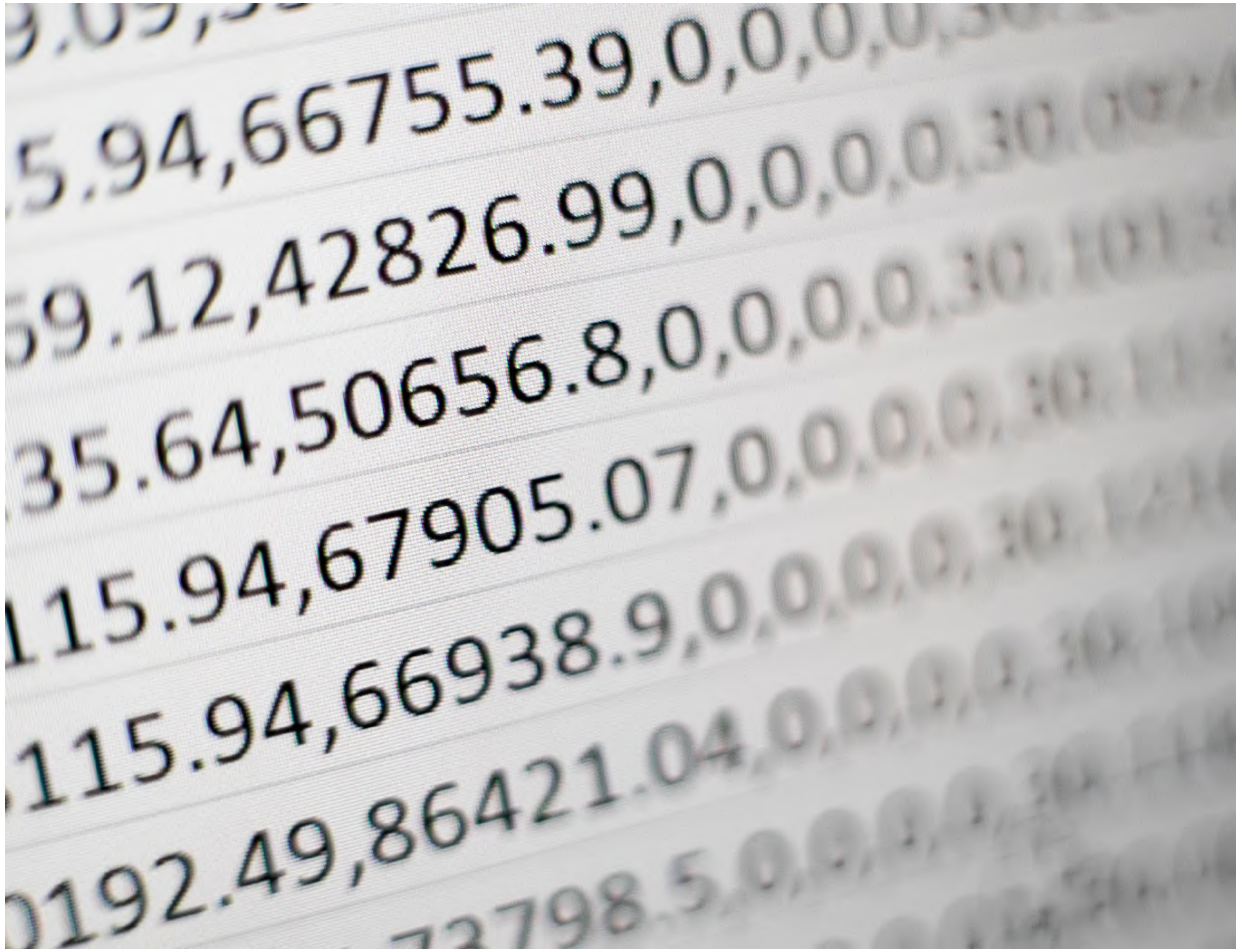
Literature Review

Sr.No	Paper Title	Authors	Summary
4.	Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science, 4*(3), 183–201.(2018).	Henrique, B. M. Sobreiro, V. A. Kimura, H.	• It is possible to obtain smaller prediction errors in the test set than in the training set when using a linear kernel • it is possible to build risk management models using SVR-based estimates
5.	Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry	Ying LU Yan-yan Song	• Decision Tree can be useful in regression and is a simple method • The main disadvantage is that it can be subject to overfitting and underfitting, particularly when using a small data set.
6.	Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 53 (2021)	*Alzubaidi, L. Zhang, J. Humaidi, A.J. et al.*	• DNN's are computational heavy and requires high power GPU • Number of neurons can be decided by trial and approach

