# december 7th, 2024

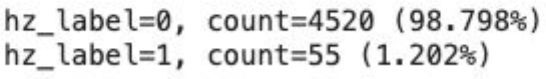
exoplanet classification

# neural networks classifier

- goal: classify exoplanet habitability - binary classifier

# training data processing

- join NASA 09-15-2024 data with HWC data from PHL.
- HWC data has a “*P_HABITABLE*” data field that can be used as label
- training data preprocessing:
  - remove data fields that are not relevant to training
  - drop data fields with too much missing values
  - for categorical data fields:
    - filling missing values with mode
    - encode with LabelEncoder
  - for numeric data fields:
    - filling missing values with MICE imputation
  - use SMOTEENN to oversample and downsample to overcome sample imbalances



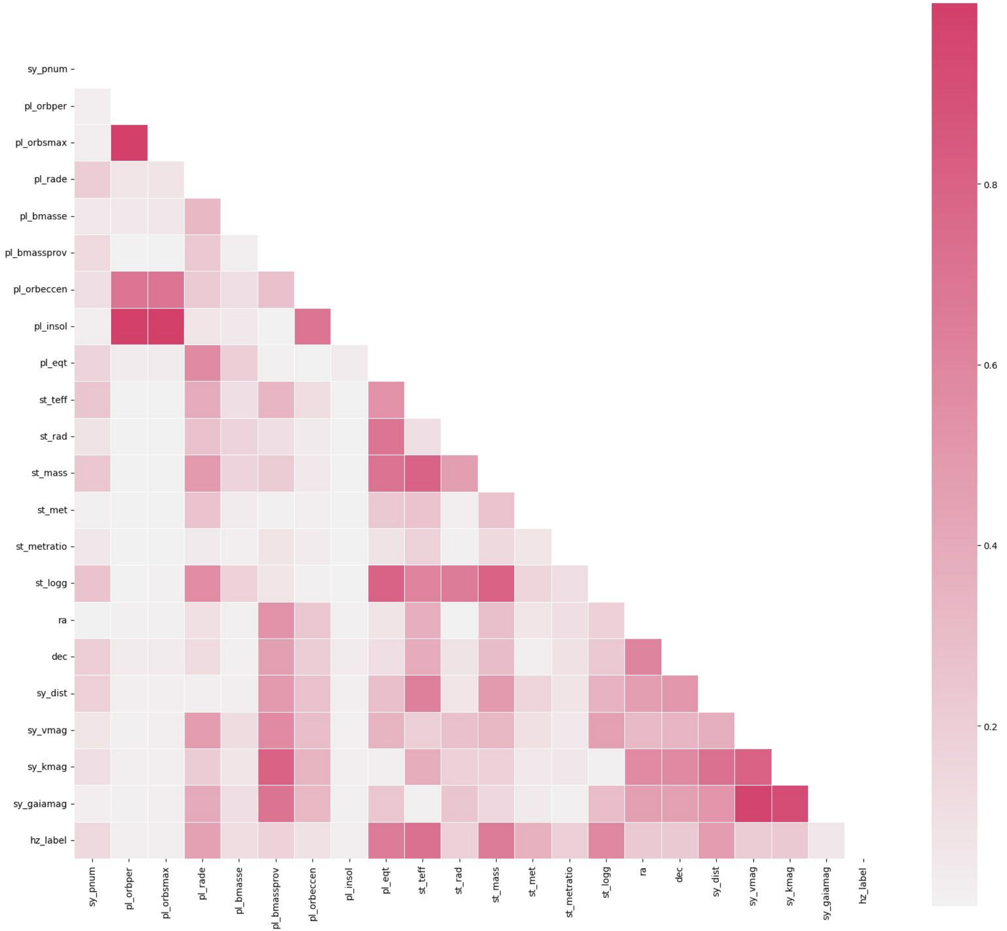
```
hz_label=0, count=4520 (98.798%)
hz_label=1, count=55 (1.202%)
```

