POSTER P01-05

Machine learning based wafer sort yield prediction based on wafer acceptance test data

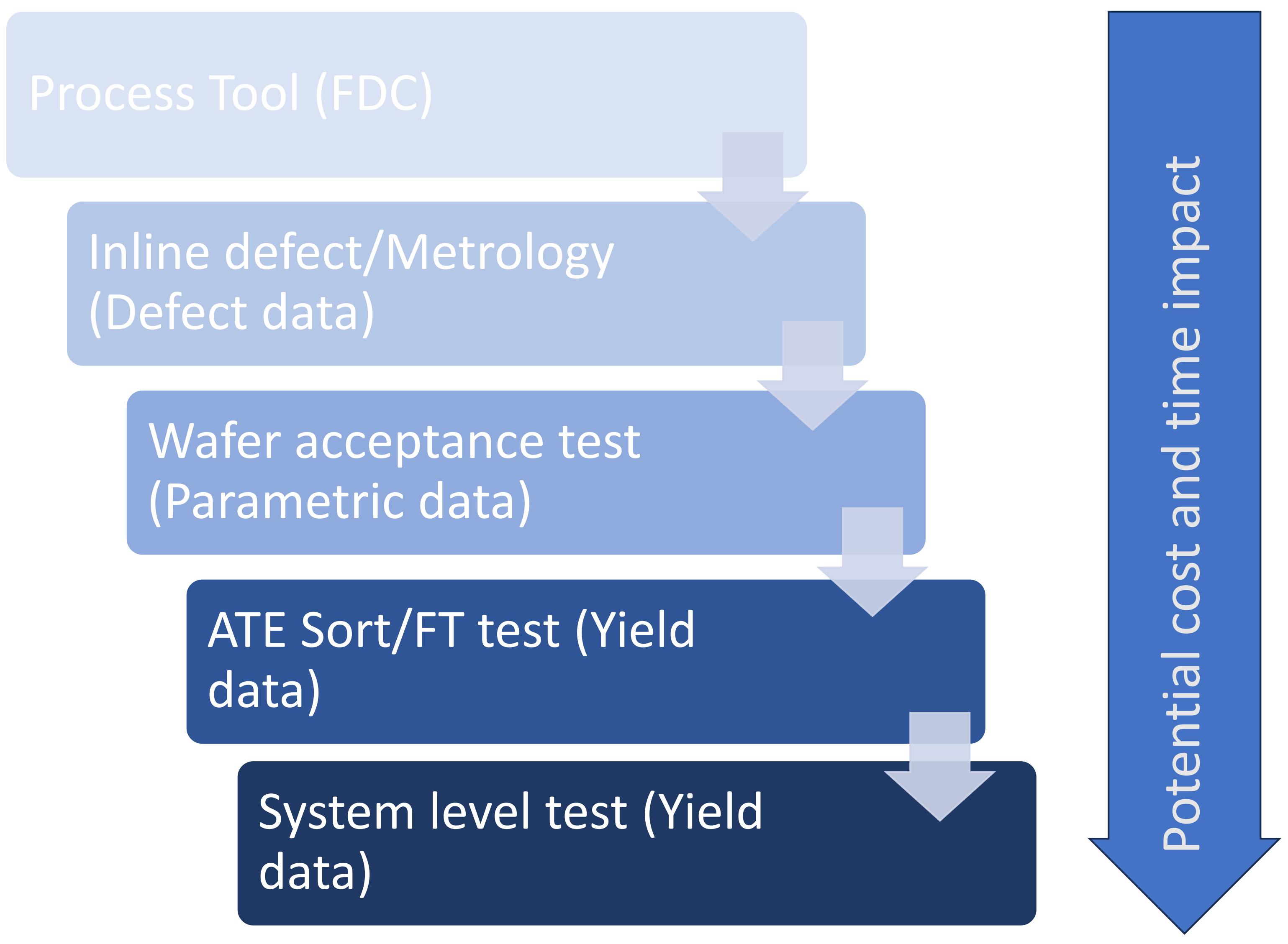
YH Chan, MH Thor, Amalini M, JY Chua, YK Teoh, CQ Chen
GlobalFoundries

