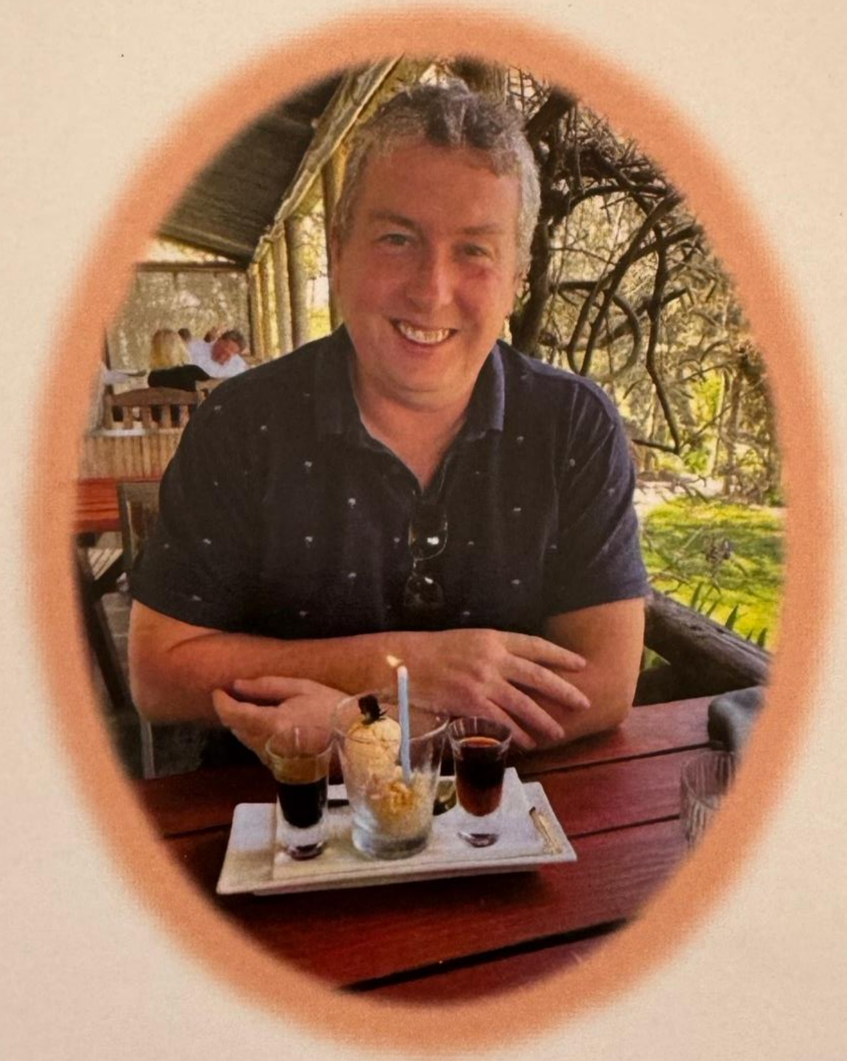
CLASSIFYING MARITIME VESSEL BEHAVIOUR FROM **AIS** USING STATISTICAL FEATURES

James Cormack, Matthew Roughan and Hung Nguyen

University of Adelaide

Nov, 2024



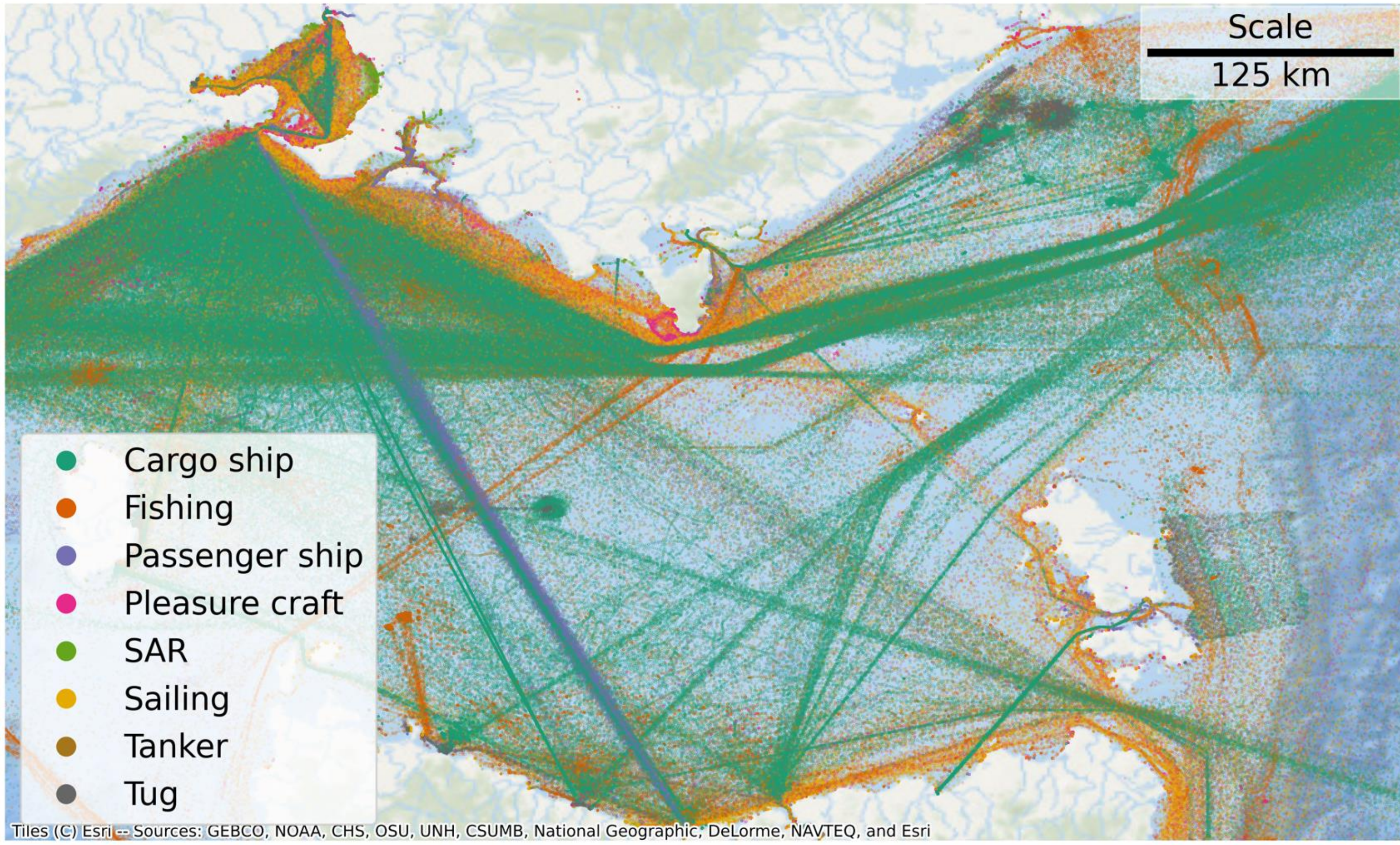
James John Cormack

1st October 1975 ~ 20th September 2024

MARITIME BEHAVIOUR CLASSIFICATION MOTIVATION

- Early SAR (Search & Rescue) detection of ships and boats in trouble
  - Fast response is crucial to good outcomes
- Maritime policing
  - Illegal fishing
  - Illegal immigration
  - Trafficking
  - Piracy
  - Unexpected military activity

