

CLASSIFYING VARIABLE STARS

TEAM "THE DOOMED KID"

Anshuman Singh

MD tanzil



PROBLEM STATEMENT

Classifying periodic variable stars amongst 4 categories:

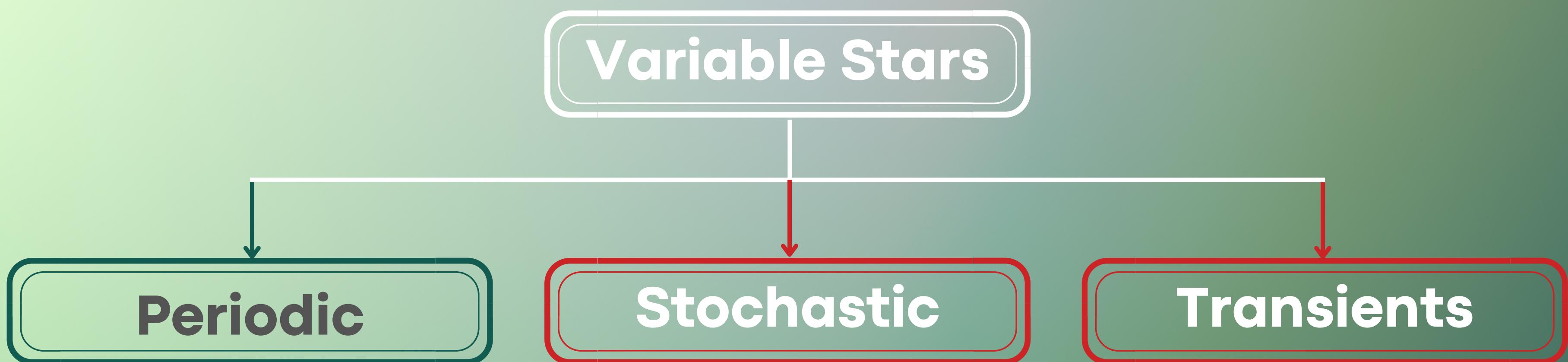
1. RR Lyrae
2. Cepheid
3. Long Period Variable Stars
4. Eclipsing Binary

DATASET

GAIA DR3

EXTRACTION

Astroquery
~50 GB



BEFORE

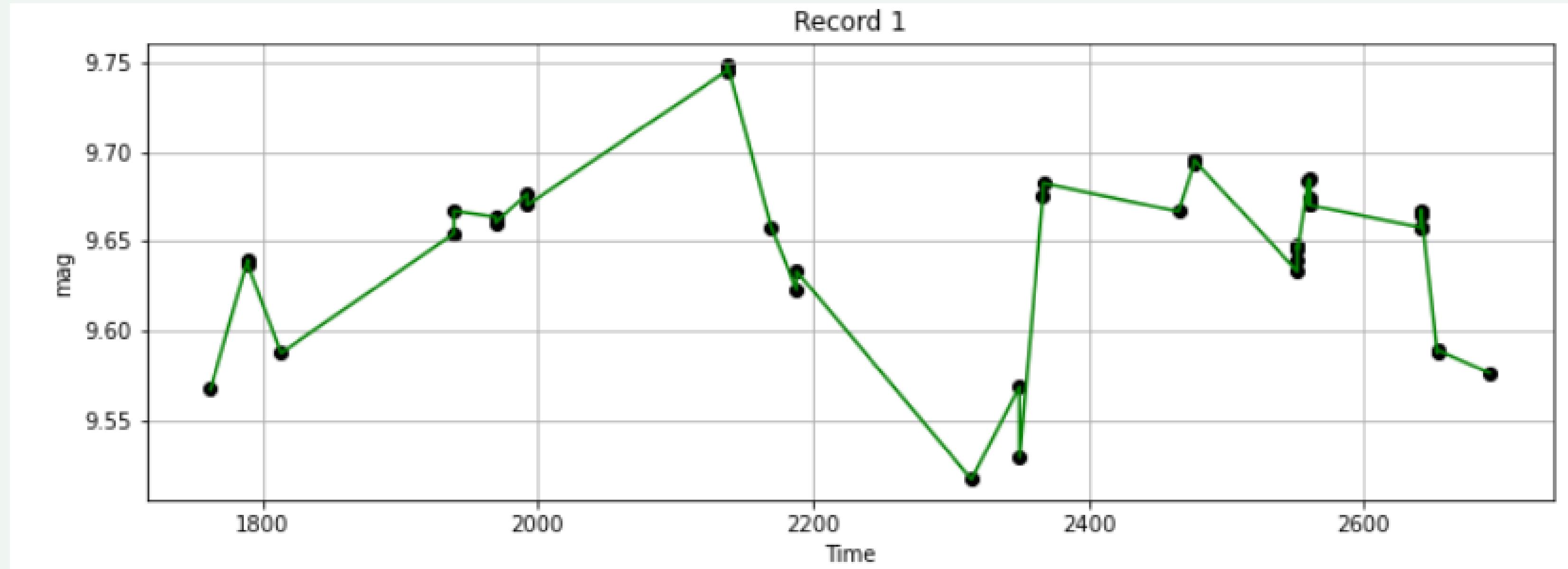
Class	Number of Records
Cepheid	16141
RR Lyrae	297778
Long Period	2325775
Eclipsing Binaries	2184356