Data Selection and preprocessing

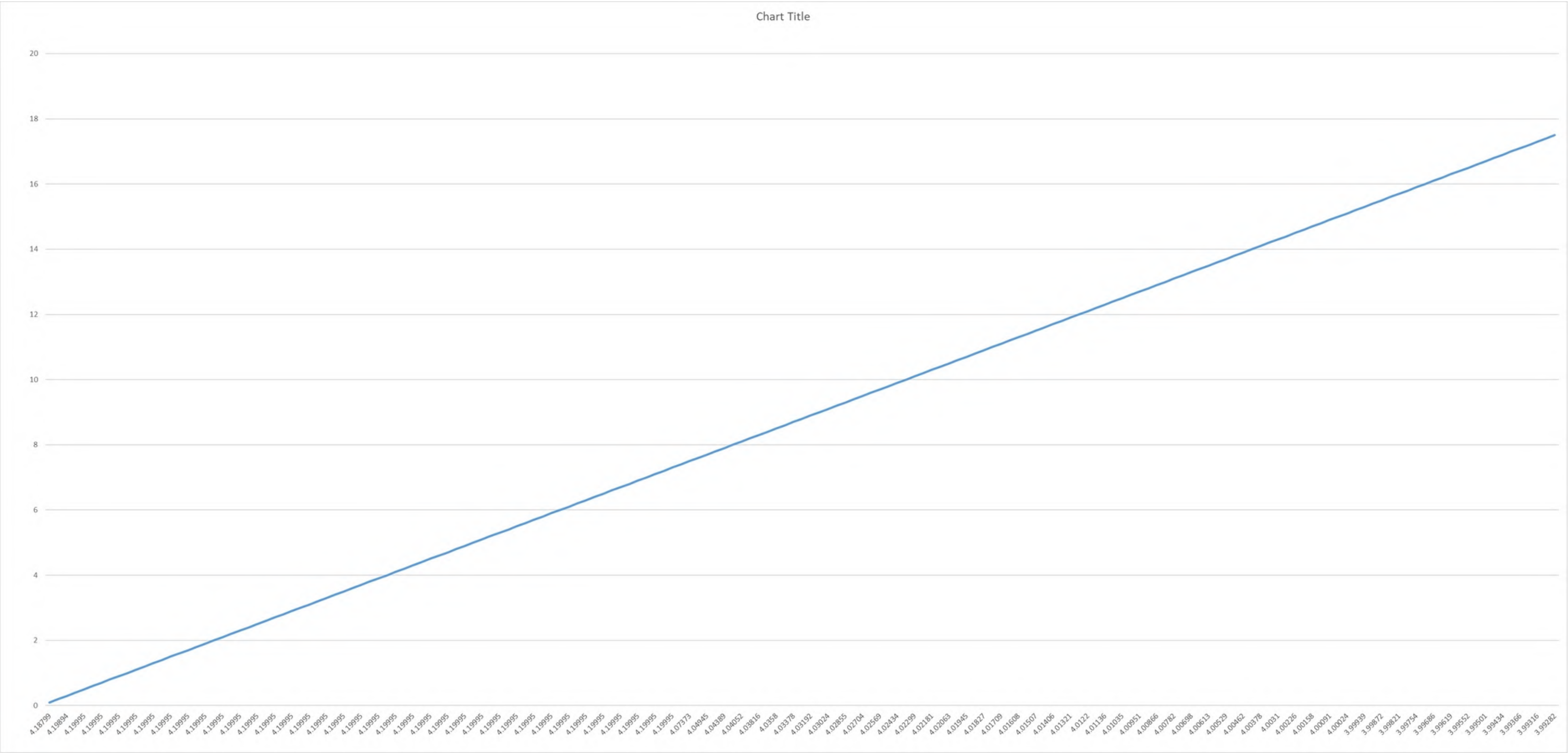
Dataset details

- eVTOL Dataset authored by Alexander Bills, Venkatasubramaniam Vishwanathan, Shashank Sripad, Evan Frank, Devin Charles, William Leif Fredericks is used to train the model
- Can be used to develop algorithms for the electric aircraft batteries
- Battery used Sony Murata - 18650 3000mAh
- Funded by Airbus
- Dataset includes varios cycles at different VTOL operating conditions.



Previous Data vs Current Dataset per cycle

Time(s)



Voltage (V)

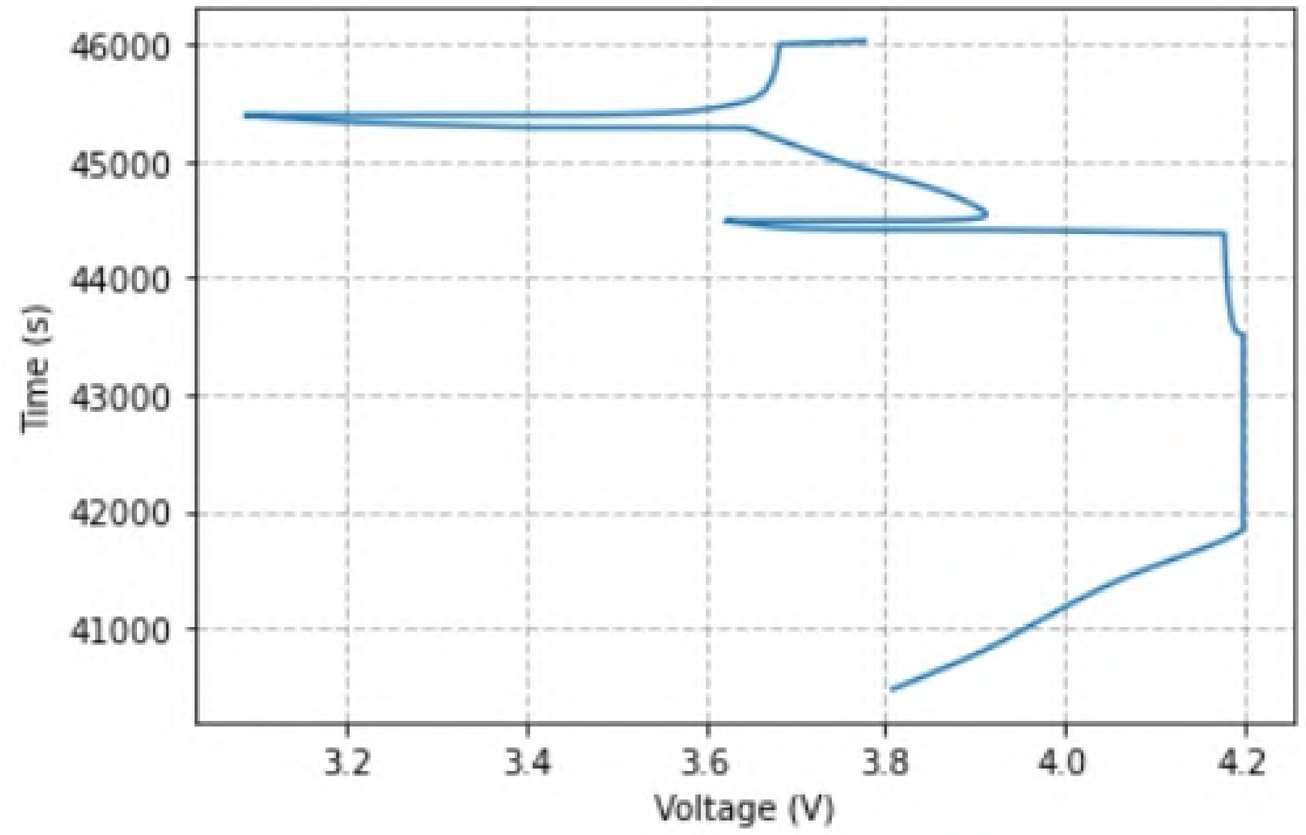


Fig. 1 - Data Cycle Comparision

Data Pre-processing

- Cleaned the data and removed unneccesary columns

- Selected 5 conditions
  - Baseline
  - Extended cruise (1000 sec)
  - 10% power reduction during discharge (flight)
  - Charge voltage reduced to 4.0V
  - Thermal chamber temperature of 20 degrees C

- Each model was trained for first 5 cycles to study the battery's behaviour for each cycle.

- Added features like
  - Mean, Median, Standard Deviation, Variance of Voltage-Temperature-Current
  - Power
  - Resistance
  - Conductance
  - Temp Change

- Normalized the dataset

- Removed columns with null values

Results

Feature Correlation

- Heatmaps are used to identify the correlation between two or more variables.
- If two or more features in a dataset are highly correlated, it can cause problems with the model's ability to learn and make accurate predictions.
- If a model is trained on a dataset with correlated features, it may be less able to generalize to other datasets with different correlations between the features

