Blob-match: Machine learning for cross-identification of radio surveys

James Gardner, supervisors: Cheng Soon Ong, Matthew Alger

October 22, 2019

What's out there in the universe?

Figure: Giant Metrewave Radio Telecope and Very Large Array (TGSS) 150 MHz / resolution 25″ vs. (NVSS) 1.4 GHz / resolution 45″

Cross-identification

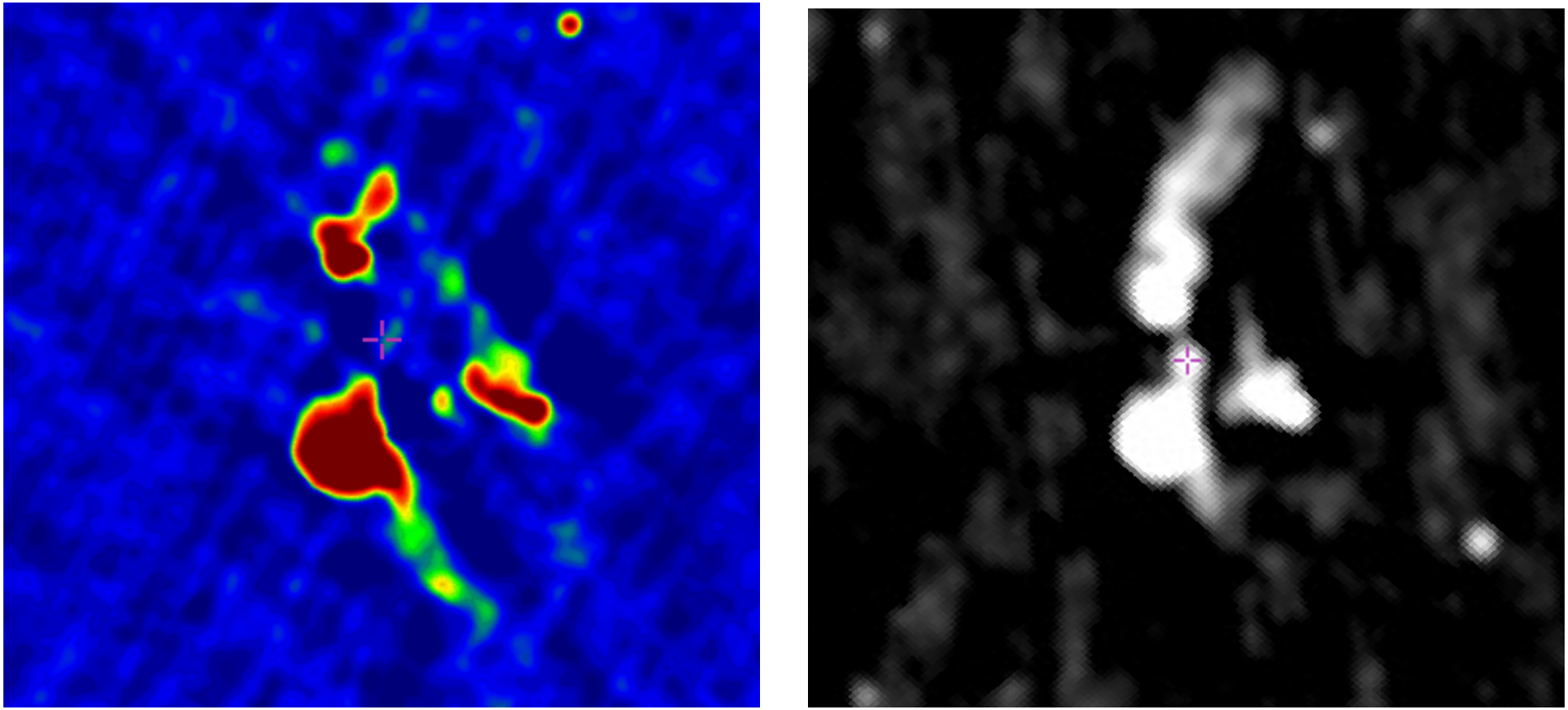


Figure: Sources (blobs) of 3c40 in TGSS and NVSS

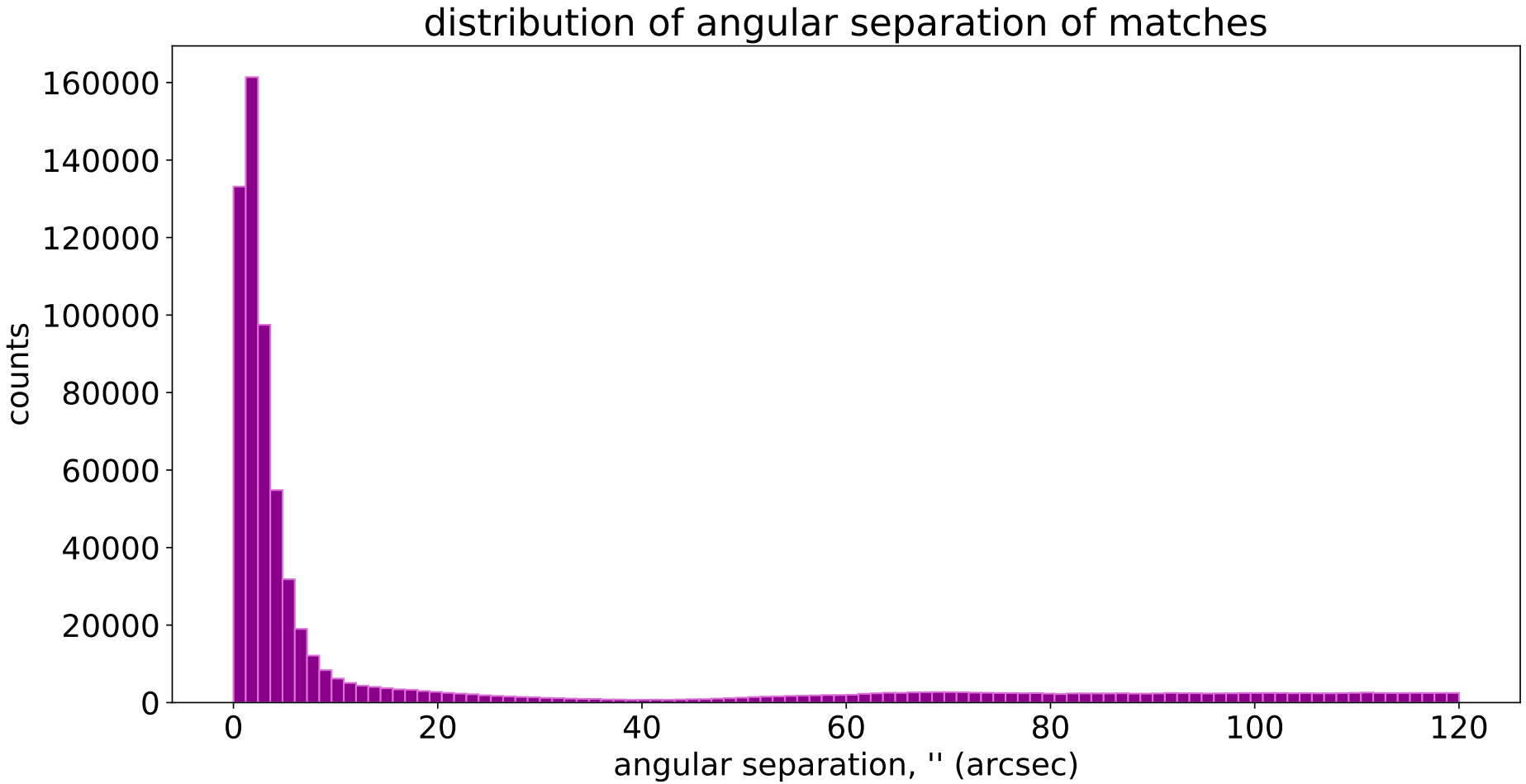
Conclusions to come

Reproduction of existing positional matching results

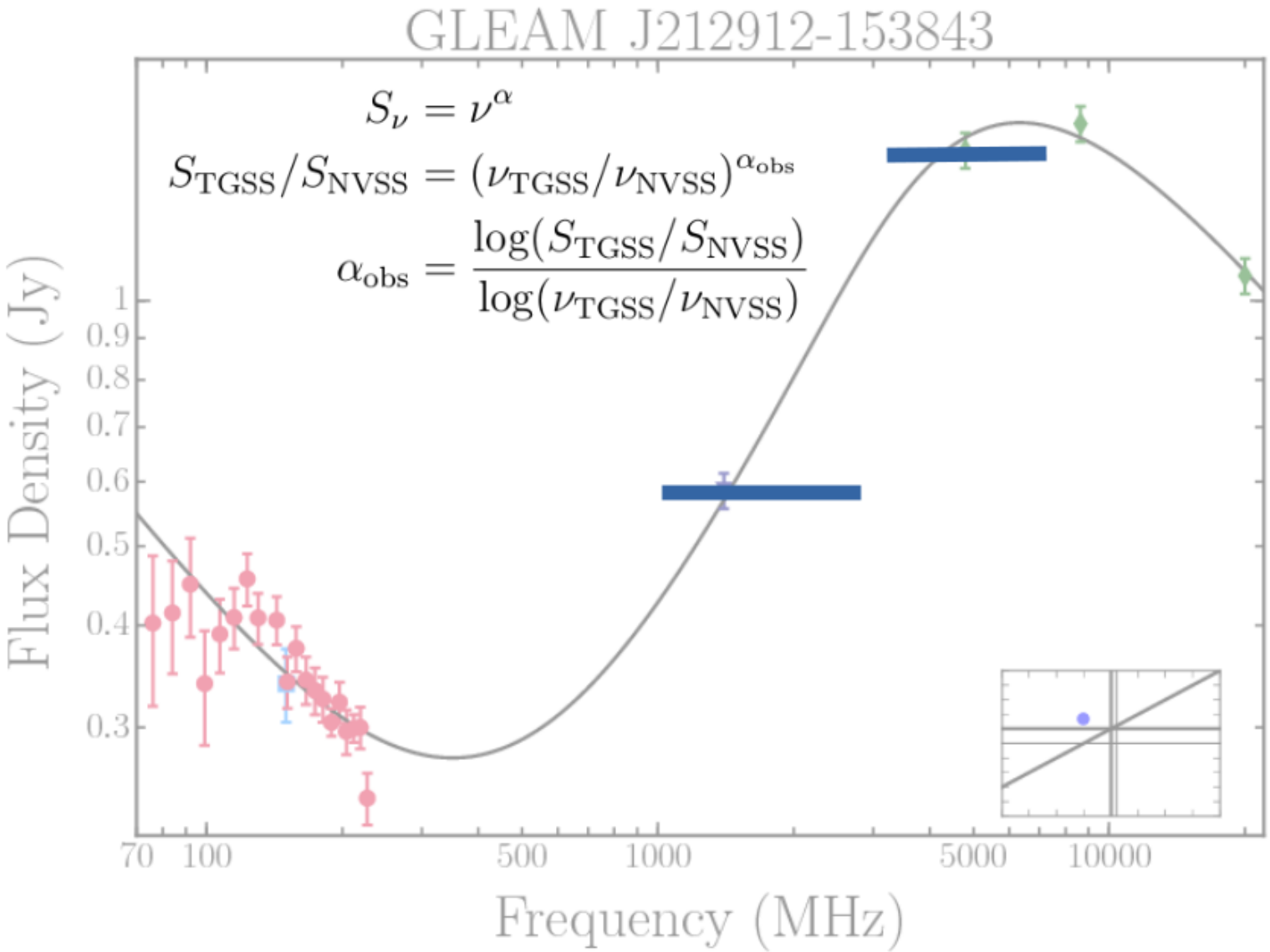
Success of classifier trained on positional matching