FILTERING

AFTER

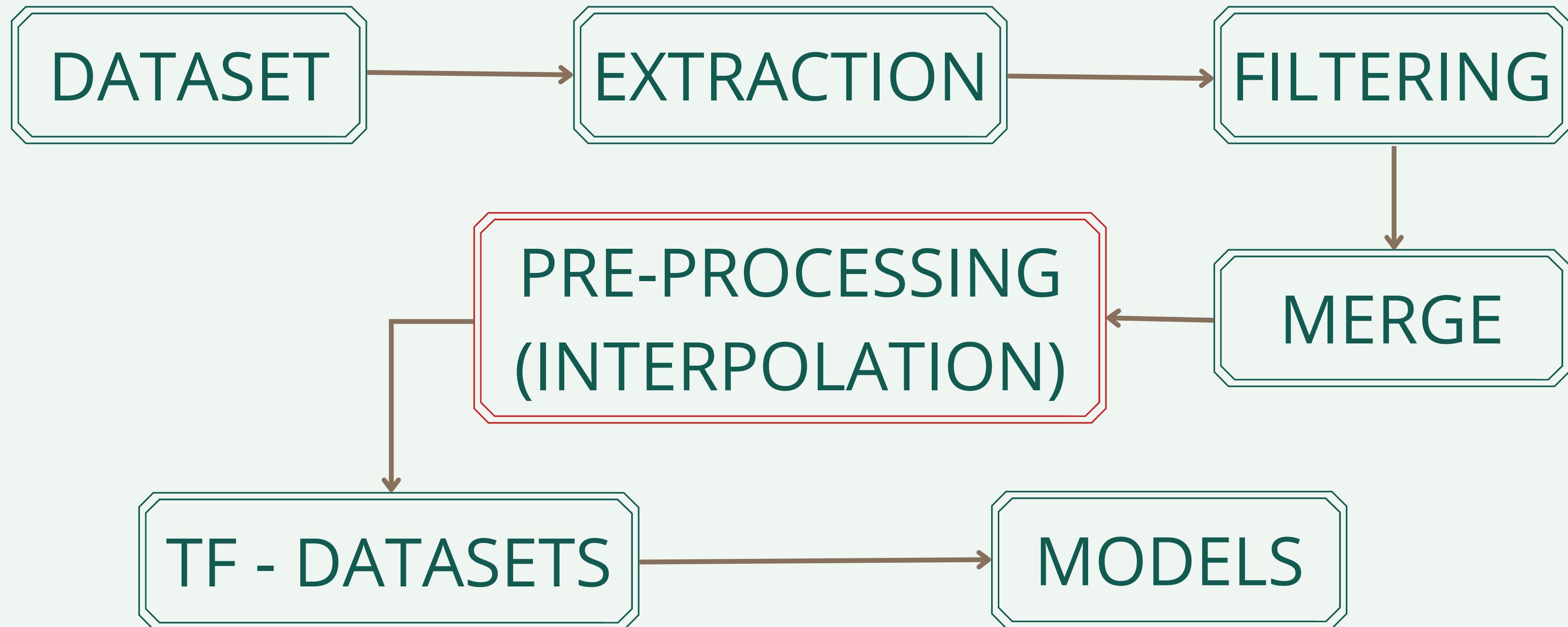
Class	Number of Records
Cepheid	15021.0
RR Lyrae	100000.0
Long Period	100000.0
Eclipsing Binaries	100000.0

EDA



LIGHT CURVE OF AN LPV STAR OF BAND "G"

APPROACH



DIFFERENT MODELS

- BASE MODEL - CNN
- RECURRENT NEURAL NETWORKS
- GATED RECURRENT UNITS
- LONG-SHORT TERM MEMORY MODEL
- MULTI LAYER LSTM
- MULTI LAYER LSTM WITH INTERPOLATED LIGHT CURVE DATA
- MULTI LAYER LSTM WITH INTERPOLATED LIGHT CURVE DATA
AND TIME DIFFERENCE AS INPUT