Introduction

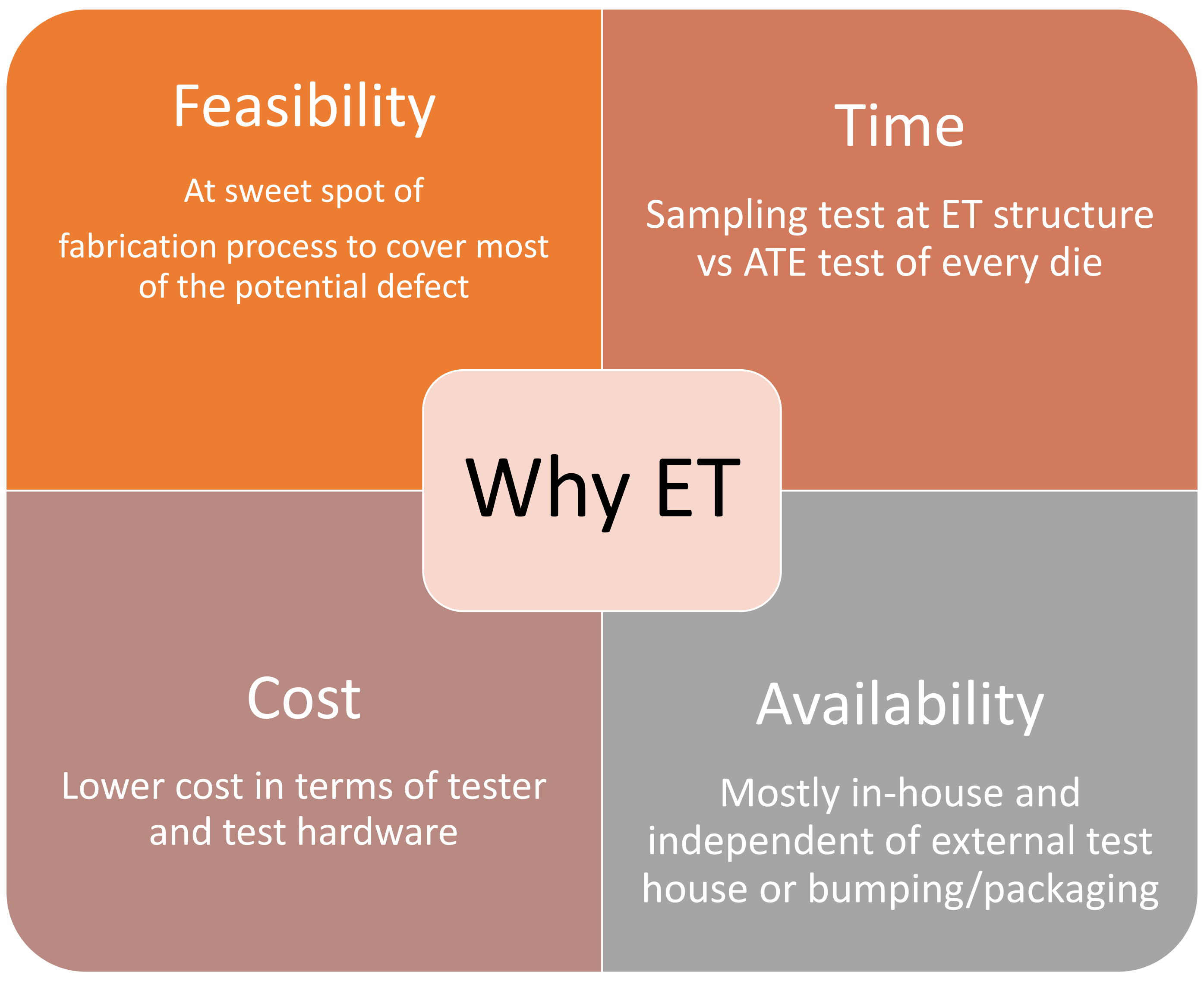
- Automatic Test Equipment (ATE) or System Level Test (SLT) provide best coverage to assess yield but feasible only at the end of fabrication cycle with additional cost and time

