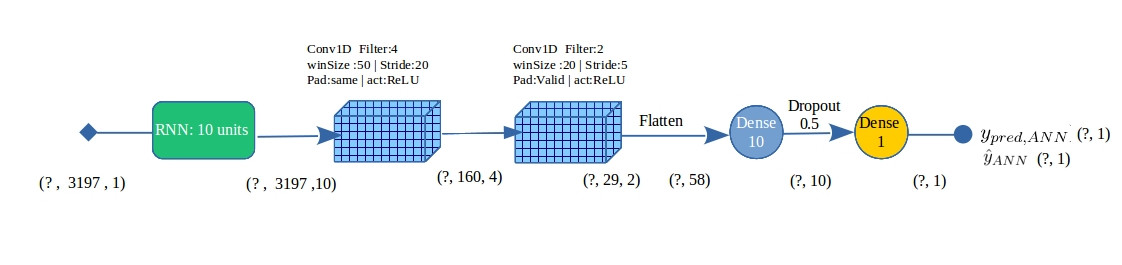
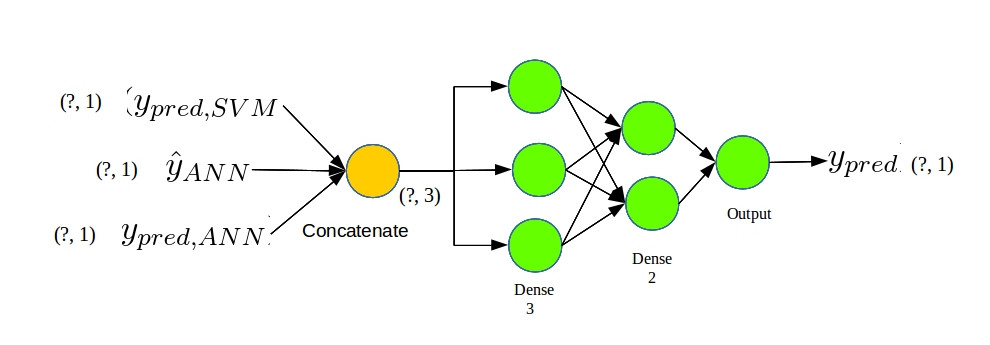
# ARCHITECTURE



# METHODOLOGY

The CSV is parsed into a usable format and detrending by using median-filter smoothed curve is done and top 2 % of the maximum values are removed. The datapoints are then normalized and scaled separately. The samples are converted into frequency domain by Fourier Transformation. One-third of the given data is separated as validation data for Phase-1.   
Phase1: SVM & ANN  
 The fourier transformed data , after SMOTE sampling and resampling used as input for SVM with linear kernel to obtain a model and this model is saved.  
The Detrended clipped and scaled data is resampled and used as input feed to the ANN model consisting of a RNN and two 1-Dimentional convolution layer , and then into a fully connected layer and a dropout layer before the class is predicted.   
Best model among many iteration is manually chosen and the model saved is used for Phase-2.

Phase 2:Ensembling.  
The saved ANN model is used to obtain the predictions and proability of existence of exoplanet to serve as feature input to the ensembler. Similarly, the saved SVM model is used to obtain the predictions.   
The obtained features are concatenated and are again split as different training and validation sets to train the Ensembler, and the best model among all iterations is saved for the testing.

During testing the same preprocessing except the train-validation split was done.  
The data is made to run via the saved SVM and the saved ANN models to obtain features and is then passed to the Ensembler to get the Final Predictions.

# NOVELTY

The presence of periodic dips in the observed flux/light intensity is considered as conclusive evidence for existence of exoplanets. Hence the detection of this feature is crucial. The given timeseries is observed to have nonzero trend and high noise. Inorder to tackle these issues, Detrending and Scaling was done . The range of the fluctuations is different for each sample but is made same by scaling and normalization. The presence of periodicity can be detected by Fourier Transforms (peaks in frequency domain). SVMs are one of the oldest and smart classification method that applies the kernel trick to project data intp higher (N) dimensional planes and classifying them with a (N-1) dimensional hyperplane.

RNNs are extensively used in NLP tasks for understanding the semantic meaning of tensor sequences , in this case, timeseries data. The obtained tensor is passed into two 1-D convolution which convolves along timestamps. CNNs are well known and used for pattern recognition task and may be effective to find the dips in the encoded timeseries. The dropout layer airs regularization and helps prevent overfitting.

MLP based ensemble layer is trained using the output of two other models. This gives the provision to filter and use the best of both models to predict the final class.

# REFERENCES

Fourier Transforms (Bloomfield, 2004)  
SMOTE sampling(Chawla et al., 2002)  
Recurrent Neural Networks (Dieng et al., 2016)   
CNN (Karpathy,2015 cs231n)  
Dropout regularization (Srivastava et al., 2014)  
SVM: https://www.svm-tutorial.com  
Ensemble (Ghosal et al., 2017)  
Outlier clipping https://www.kaggle.com/aleksod/0-75-precision-0-60-recall-linear-svc  
KOIs: https://en.wikipedia.org/wiki/Kepler\_object\_of\_interest  
Transit: https://en.wikipedia.org/wiki/Transit\_(astronomy[)](https://en.wikipedia.org/wiki/Transit_(astronomy))  
Transit photometry: https://en.wikipedia.org/wiki/Methods\_of\_detecting\_exoplanets