Failure of naive transitive partitioning

Positional matching



Spectral index



Spectral index comparison

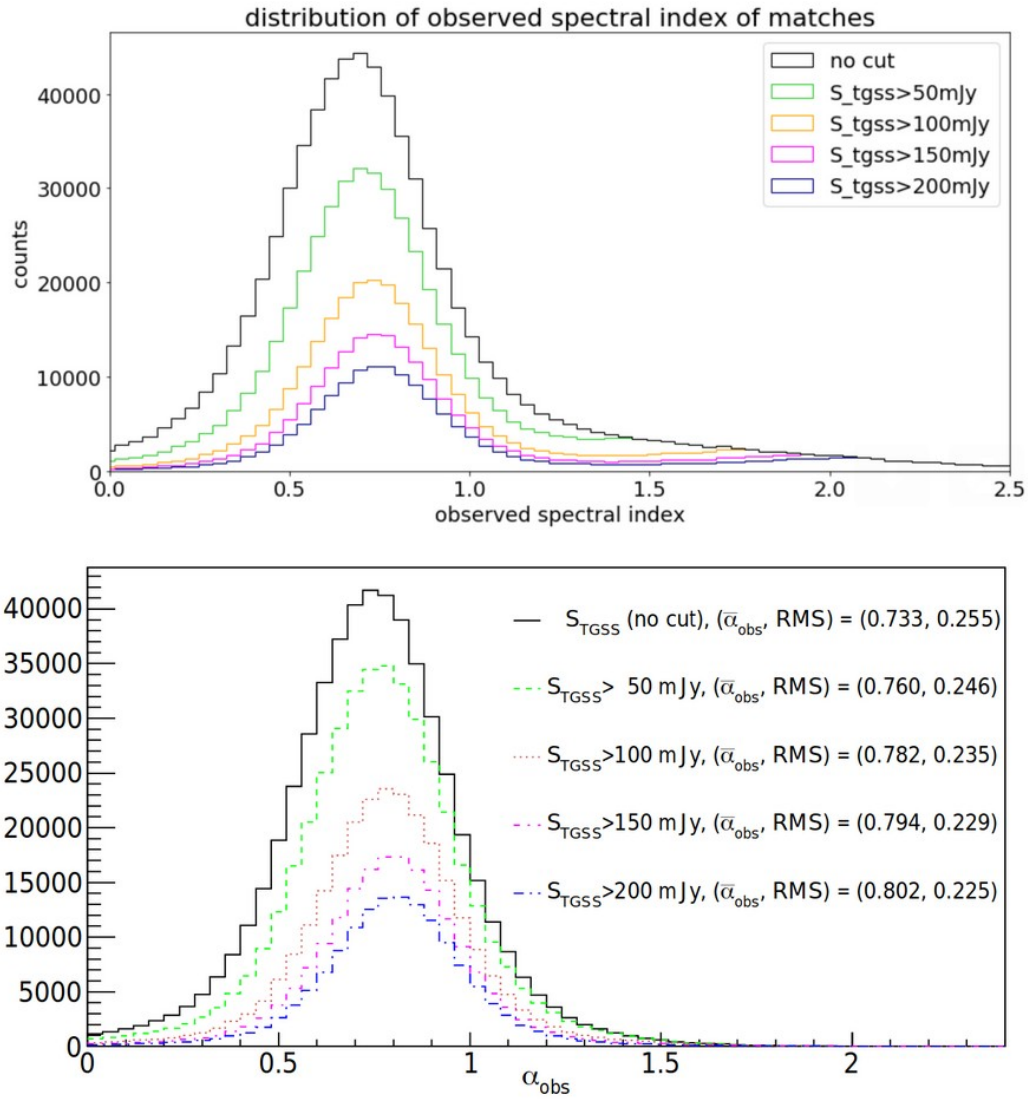


Figure: comparison to Tiwari 2019

Supervised machine learning

$$f : X \longrightarrow Y$$