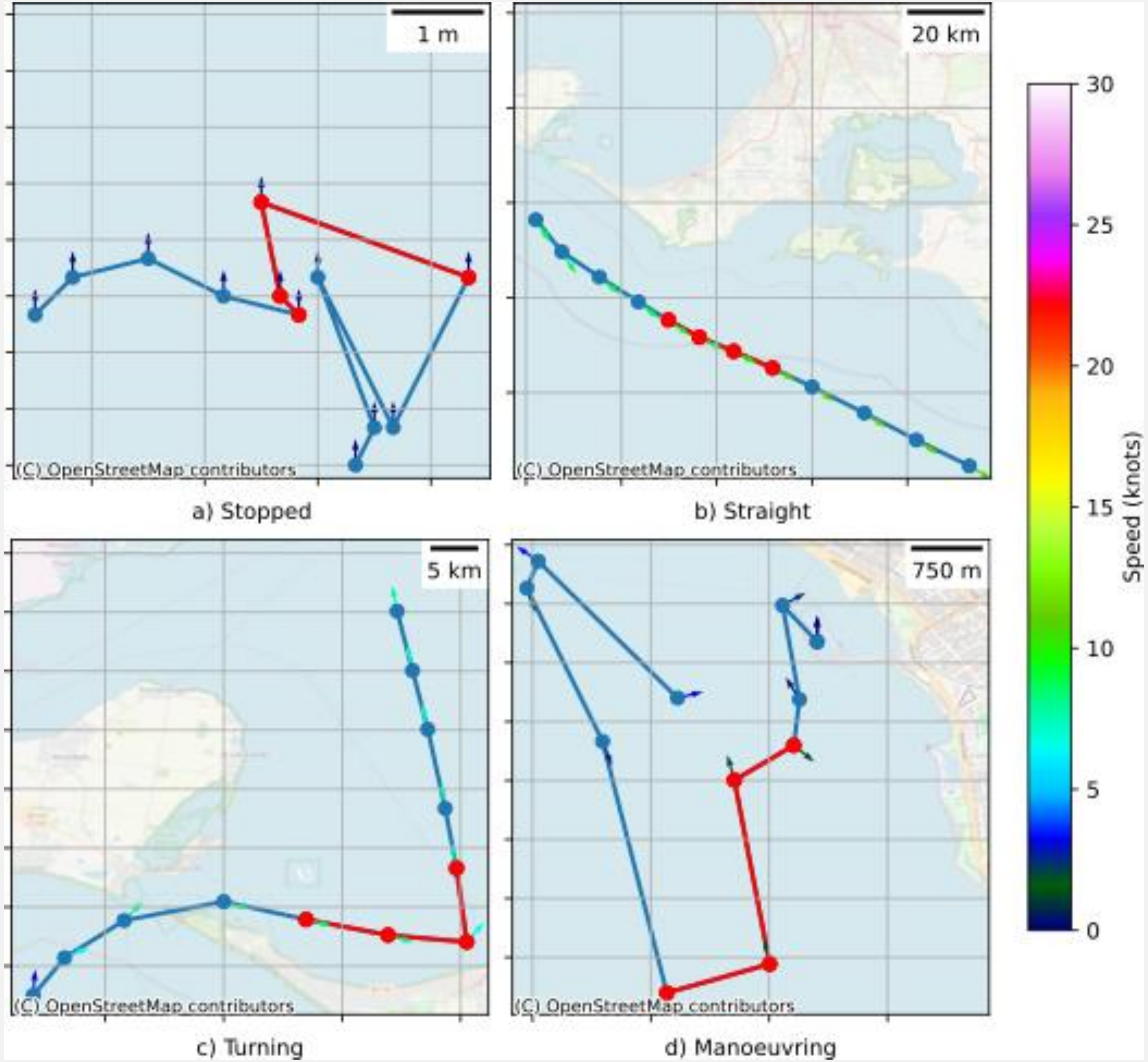
AIS DATA



- Australia
  - Large coastline (6th longest in the world
  - Remote locations
- AIS = Automatic Identification System
  - Transceivers on ships send information at intervals
  - Information = position, course, speed, ID, name, type, …
  - Received onshore base stations or by satellites (or other ships)
  - Primary use is collision avoidance
- AIS (publicly) available for many ships (over 300,000)
  - Required for larger ships (anything over 300 gross tonnage
  - Required for all commercial passenger ships
  - Since 2006 possible for smaller ships including many leisure craft