# feature correlation analysis

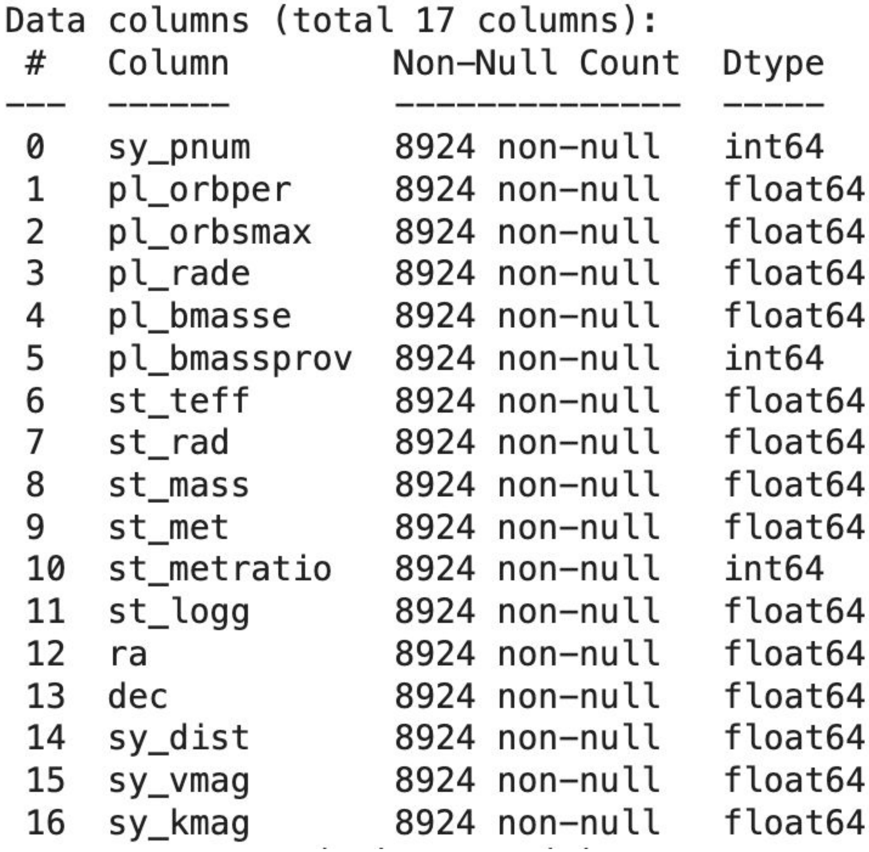
correlation analysis. remove highly correlated features:

- pl_orbeccen
- pl_insol
- sy_gaiamag

end up with 17 features in the training data

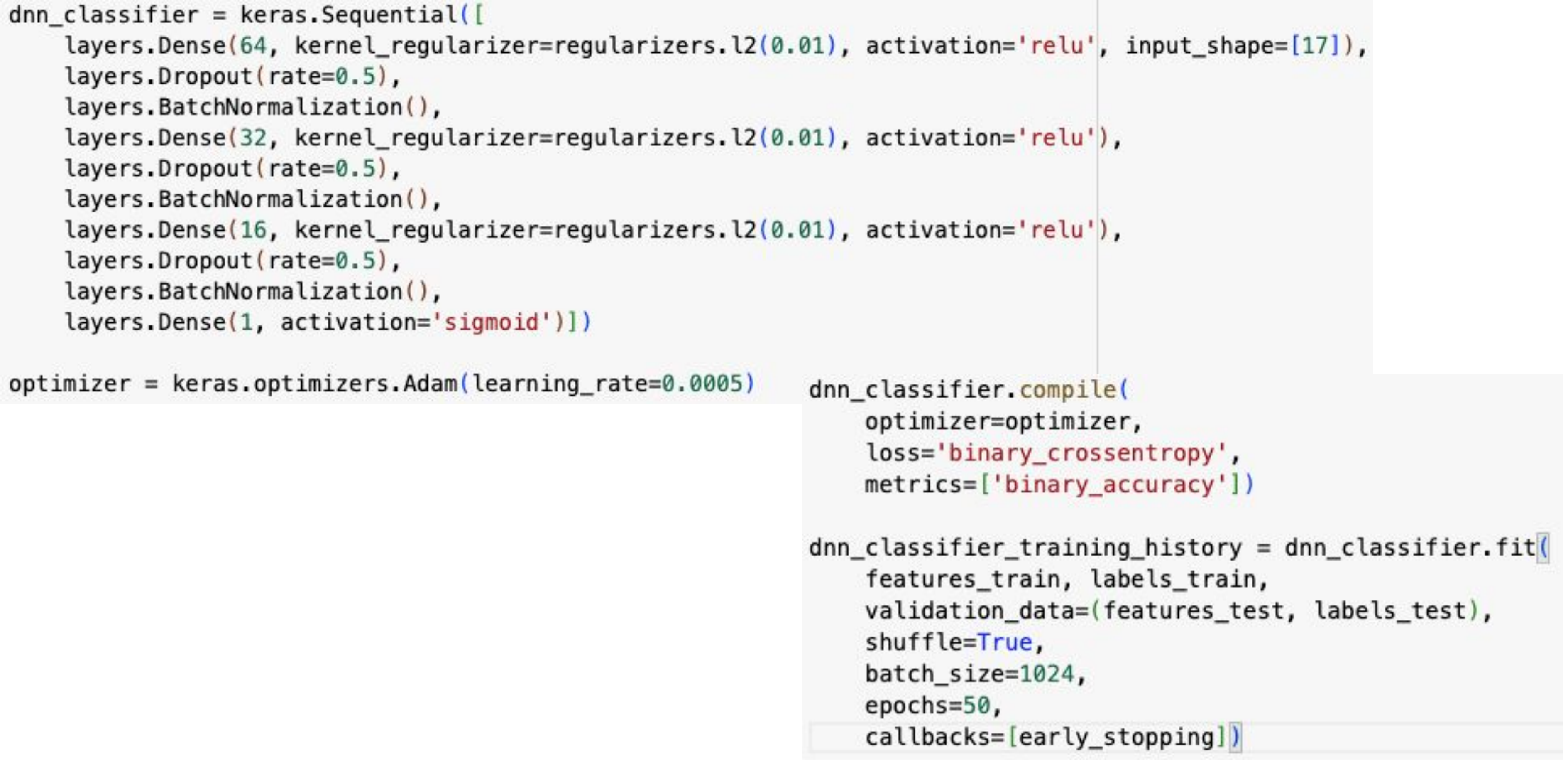


# training features



```
Data columns (total 17 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   sy_pnum      8924 non-null   int64  
 1   pl_orbper    8924 non-null   float64
 2   pl_orbsmax   8924 non-null   float64
 3   pl_rade      8924 non-null   float64
 4   pl_bmasse    8924 non-null   float64
 5   pl_bmassprov 8924 non-null   int64  
 6   st_teff      8924 non-null   float64
 7   st_rad       8924 non-null   float64
 8   st_mass      8924 non-null   float64
 9   st_met       8924 non-null   float64
 10  st_metratio  8924 non-null   int64  
 11  st_logg      8924 non-null   float64
 12  ra           8924 non-null   float64
 13  dec          8924 non-null   float64
 14  sy_dist      8924 non-null   float64
 15  sy_vmag      8924 non-null   float64
 16  sy_kmag      8924 non-null   float64
```