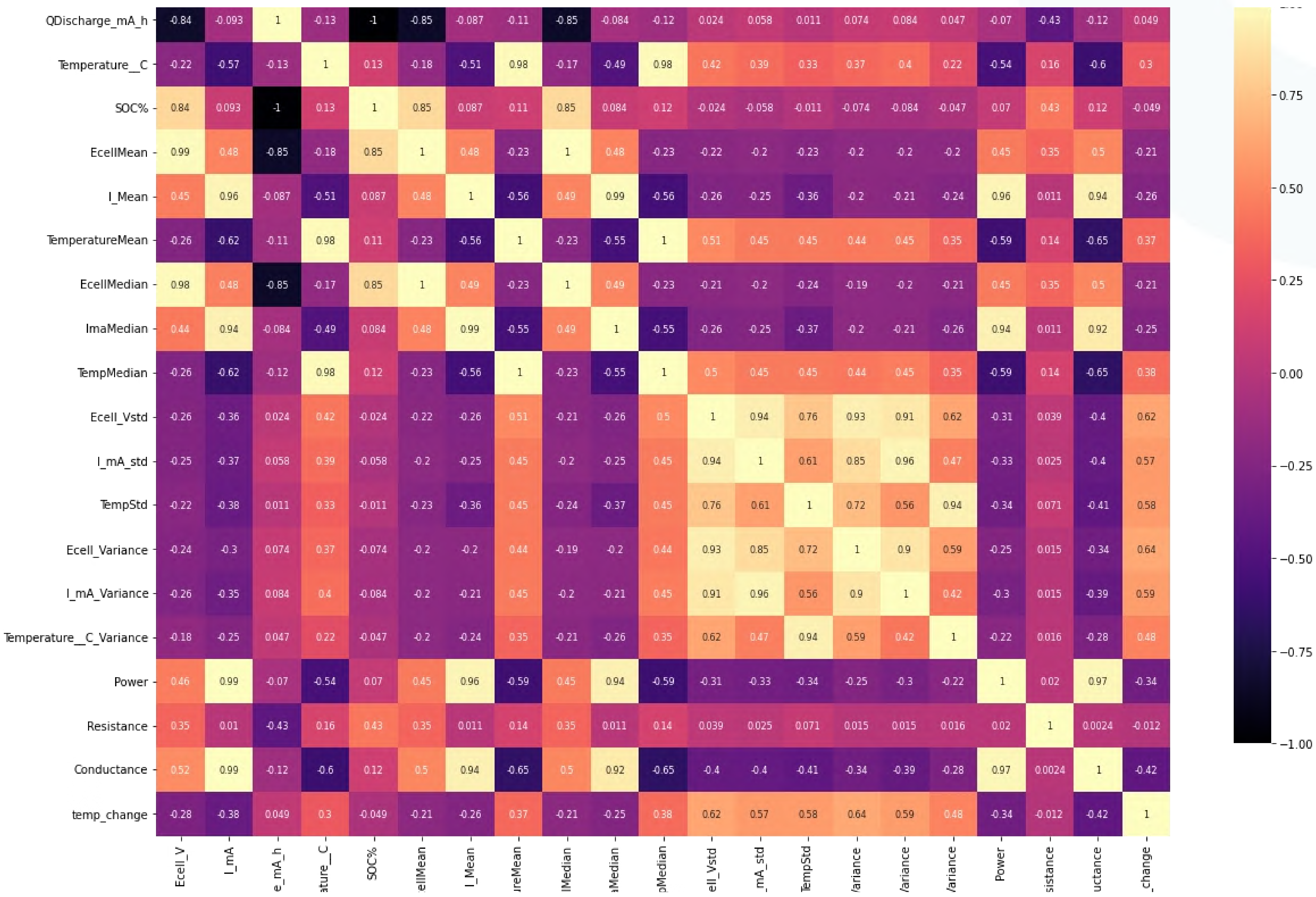


Fig. 2 - Correlation heatmap of our dataset

Analysis

Evaluation Metrics

- **Root Mean Squared Error (RMSE)**: A commonly used metric to evaluate the performance of a regression model. It measures the difference between the predicted and actual values, squared and averaged over all samples, then taking the square root. Lower values indicate better performance.

- **R-squared (R2) Score**: A statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. It ranges from 0 to 1, with higher values indicating better fit between the predicted and actual values.

- **Mean Absolute Error (MAE)**: Another commonly used metric to evaluate the performance of a regression model. It measures the average absolute difference between the predicted and actual values. Lower values indicate better performance.

Analysis

R-Squared Comparison of models with & without features (60:40)

