**Paper 1:**

Yildiz et al. performed Regression and ML analyses on commercial building electricity load forecasting. The data is obtained from Kensington Campus and Tyree Energy Technologies Building (TETB) at University of New South Wales (UNSW). Importance of forecasting is mentioned in the paper together with brief review of Thermal models, and Auto Regressive models. Hourly electricity load data and minute interval weather data including Ambient Dry Bulb Temperature (DBT) and Relative Humidity (RH) is obtained from Campus and TETB electricity meters and local weather station respectively. Complimentary weather data is obtained from Sydney Observatory Hill Weather Station which is 7 km away from UNSW.

Comprehensive Regression analysis is provided. Input parameter are referred to as influence parameter as they influence the target parameter (load). Regression models and how to improve them is provided. Single vs multivariate regression models is considered. ML models considered are ANN, SVM, and Regression Trees. The authors analyze different ML algorithms for both campus and single building separately. Also they perform two type of Short term elecctricy load forecasts (STLF). First is day ahead Hourly forecast and second is Daily peak forecast. This is done for both buildings. Furthermore, data is analyzed for four seasons and for year 2013-2014. Summer months are December, January and February. Autumn months are March, April and May. Winter months are June, July and August. Spring months are September, October and November.

Influence parameters considered for hourly forecast are Previous day same hour load, Previous week same hour load, Previous 24 h average load, Working day/holiday binary indicator, DBT, RH, and Day of week. Similarly for peak load forecast; influence parameters are Previous day peak load, Previous week peak load, Previous week minimum load, Holiday/Business day binary indicator, DBT and hour of the day. For performance evaluation of models, the metrics used are $R^2$, $R^2_{adj}$, RMSE(%), MBE(%), and MAPE.



Results shows that ANN with Bayesian Regiulation Backpropagation is the best ML model. Furthermore, regression models performed fairly well compared to advanced ML models. Moreover, almost all model perform better prediction for overall campus load compared to single building load. Also, average day ahead hourly forecasts have higher accuracy compared to daily peak demand forecasts.

I can use the methods for improving regression models for my project that are discussed in paper such as Principal Component Analysis (PCA), Sensitivity analysis, and Stepwise regression.

**Paper 2:**

Young et al. proposed ANN model for forecasting sub-hourly (15 minutes interval) electricity usage in commercial buildings. They investigate even smaller unit than an hour for STLF. Data set for study is obtained from building management system (BMS) of a commercial building complex, and data are periodically pulled into a relational database. The site consist of three office buildings and they all are managed by onbe utility billing system. One main electric meter and several sub-meters are installed. The main meter measures electricity usage, both the instantaneous power in kW with minute interval and aggregated electricity usage at every 15 minutes in kWh.

Nine ML models considered are Simple Naive model, Gaussian process with radial basis function (RBF) kernel, Gaussian process with polynomial kernel, Linear Regression, ANN, SVM with normalized polynomial kernel, SVM with RBF kernel, K-Star classiifer, and Nearest neighbor ball tree. Significant predictor variables are previous electricity usage, Interval stamp (TIF), Day indicator (DTF), HVAC operation schedule (OPC), Outdoor dry-bulb temperature (ODT), and Outdoor relative humidity (ODH). For performance evaluation of models, the metrics used are correlation coefficient, Coefficient of variance of the root-mean squared error or CV(RMSE), and Absolute Percentage Error (APE).

Results shows that ANN is the best algorithm for this study. For verification, two months (Augusta nd September) in year 2012 are used. Furthermore, three training methods are considered: Static, Accumulative and Sliding Windows. Cumulative and Sliding win dows train better than Static method. Therefore ANN with regularization algorithm for training is adopted. The model can provide a day-ahead electricity usage profile with sub-hourly intervals and daily peak electricity consumption with a reasonable accuracy.

**Paper 3:**

Mucahit et al. proposed a time series forecasting- based peak shaving algorithm for building energy management. Peak shaving helps to reduce the peak electricity demand resulting in reduced cost for end-users. A smart building is any structure that uses a central controller to automatically regulate energy demand. The peak load of the system is the highest amount of energy consumption during the day that is characterized by short time periods. To handle peak demand, peak load shaving is an attractive strategy. It involves using battery energy storage systems (BESS) where the secondary energy storage device allow a microgrid utility to shave the peak demand by chagrin g the BESS when demand is low, and discharging it when demand is high.

The publicly available dataset provided by the U.S Department of Energy on household electricity consumption is used for analysis. The data is taken from March 6, 2011, to April 5, 2012. The peak of electricity on weekdays is higher than on the weekend. Furthermore, peak of electricity occurred between 11 pm and 1 am on weekdays and between 12 pm to 3 pm on weekends. Moreover, for classifying predicted load, electricity loads are classified into seven

groups. First class ($C^1$) is associated with range from 0 to 1000 kW electricity, and last class ($C^7$) is associated with range from 5500 kW or higher electric consumption. $C^1$is best region for recharge action while $C^7$is best region for discharge action.