Supervised machine learning

$$f : X \longrightarrow Y$$

Logistic regression

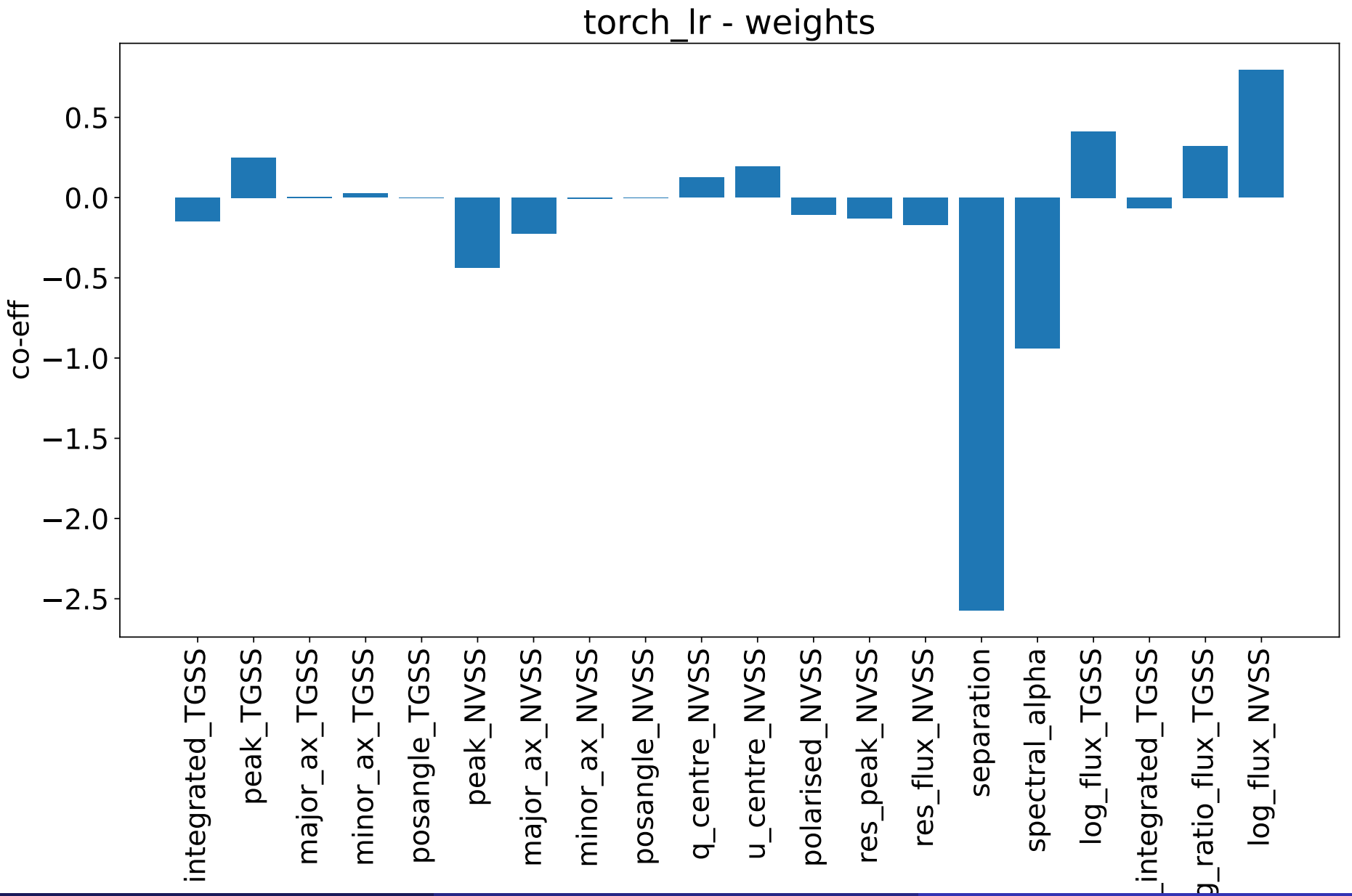
$$f(\vec{x}) = \sigma(\vec{\theta} \cdot \vec{x} + \theta_0)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