THE PROBLEM

- Classify ship behaviour from AIS data
  - Short segments so that classification is timely
  - Initially classify standard behaviours: stopped, turning, straight, maneuvering
  - Towards detecting anomalous behaviours
- Challenges
  - Data is not BIG (by modern ML standards)
    - Small number of variates
    - 1000's of tracks, but that isn't massive
  - Data isn't 'nice'
    - Irregular time intervals
    - Missing data


1 m
(C) OpenStreetMap contributors
a) Stopped
20 km
(C) OpenStreetMap contributors
b) Straight
5 km
(C) OpenStreetMap contributors
c) Turning
750 m
(C) OpenStreetMap contributors
d) Manoeuvring
Speed (knots)
0
5
10
15
20
25
30

APPROACH

- Standard supervised Machine Learning (ML)
  - Segments of 4 AIS data points (approximately 1 hour)
    - Enhanced features created to expand the input
    - >60 features
  - ML classifier approaches:
    - Multi-layer Perception (MLP)
    - Random Forest (RF)
    - k-Nearest Neighbour (k-NN)
  - Tested different feature set/classifier combinations

DATA DETAILS

- Data from Australian Maritime Safety Authority (AMSA)
  - Publicly available, anonymized
  - Approximate 15-minute sampling interval (reduced from original in some cases)
  - Some missing data
  - Accuracy unknown, but assuming GPS origin ~10m
  - Includes different types of vessel: cargo, tanker, passenger, pleasure craft, fishing, sailing, tug, and SAR (search and rescue)
- Dataset
  - Jan 2021 – March 2023
  - Bass Strait region (-41.2 to -37.7 latitude, 143.6 to149.3 longtitude)
    - High shipping volume
    - Challenging environment

GROUND TRUTHING

- Primary data was manually classified
  - GUI interface developed to allow track classification
  - Uses 4 key points, plus context (4 additional points on either side of the track)
    - These are not used in the ML
- Secondary classification
  - Original traces were asymmetrically distributed
  - To enhance the more limited classes, semi-automated rules were used to generate candidate tracks