# neural networks classifier

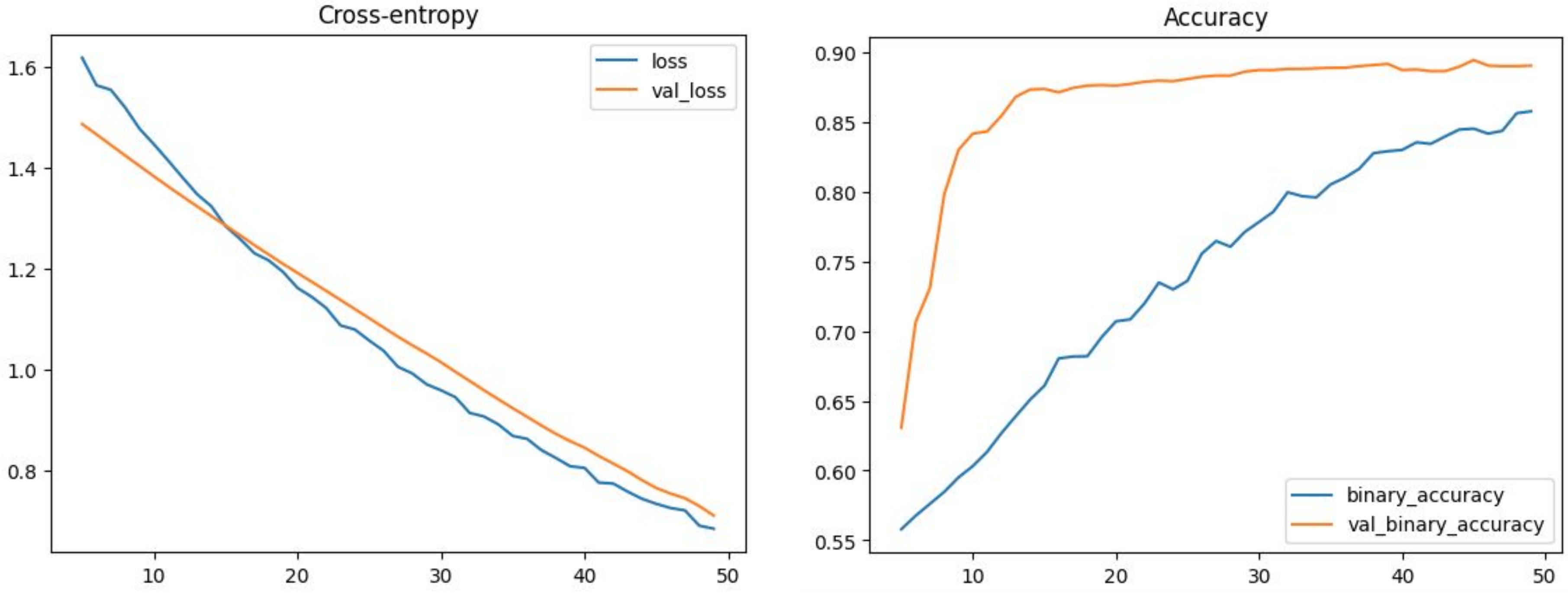


```
dnn_classifier = keras.Sequential([
    layers.Dense(64, kernel_regularizer=regularizers.l2(0.01), activation='relu', input_shape=[17]),
    layers.Dropout(rate=0.5),
    layers.BatchNormalization(),
    layers.Dense(32, kernel_regularizer=regularizers.l2(0.01), activation='relu'),
    layers.Dropout(rate=0.5),
    layers.BatchNormalization(),
    layers.Dense(16, kernel_regularizer=regularizers.l2(0.01), activation='relu'),
    layers.Dropout(rate=0.5),
    layers.BatchNormalization(),
    layers.Dense(1, activation='sigmoid')])

optimizer = keras.optimizers.Adam(learning_rate=0.0005)
```

```
dnn_classifier.compile(
    optimizer=optimizer,
    loss='binary_crossentropy',
    metrics=['binary_accuracy'])

dnn_classifier_training_history = dnn_classifier.fit(
    features_train, labels_train,
    validation_data=(features_test, labels_test),
    shuffle=True,
    batch_size=1024,
    epochs=50,
    callbacks=[early_stopping])
```