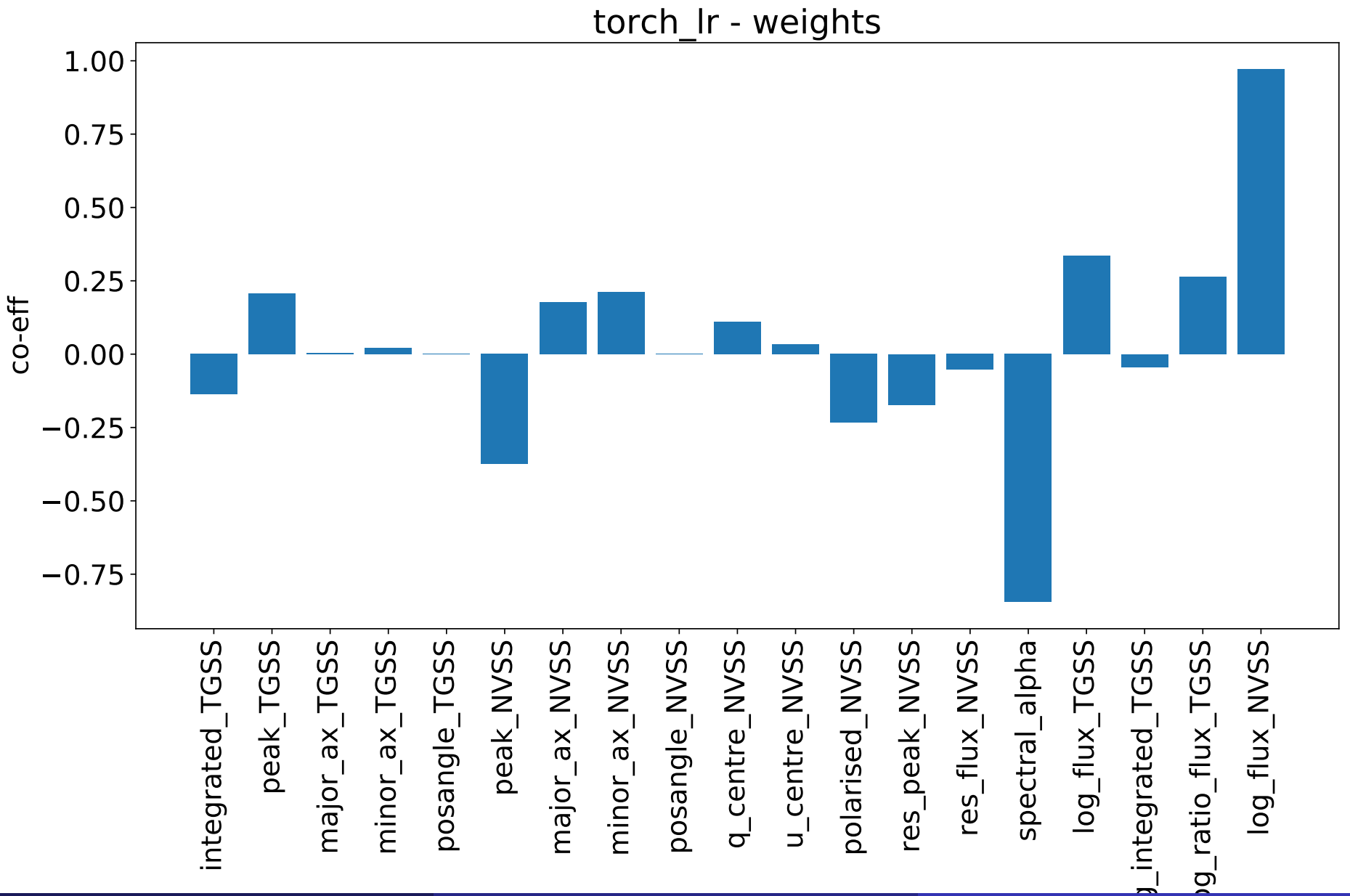
Binary cross entropy

$$
\begin{aligned}
H(y) &= \mathbb{E}_y[-\log_2(y)] \\
H_f(y) &= \mathbb{E}_y[-\log_2(f)] \\
&= -(y \log_2(f) + (1-y)\log_2(1-f))
\end{aligned}
$$