The ML models used in regression analysis for one-step ahead forecasting are as follows: Naive Baseline (taking load/label previous week load as predicted value for today), Random Forest (RF), Light Gradient Boosted Machine (LGBM), and Long-Short Term Memory Netwrorks (LSTM). Performance metrics for regression analysis are Normalized deviation (ND), Normalized Root

Mean Squared Error (NRMSE), and MAPE. Furthermore, ML models used in classification analysis for load forecasting are as follows: K-Nearest-Neigbours with Dynamic-Time-Wrapping (KNN-DTW), RF, XGBoost, LGBM, LSTM, Fully Convolutional Networks (FCN), and Residual Networks (ResNet). Performance metrics for classification analysis are Accuracy, F1-score, Precision, and Recall.

The results shows thats RF is the best algorithm for both regression and classification tasks. The RF is then used to develop a forecasting-based peak shaving algorithm, which first predict the peak period, and these predictions are then used to determine the decision of charging and discharging the battery. I can apply the knowledge of ML models used in study since I have both regression nd classification analysis in my project.

**Paper 4:**

Kody et al. performed Heating, Cooling, and Electrical Load forecasting for a large-scaledistrict energy system. The study covers a large-scale district energy system that simulateouly produce electricity, heating and cooling for a large University of Texas campus at Austin. The Hal C.Weaver power plant and associated facilities provides all the cooling, heating and electrical needs for the campus. It is connected to city grid but only used in case of emergency. The load 12 profiles for cooling and electric loads fluctuate significantly while that of heating load remains almost constant for summer-time condition for example in August. Furthermore, Cooling load fluctuate more compared to electrical load. In winter-time condition, e.g in February, heating load profile increased while cooling load profile decreased compared to summer-time condition. Moreover, in winter, heating load fluctuates much more making it difficult to forecast them accurately.

The input parameters are dry bulb temperature and RH (weather variables). The correlation analysis in the study provide useful insights regarding relations between input varaibles and three different loads and also among three different loads as well. I can also apply the visualization similar to figure 4 in study as it give clear pictures which variables are related to other strongly. Each load 9Electrical, Cooling and Heating) is strongly correlated to ambient dry bulb temperature and less so with ambient relative humidity. As expected, cooling and electrical loads are positively correlated to temperature and negatively correlated to humidity, while the reverse is true for heating. Furthermore, all the loads are highly correlated with each other suggesting they all undergo similar variations.

ANN is still the dominant methodology for forecasting building energy loads. Moreover, time series analysis models are an improvement upon ANNs in case of time dependency of data points e.g, energy load forecasting. Time series analysis models are categorized into two groups, time-domain and frequency-domain. Therefore, three models namely ANN, Linear Auto Regressive with eXogenous input (ARX), and Nonlinear ARX. Furthermore, all three models are developed first using weather input only and then both weather and time inputs. Coffectient of determine ($R^2$) is a performance metric. The results shows that for one day ahead prediction for cooling, heating and electrical loads, NARX with both weather and time variables outperform all other models.

**Paper 5:**

Zakria et al. performed energy output forecasting analysis of Hybrid photovoltaic (PV)-wind system using feature selection technique for smart grid. The motivation is to enhance smart grid by efficiently predicting energy produced by renewables energy system. Weather factors have significant impact on power output of hybrid system. The weather factor (input parameters) selected for study are Solar irradiation (Solar power per unit area), Wind Speed, Ambient Temperature, Humidity, Precipitation, Atmospheric pressure, and wind direction. Solar irradiation, Wind Speed, Ambient Temperature, and Humidity have most significant impact on PV-wind power output. Target feature can be either power or energy. Goal is to determine energy demand trend over long period.

Historical hourly weather dataset is gathered using calibrated sensors at Middle East Technical University, North Cyprus Campus from Jan 1$^{st}$to Dec 26$^{th}$, 2015. Seven regression models are used in study namely: Extra tress Regressor, AdaBoost, SVR, K-Neigbors Regressor, Gaussian Process Regressor, MLP Regressor and Linear regressor. Performance metrics are MSE, MAE, $R^2$, and computational time. All seven models are first compared without any feature selection technique and then using Recursive Feature Elimination using Cross Validation (RFECV) technique. Feature Selection method such as RFECV can improve overall computational efficieny.

Results shows that Linear Regression is the best model for both analysis with and without feature selection technique. Also analysis showed that attributes are linearly dependent on each other. Development model stages (Figure 4) and Pair Plot (Figure 7) are together useful information that I can apply in my project. Other information I learn is that for every 1 degree in temperature, there is 5% decrease in RH. Furthermore, For RH > 50%, retaltionship between temperature and RH become linear.