RESULTS

BASE MODEL - CNN

65.08

LSTM

97.68

MULTI LAYER LSTM

98.16

MULTI LAYER LSTM
+ INTERPOLATION

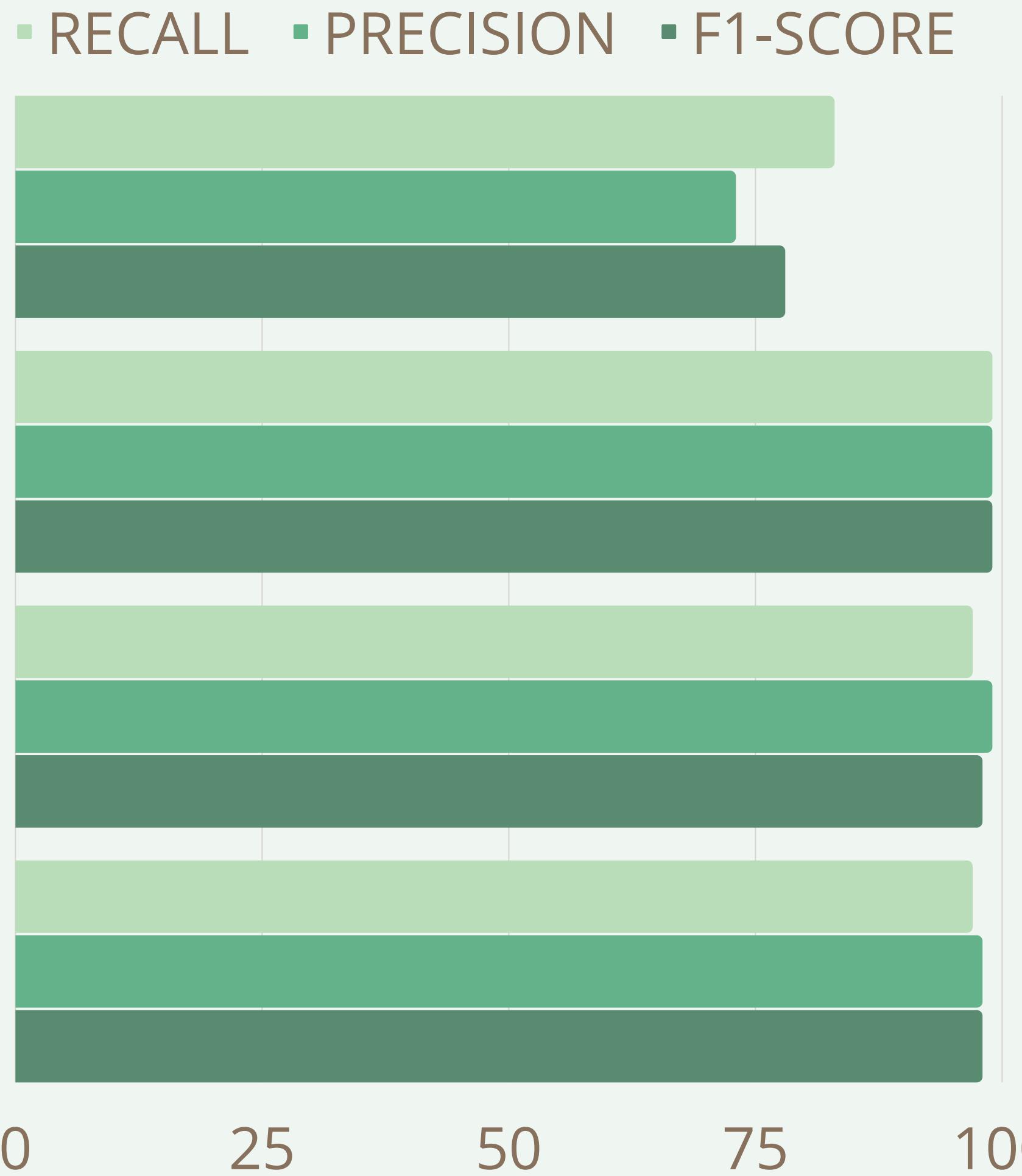
97.30

MULTI LAYER LSTM
+ INTERPOLATION
+ INPUT TIME DIFFERENCE

97.31

ACCURACY

BEST MODEL



CONCLUSION

- The best model for G band input is Multi Layer LSTM.
- Considering the data from other bands, it does not generalize well.
- The Multilayer LSTM model with interpolation and time difference as input generalizes better.

FUTURE WORK

- Include more classes and sub-classes
- Try to input the data of the other 2 bands – RP, BP.
- Combine the data of different bands to obtain an improved result.
- Experiment with LSST data.