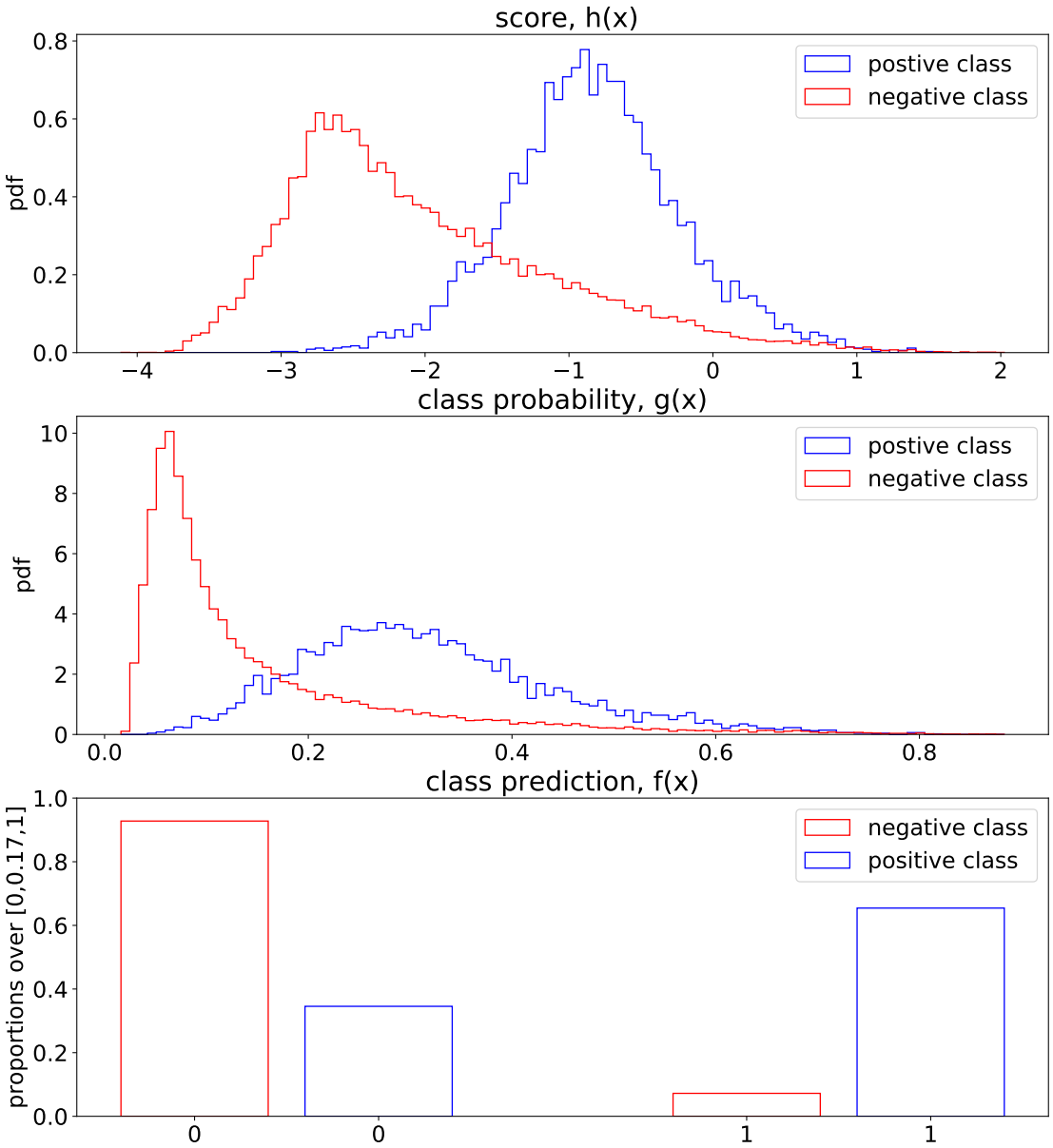
Logistic regression weights



Logistic regression weights - no separation



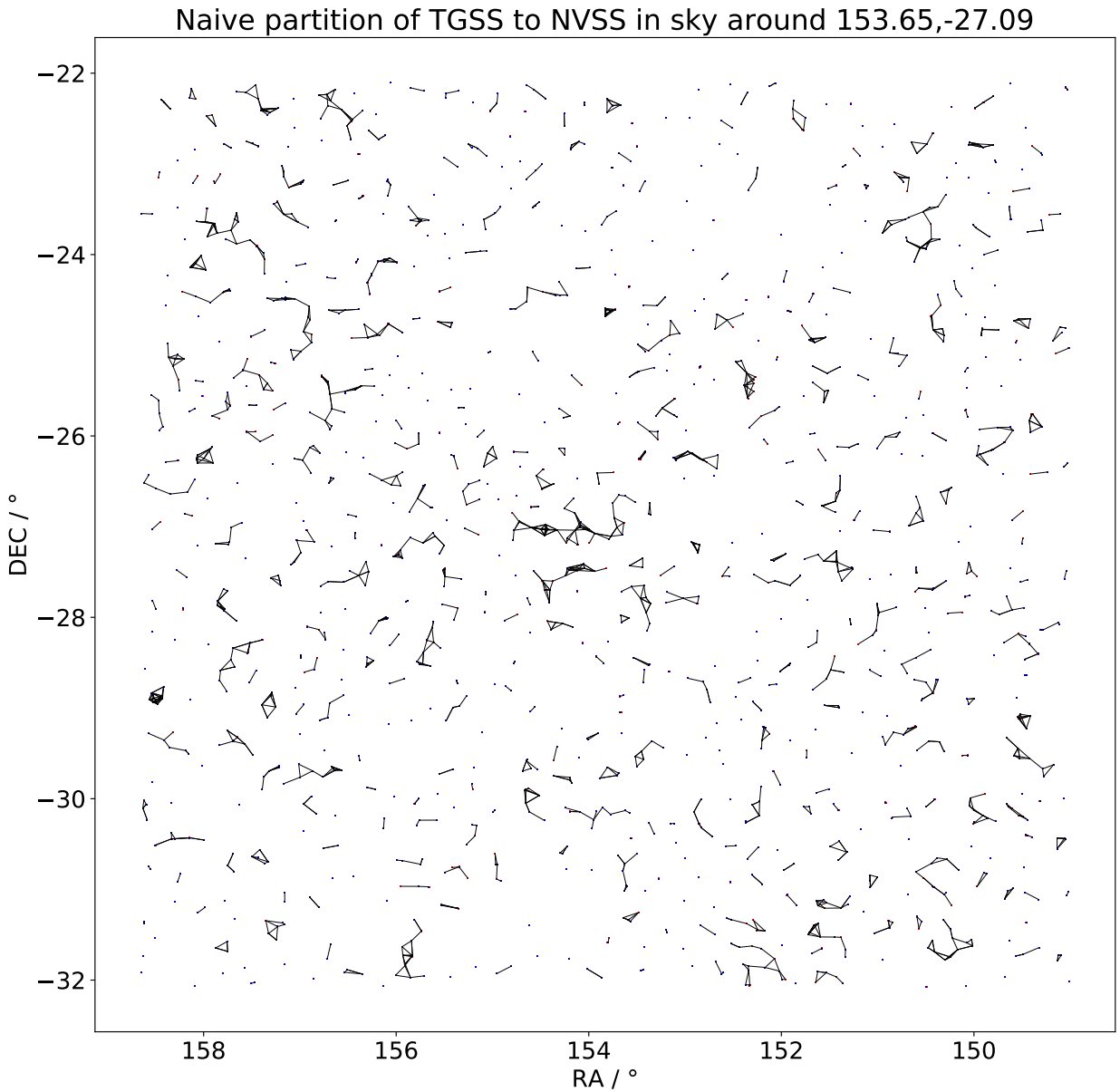
Logistic regression predictions



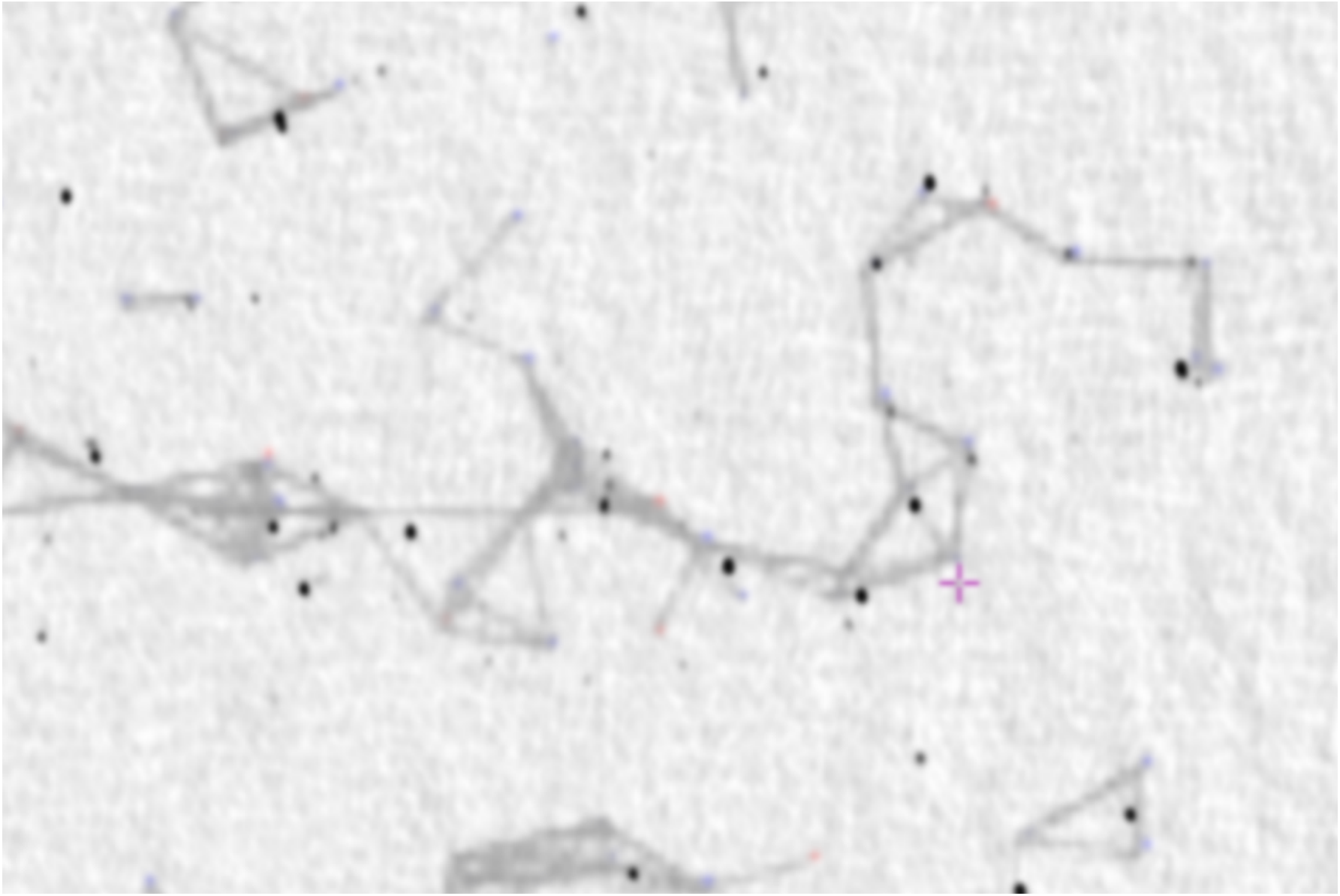
Logistic regression accuracy

		accuracy	precision	recall
all features	over patch catalogue	75.7	42.3	90.2
all features	over manual labels	80	80	100
no separation	over patch catalogue	70.3	36.9	88.2
no separation	over manual labels	80	80	100

Naive transitive partitioning



Partition overlay



Conclusions

Reproduction of existing results

Success of classifier trained on positional matching

Failure of naive transitive partitioning

Blob-match: Machine learning for cross-identification of radio surveys

James Gardner, supervisors: Cheng Soon Ong, Matthew Alger

October 22, 2019