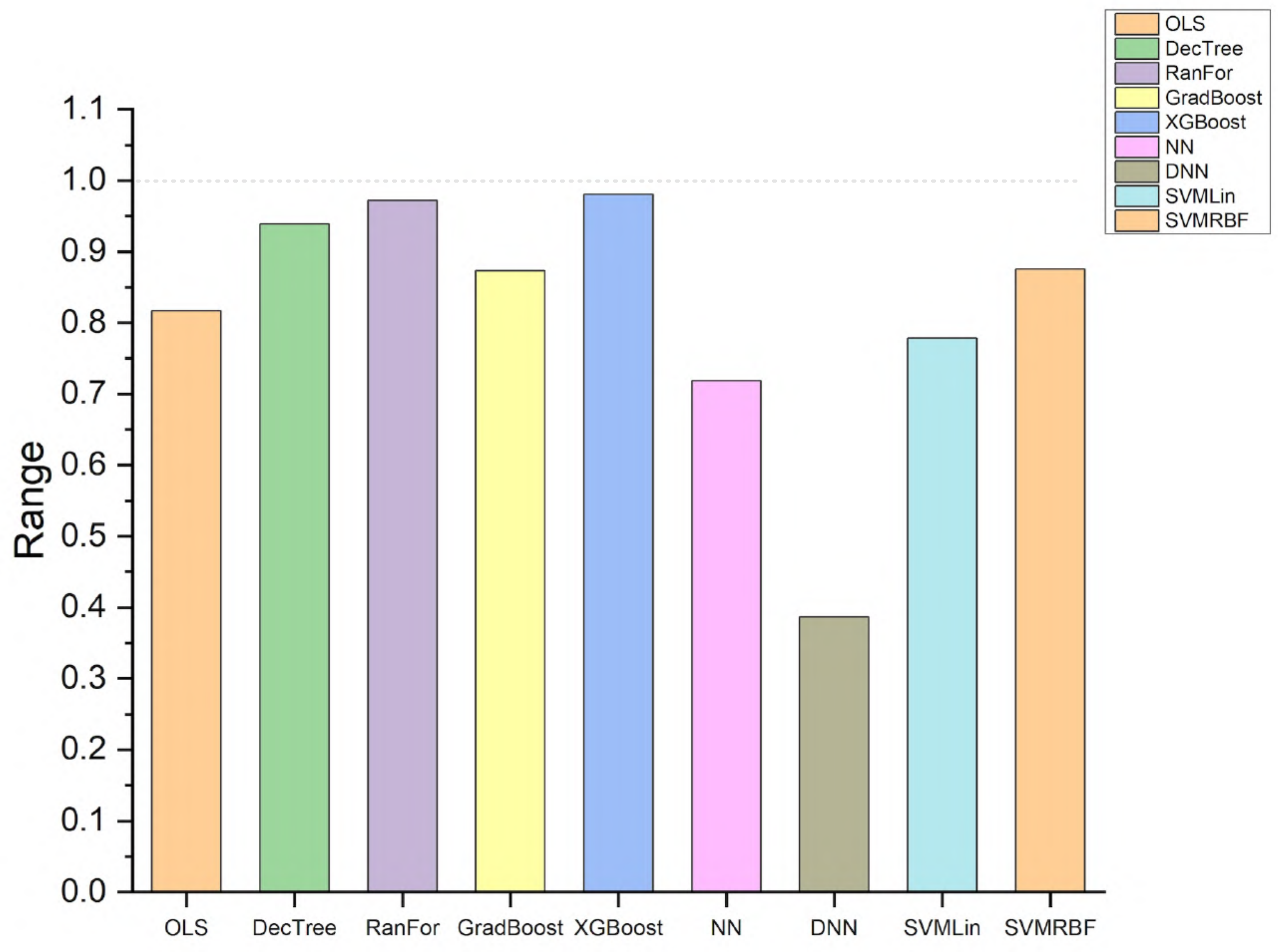


Fig3. Rsquared without features

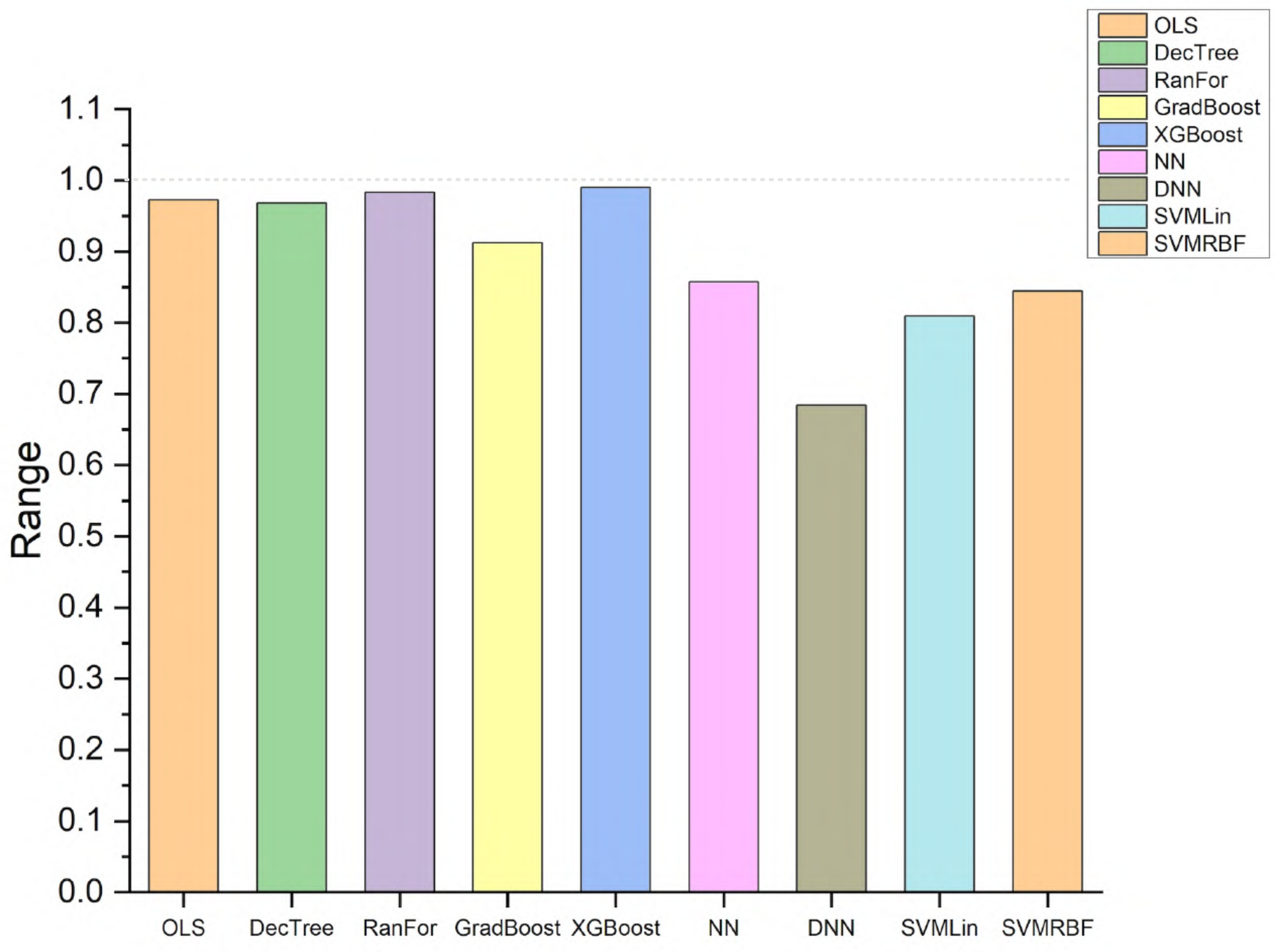


Fig4. Rsquared including features

Analysis

RMSE Comparison of 80:20 and 60:40 training ratios with features

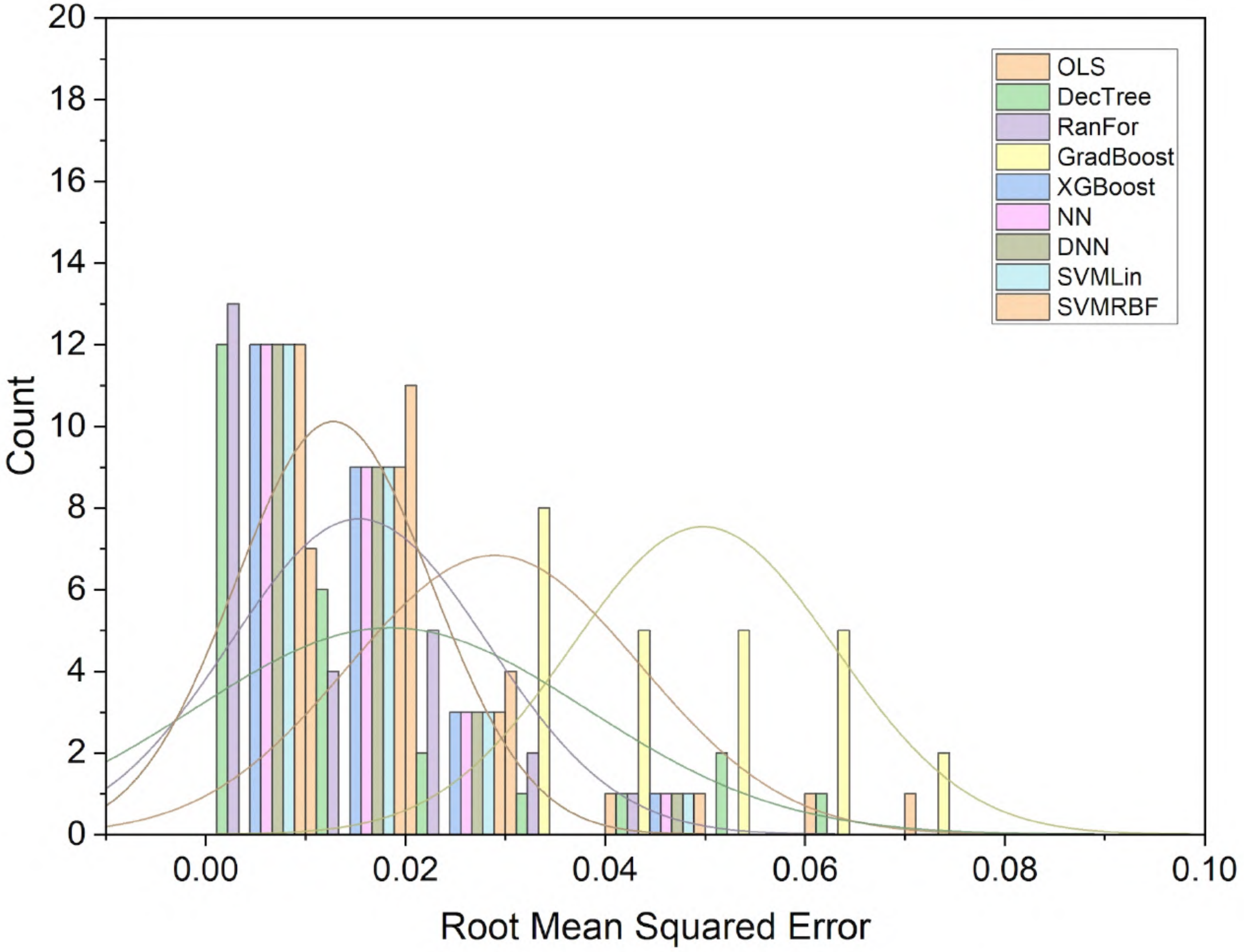


Fig5. RMSE of 60:40

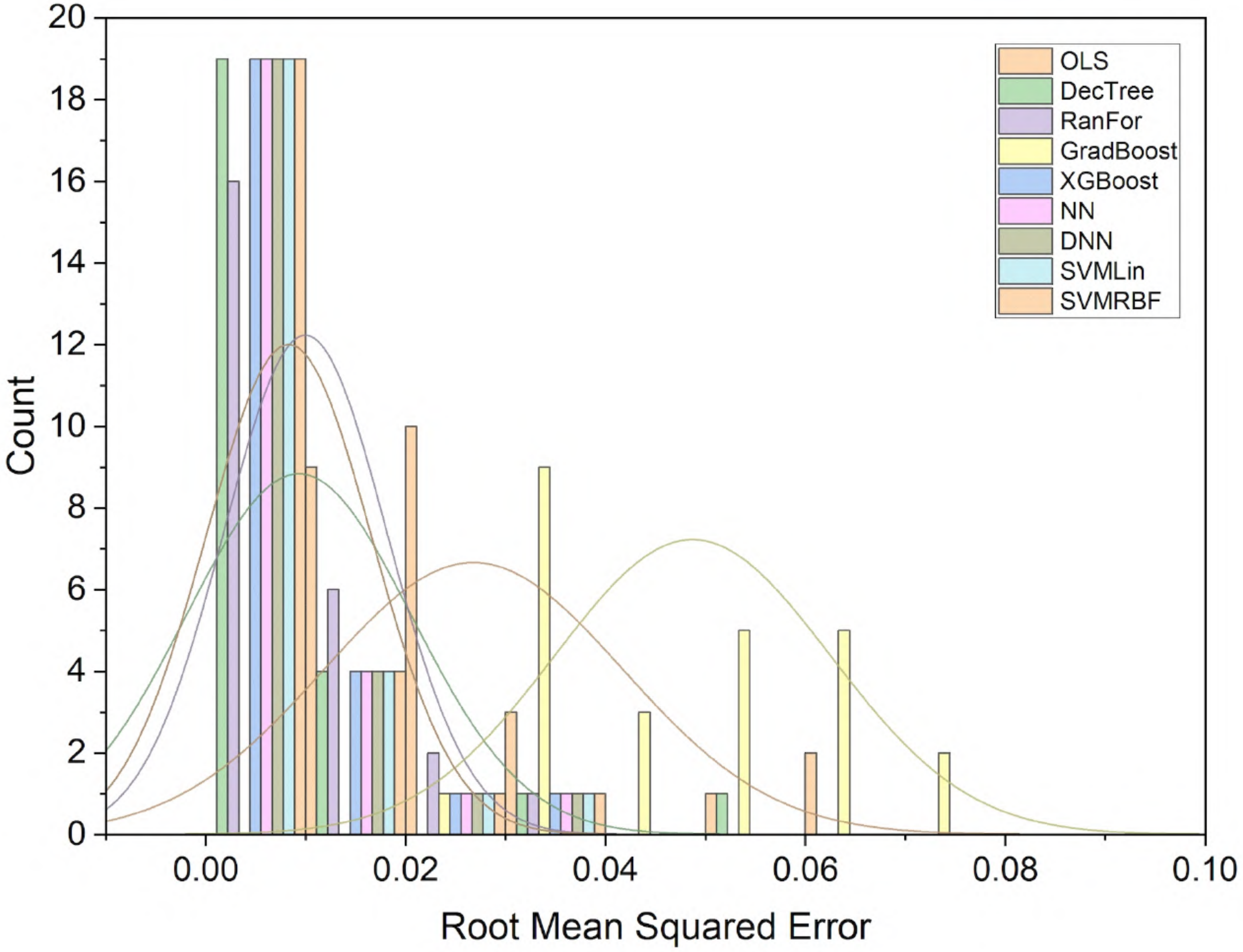


Fig6. RMSE of 80:20

Choosing 5 algorithms by comparing their MAE and RMSE

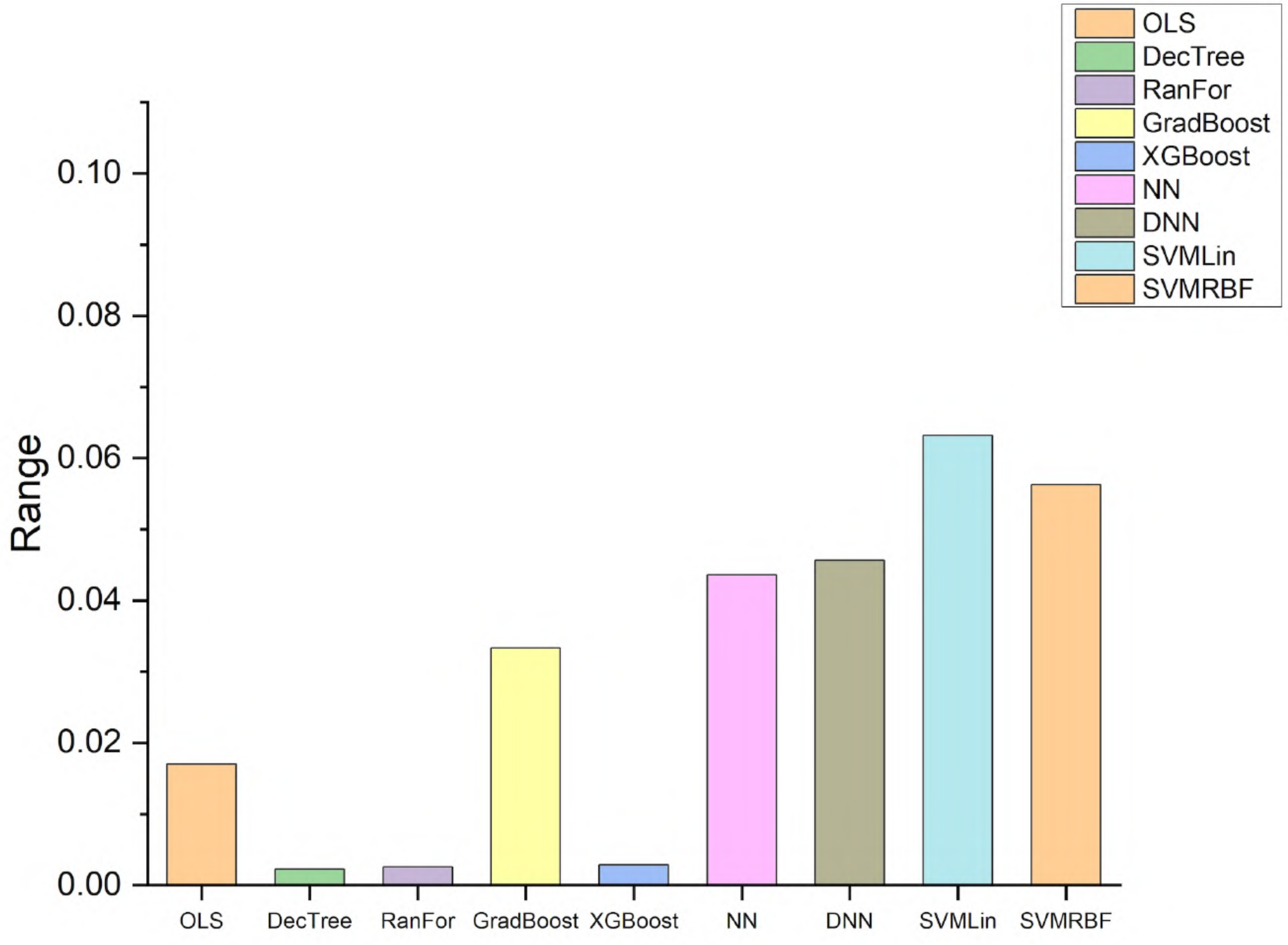


Fig7. MAE at 80:20 Ratio

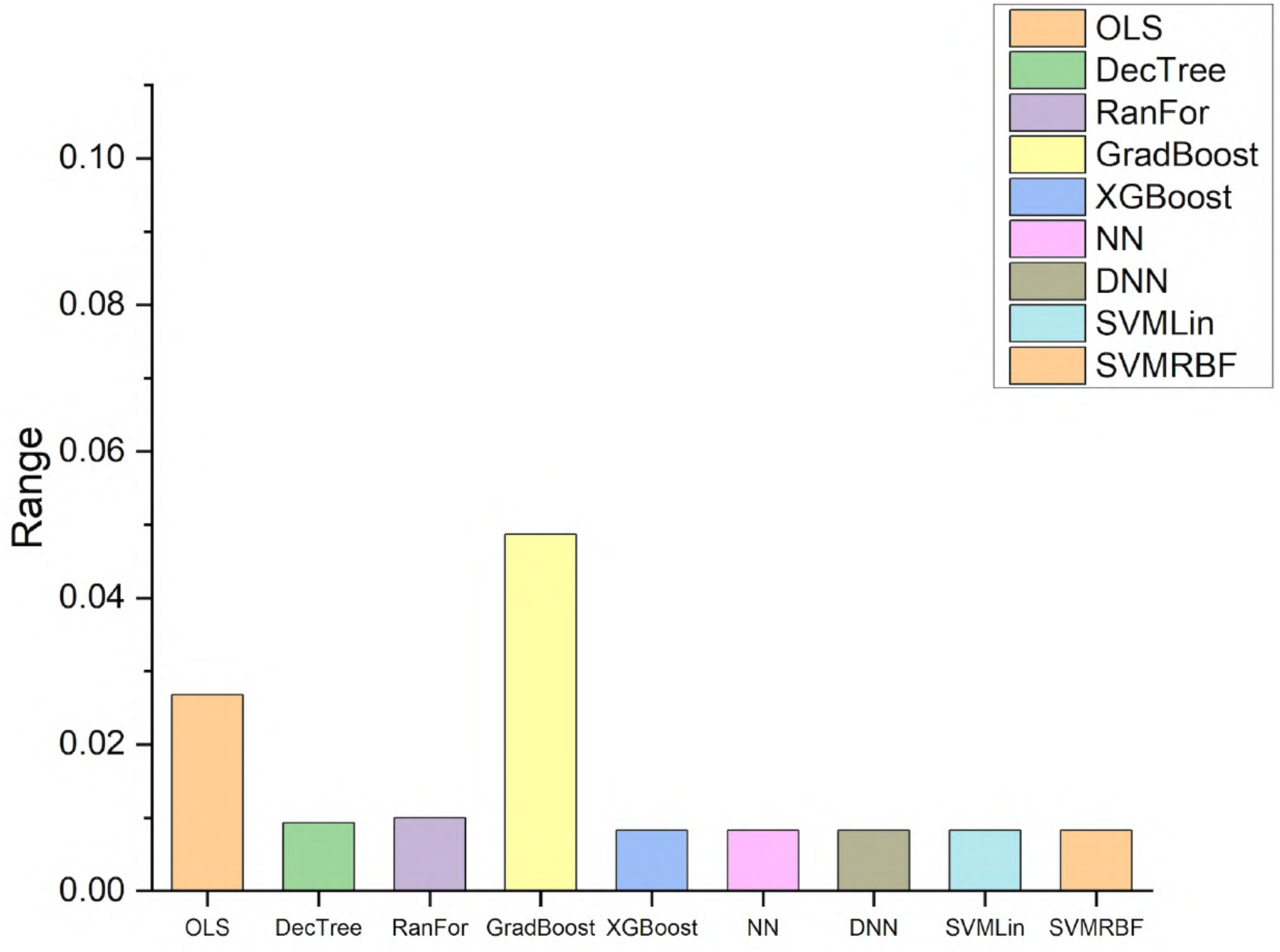


Fig8. RMSE at 80:20

Choosing the accurate algorithm by individual cycle study

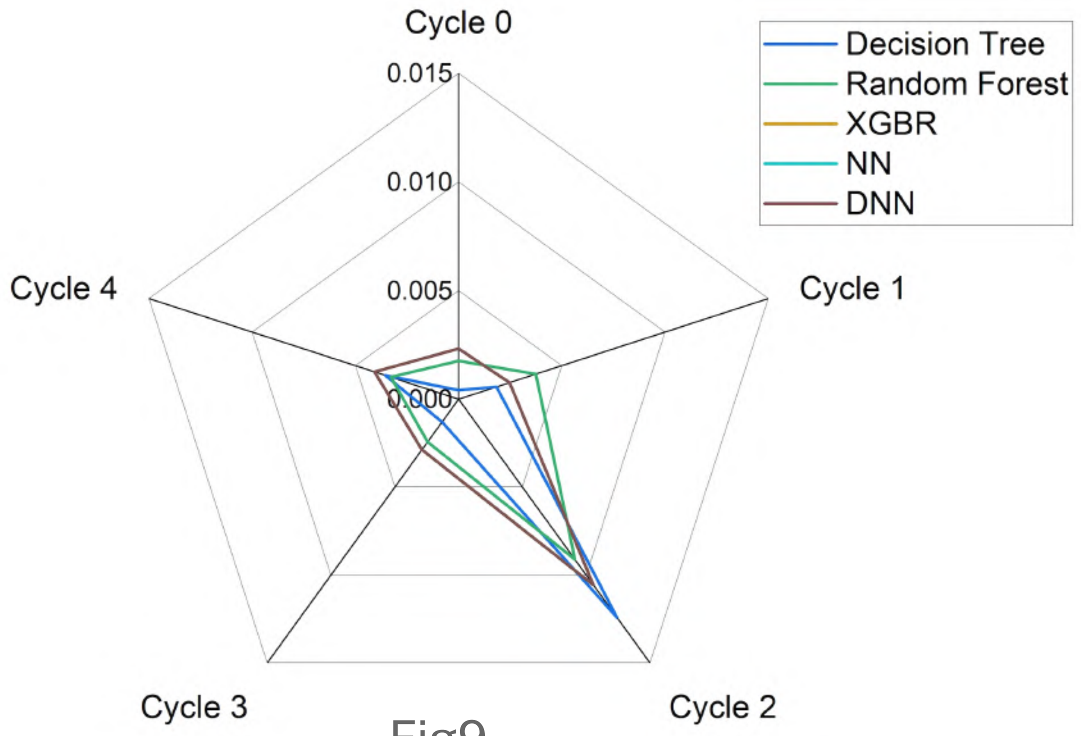


Fig9.
Condition 1