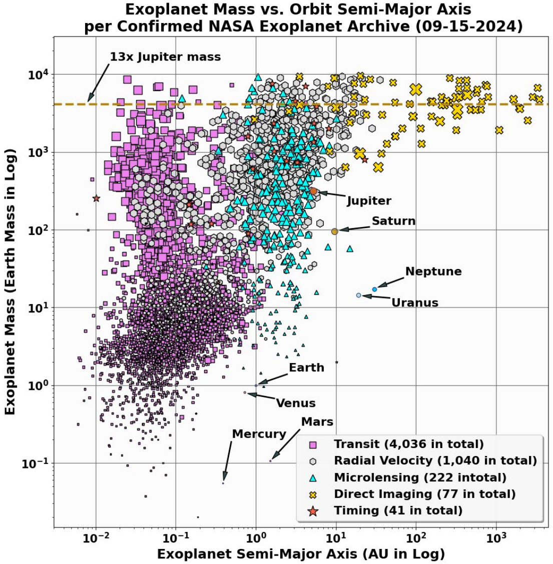
# neural networks classifier

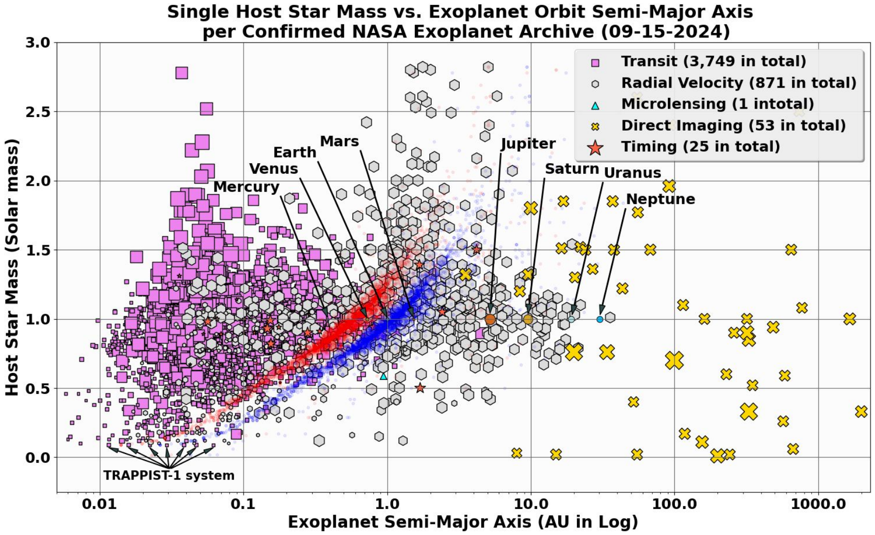
```
Best Validation Accuracy: 0.8944
```



# future work

- fine tune neural networks classifier
  - simpler model architecture: less layers, less connected units
  - hyperparameter tuning (learning rate, batch size, etc.)
- explore graphs related to Seager’s paper


Exoplanet Mass vs. Orbit Semi-Major Axis
per Confirmed NASA Exoplanet Archive (09-15-2024)
13x Jupiter mass
Exoplanet Mass (Earth Mass in Log)
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
$10^{-1}$
Jupiter
Saturn
Neptune
Uranus
Earth
Venus
Mars
Mercury
■ Transit (4,036 in total)
○ Radial Velocity (1,040 in total)
▲ Microlensing (222 intotal)
✖ Direct Imaging (77 in total)
★ Timing (41 in total)
$10^{-2}$
$10^{-1}$
$10^0$
$10^1$
$10^2$
$10^3$
Exoplanet Semi-Major Axis (AU in Log)


Single Host Star Mass vs. Exoplanet Orbit Semi-Major Axis
per Confirmed NASA Exoplanet Archive (09-15-2024)
Host Star Mass (Solar mass)
■ Transit (3,749 in total)
○ Radial Velocity (871 in total)
▲ Microlensing (1 intotal)
✖ Direct Imaging (53 in total)
★ Timing (25 in total)
Mars
Earth
Venus
Mercury
Jupiter
Saturn
Uranus
Neptune
TRAPPIST-1 system
Exoplanet Semi-Major Axis (AU in Log)