Label	Samples
Turning	1151
Stopped	1151
Straight	1151
Manouevring	746

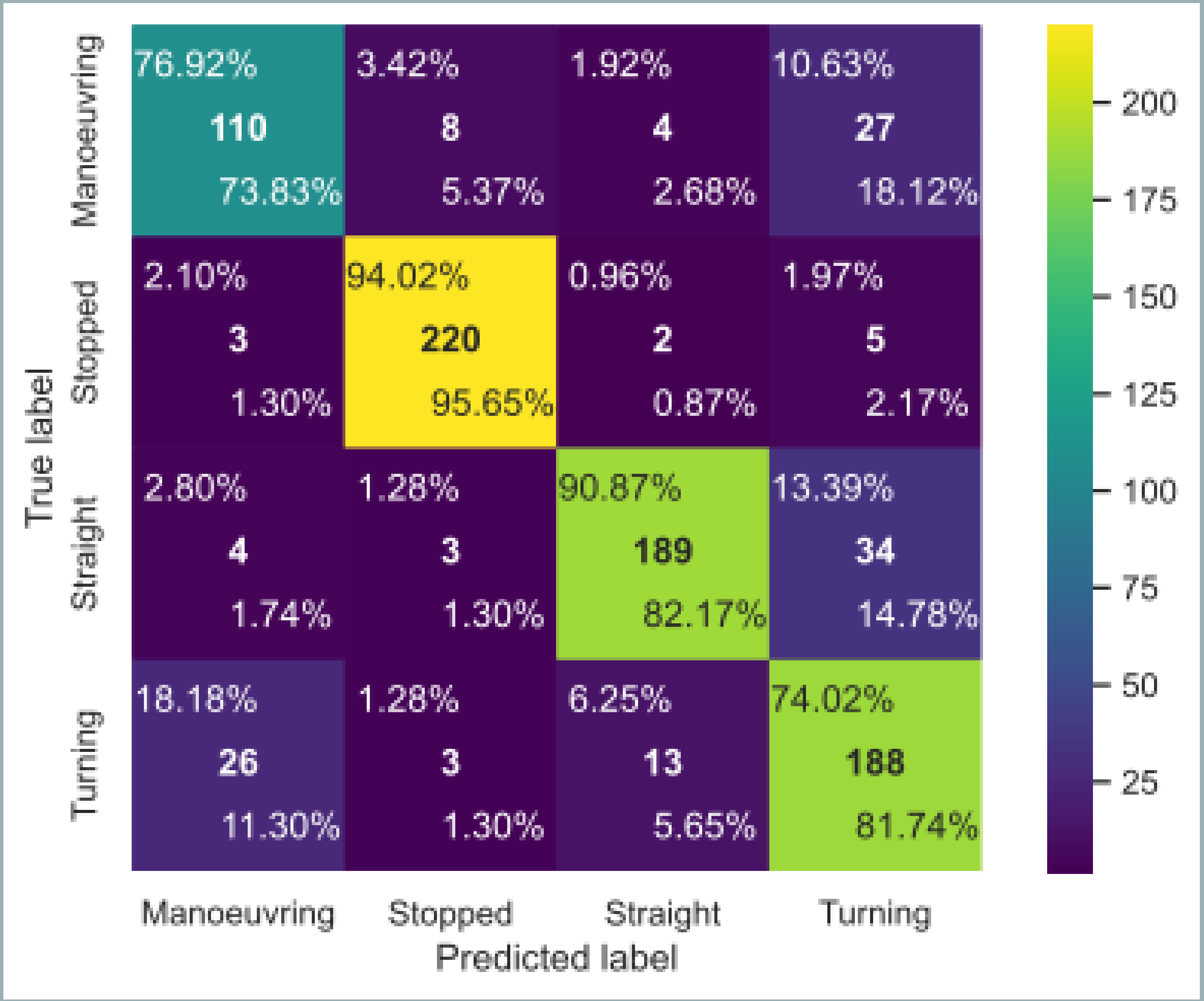
64/20/16 split

RESULTS

Best approach = MLP

with top-20 (Gini) features

with F1 = **83.6%**

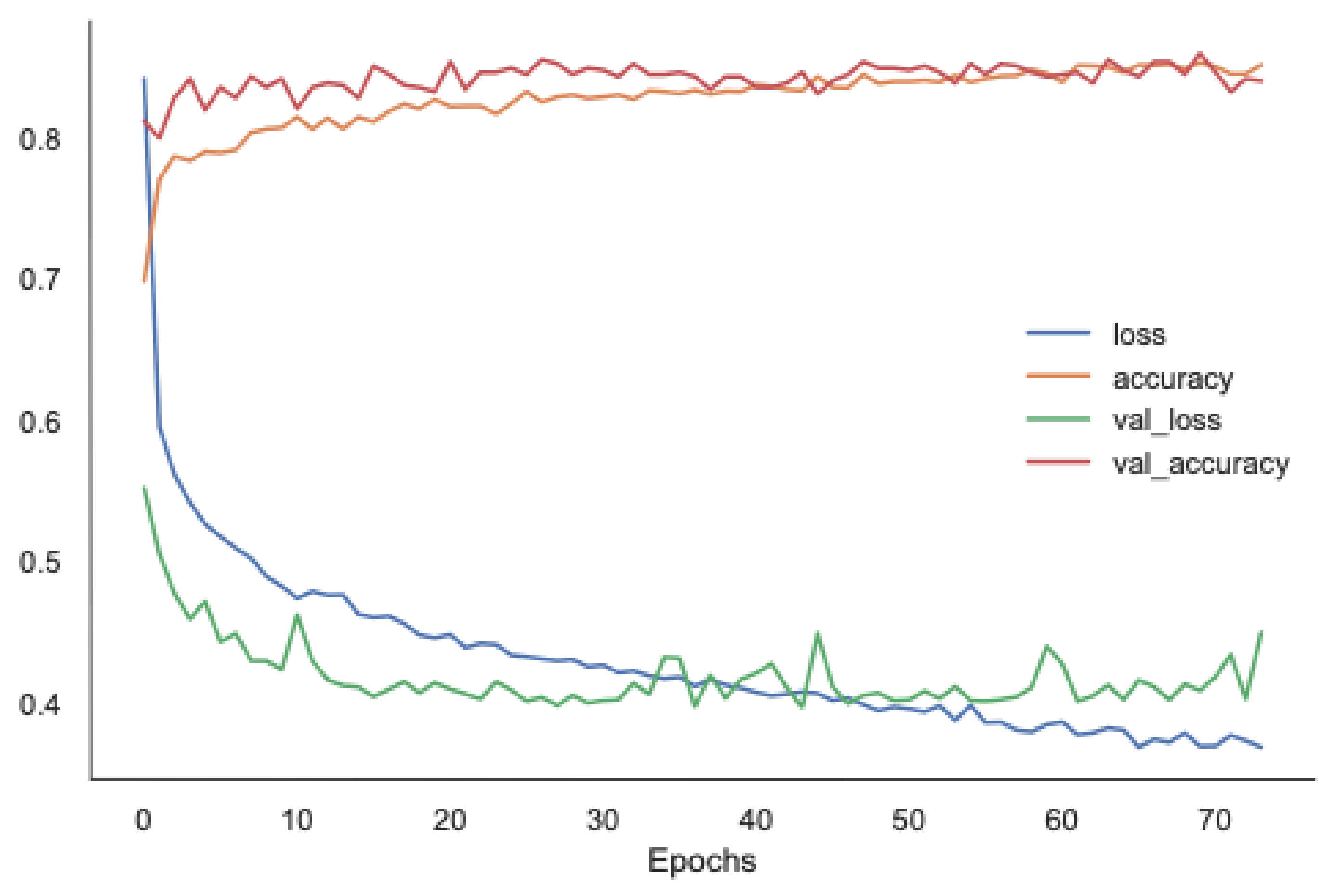


CONCLUSION

- We can obtain high accuracy classification with very limited data
  - F1 of 83% from 4 data points

- Future
  - ML not extensively explored
  - Data could be expanded in several ways
  - Next stage is detecting anomolies

TRAINING


0.8
0.7
0.6
0.5
0.4
loss
accuracy
val_loss
val_accuracy
0
10
20
30
40
50
60
70
Epochs

FEATURE EXAMPLES

- **Avg_Interval:** Average interval between all track points.
- **Pcw_Std_Course:** Standard deviation of Course over all track points.
- **Pcw_Avg_Accel:** Average acceleration over all track points.
- **Pcw_Max_Speed_Change:** Maximum change in Speed over all track points.
- **Pcw_Avg_Dist2Coast:** Average distance to coast over all track points.
- **Str_Change_In_DOT:** Difference in DOT between the first track point and the last track point.
- **Pcw_vs_Str_Speed_Diff:** Difference between the piecewise average speed and the straight line average speed.
- **Pcw_Avg_Course_vs_DOT:** Difference between the piecewise average course and the piecewise average DOT

DOT = direction of travel

GINI IMPORTANCE

GINI order	Description
1	Distance travelled - distance between first and last track points
2	The average distance travelled between segments
3	The total distance travelled over all segments
4	The average change in direction over all segments
5	The cumulative change in direction over all segments
6	The average velocity over all segments
7	The number of stops velocity < 0.5knots
8	The standard deviation of the direction of travel over all segments
9	The average velocity between the first track point and the last
10	The standard deviation of the course over all segments

COMPARISON OF CLASSIFIERS

Classifier	Feature Set	F1-Score
MLP	Top 20	83.6%
MLP	Top 15	83.6%
MLP	Top 20	83.5%
RF	All	82.9%
KL-NN	Top 10	80.4%