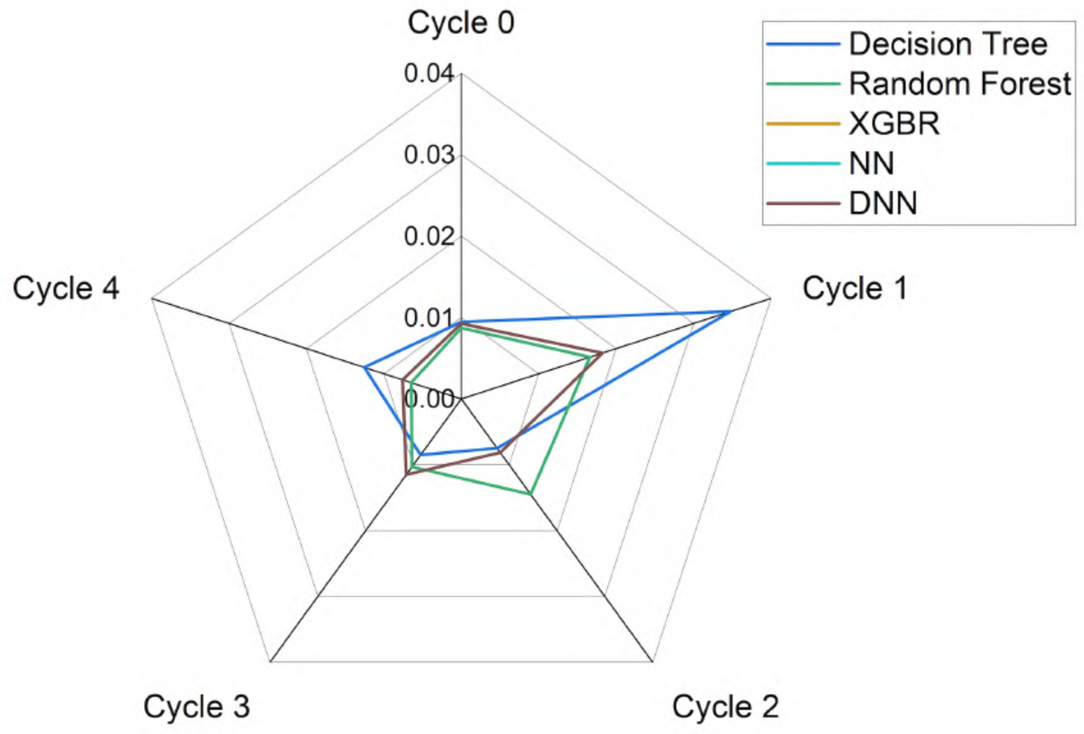


Fig11.
Condition 3

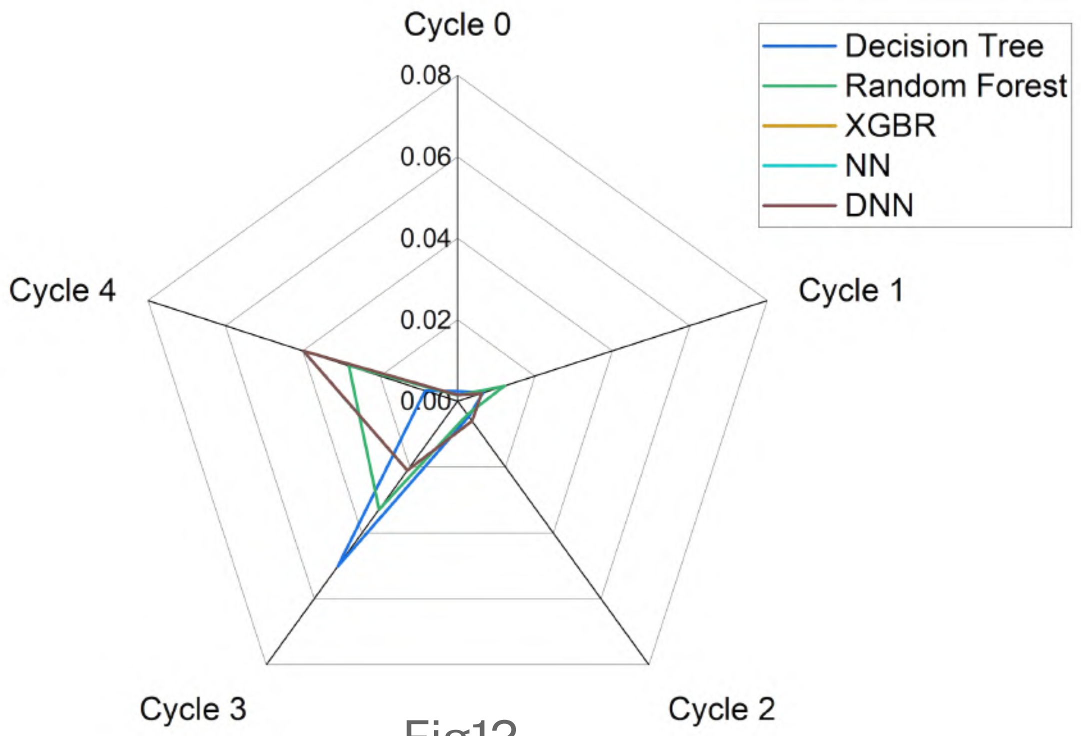


Fig12.
Condition 4

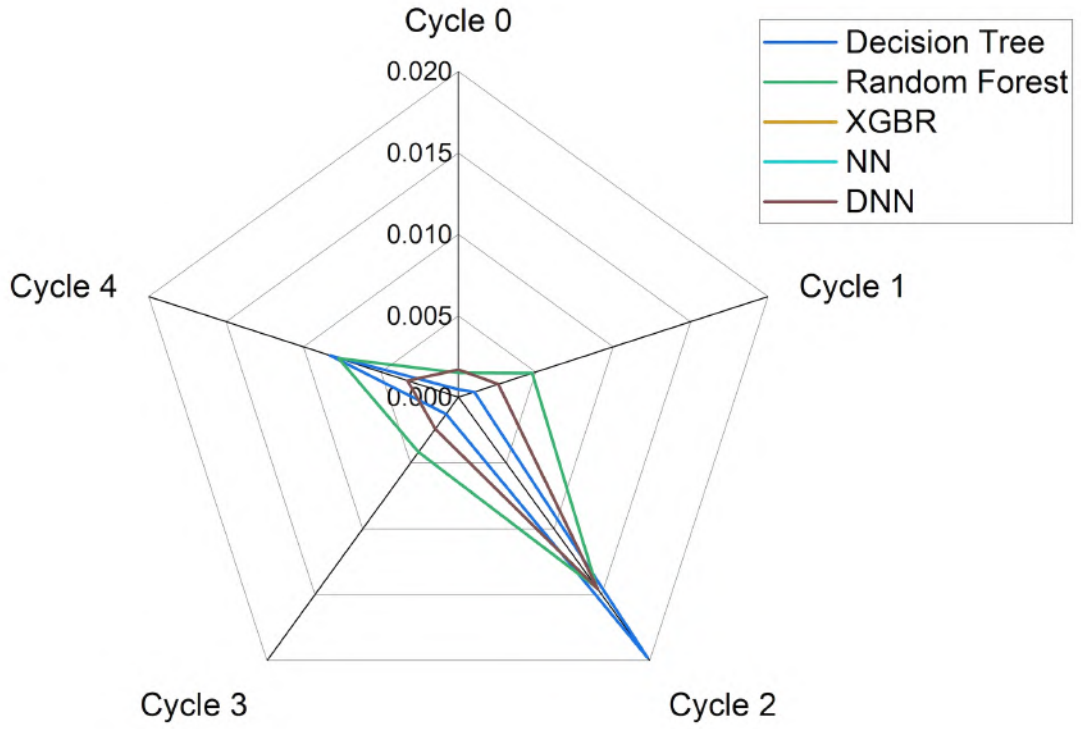


Fig10.
Condition 2

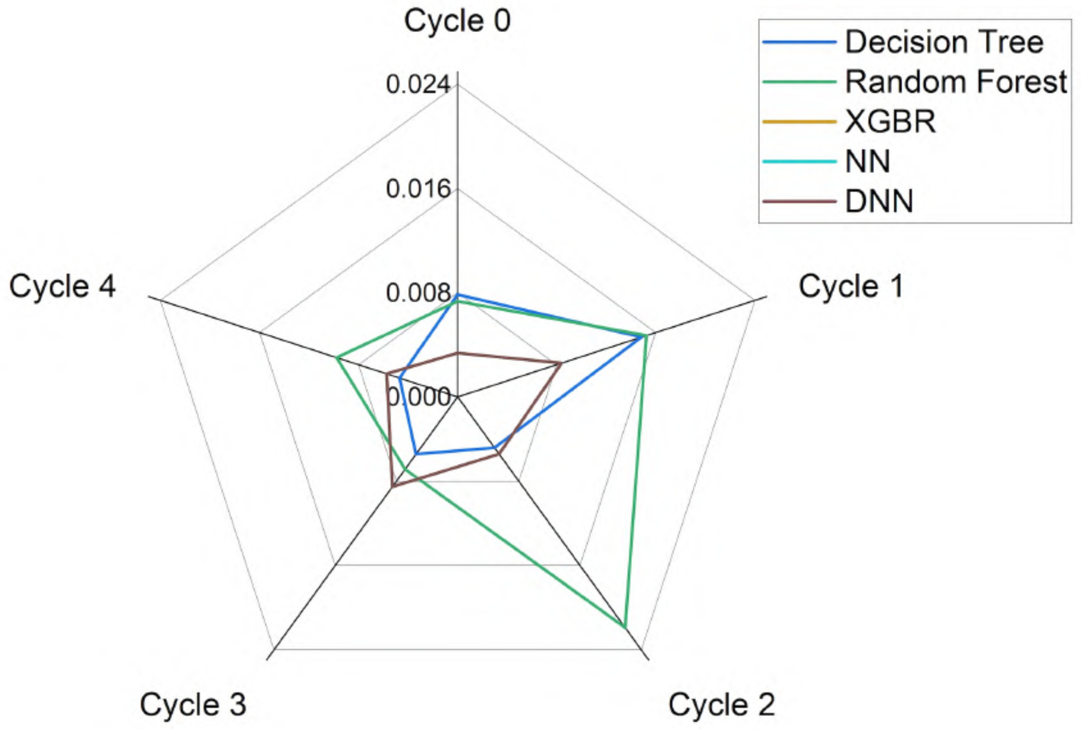


Fig13.
Condition 5

Results

Did we achieve our goals?

✓ Based on our analysis, the best training ratio was found to be **80:20**. This finding can enhance the robustness of our approach.

✓ Feature selection proved to improve the performance of all the models by around **10%**

✓ **Random Forest Regression** accurately predicted the output in all the conditions and in individual cycles.

Future Work - BMS Module

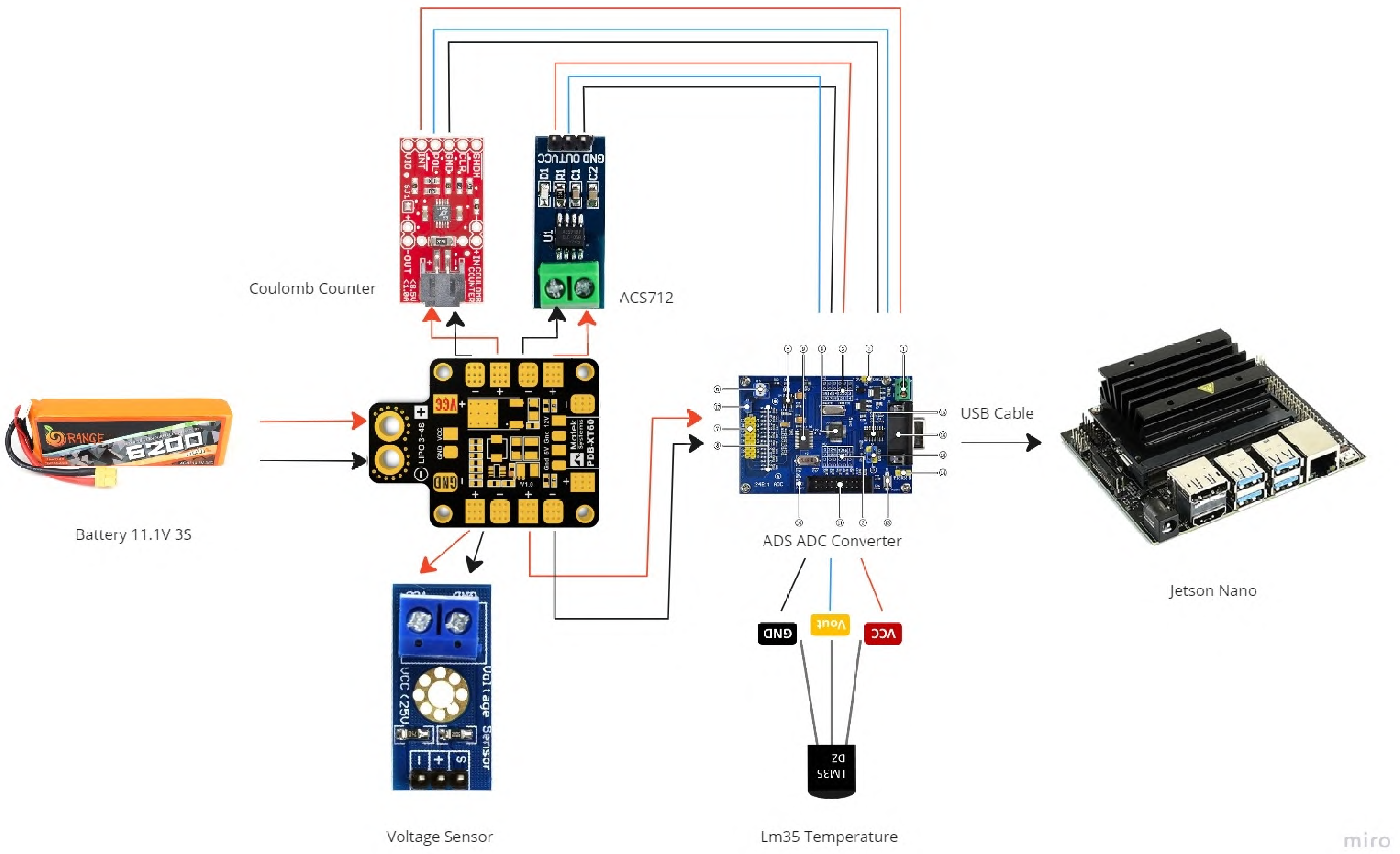