- An early yield prediction method will detect potential yield issue earlier in the manufacturing stages resulting in cost and time saving as well as better production planning

- A machine learning based methodology is developed to predict yield with good accuracy based on cost and time efficient wafer acceptance (ET) test data



Process Tool (FDC)
Inline defect/Metrology (Defect data)
Wafer acceptance test (Parametric data)
ATE Sort/FT test (Yield data)
System level test (Yield data)
Potential cost and time impact



Why ET

Feasibility
At sweet spot of fabrication process to cover most of the potential defect

Time
Sampling test at ET structure vs ATE test of every die

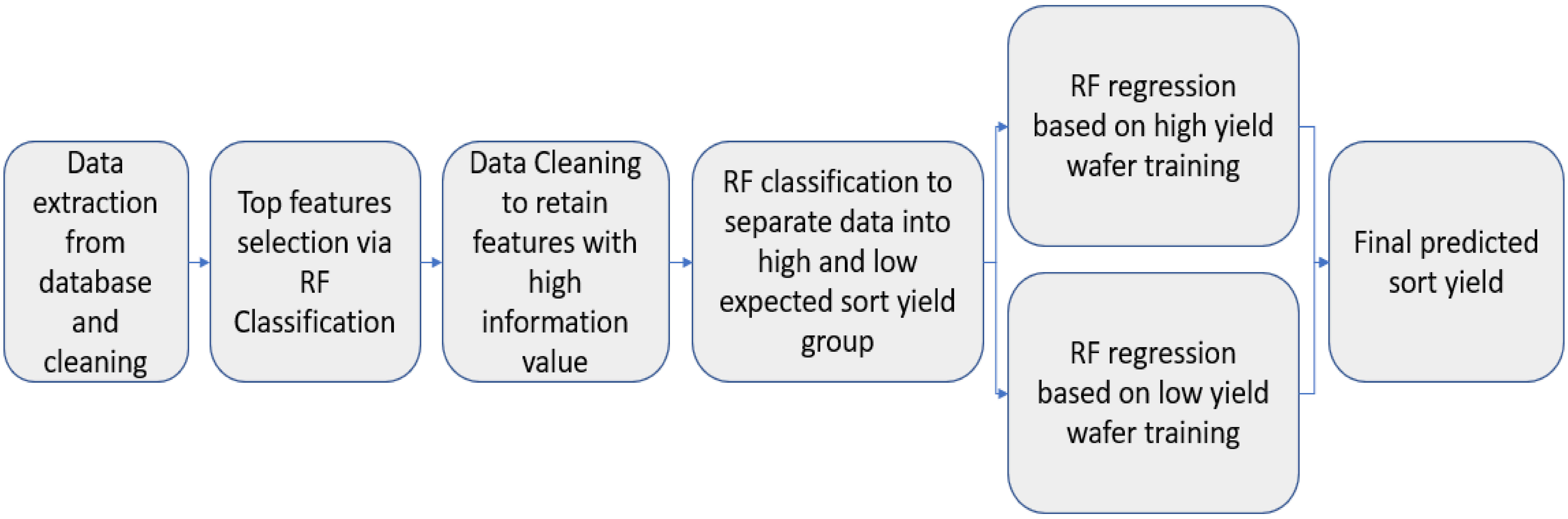
Cost
Lower cost in terms of tester and test hardware

Availability
Mostly in-house and independent of external test house or bumping/packaging

Methodology development

Overview:

- Data cleaning and features selection
- Machine learning algorithm selection / hyperparameter tuning
- Training methodology
- Result discussion



Data extraction from database and cleaning
Top features selection via RF Classification
Data Cleaning to retain features with high information value
RF classification to separate data into high and low expected sort yield group
RF regression based on high yield wafer training
RF regression based on low yield wafer training
Final predicted sort yield