Fig. 14 - BMS circuit Diagram

References

1. Vilsen, S. B., & Stroe, D.-I. (2021). Battery state-of-health modelling by multiple linear regression. *Journal of Cleaner Production*, *290*, 125700. https://doi.org/10.1016/j.jclepro.2020.125700
2. Yang, H., Cao, Y., Xie, H., Shao, S., Zhao, J., Gao, T., Zhang, J., & Zhang, B. (2020). Lithium-ion Battery Life Cycle Prediction with Deep Learning Regression Model. *2020 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 3346–3351. https://doi.org/10.1109/APEC39645.2020.9124049
3. Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785
4. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451
5. Pahlavan-Rad, M. R., Dahmardeh, K., Hadizadeh, M., Keykha, G., Mohammadnia, N., Gangali, M., Keikha, M., Davatgar, N., & Brungard, C. (2020). Prediction of soil water infiltration using multiple linear regression and random forest in a dry flood plain, eastern Iran. *CATENA*, *194*, 104715. https://doi.org/10.1016/j.catena.2020.104715
6. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88
7. Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, *4*(3), 183–201. https://doi.org/10.1016/j.jfds.2018.04.003
8. Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., Yu, W., & Yan, J. (2020). Study of cardiovascular disease prediction model based on random forest in eastern China. *Scientific Reports*, *10*(1), 5245. https://doi.org/10.1038/s41598-020-62133-5

Progress

Mid Sem 7th Semester

- Understand the working of a monitoring system and identify the problem statement
- Choose necessary sensors and hardware required for data acquisition

End Sem 7th Semester

- Understand machine learning algorithm and its working
- Understood regression and compared with other types of regression models

Mid Sem 8th Semester

- Increasing the amount of the dataset for better model performance.
- Using Machine Learning to implement complex algorithms.

End Sem 8th Semester

- Integrate hardware and Machine Learning to get the desired result of a battery driven device

Thank you