Data cleaning and features selection

- Software developed to automatically clean and normalized raw data
- Features of importance study to select optimal number of features based on data available

	High yield (>70%)		Low yield (<70%)	
	Delta	Rsquare	Delta	Rsquare
All features (614)	1.75%	0.341	2.63%	0.139
Top 500 features	1.73%	0.354	2.56%	0.129
Top 400 features	1.73%	0.372	2.02%	0.148
Top 300 features	1.71%	0.343	2.22%	0.197
Top 200 features	1.73%	0.393	2.34%	0.22
Top 100 features	1.61%	0.405	1.95%	0.21
Top 50 features	1.59%	0.423	2.00%	0.334
Top 25 features	1.62%	0.42	2.29%	0.372

Remove features with more than 50% missing data

Remove data points that have the maximum and minimum value within feature and replace them with median value

Performed normalization within feature to give each value an adjusted value between 0 and 1

Rearrange data in Column/Row format compatible to Google Colab

Machine learning algorithm selection and training methodology

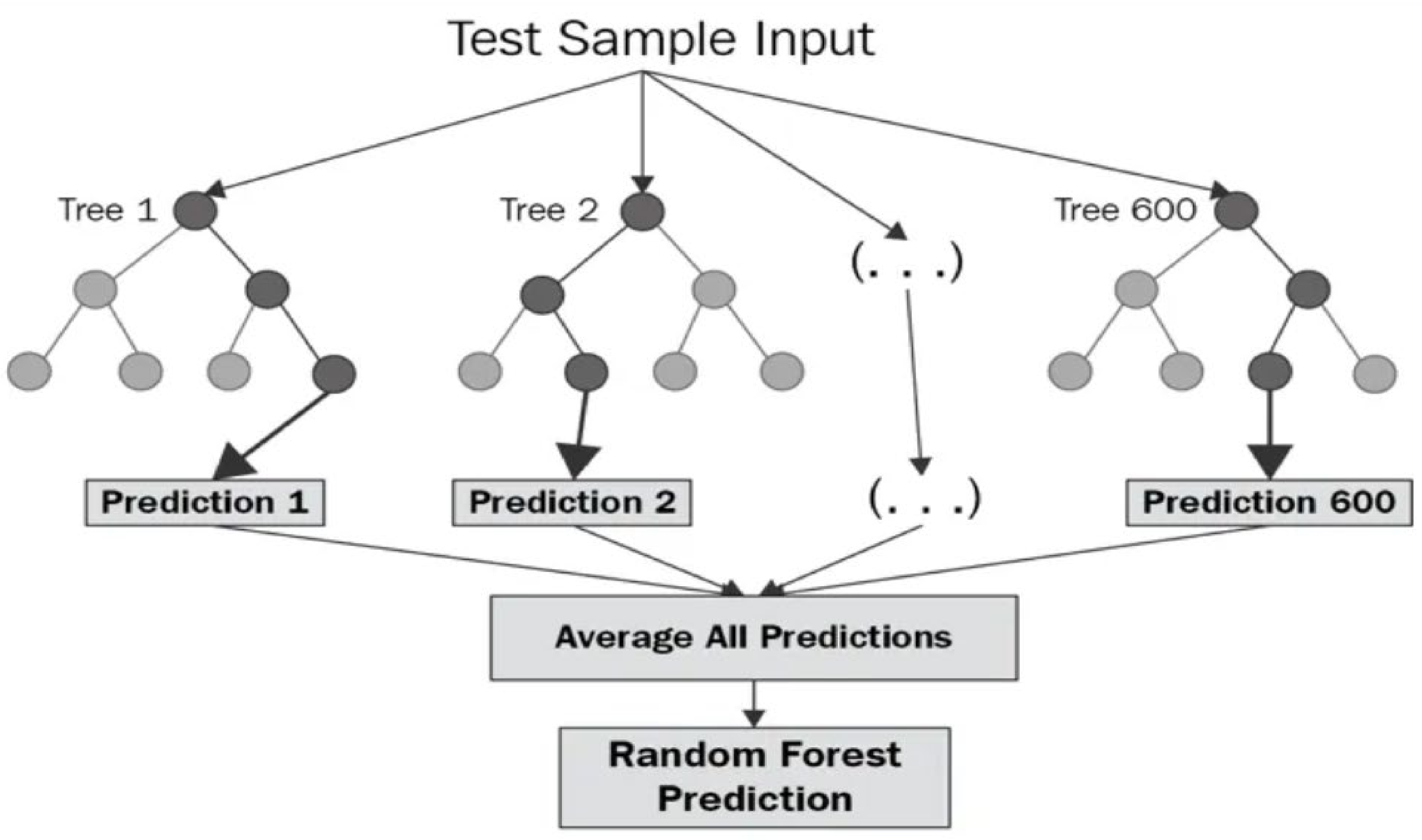
Algorithm selection:

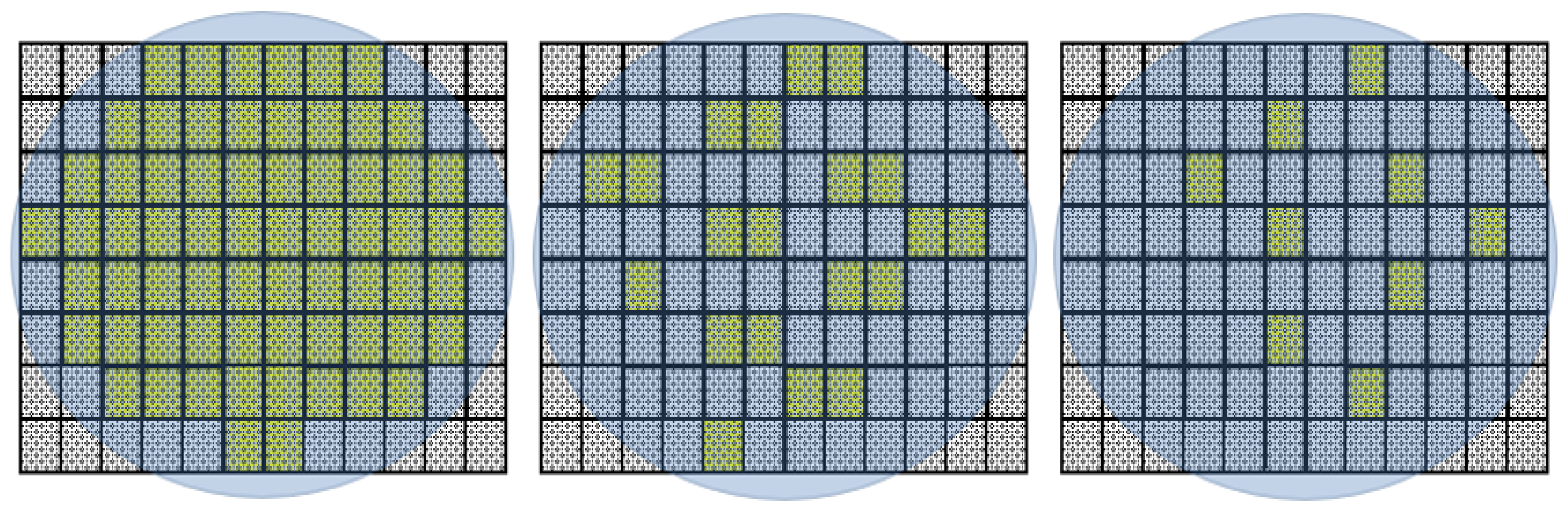
- Two high volume products were selected with 20% data retention to compare performance of various machine learning method
- Random forest method consistently yield best result in both classification and regression due to its robustness in handling noisy data, outliners and preventing overfit

Model Name	Test Accuracy
Decision Tree Classifier	67.8%
Logistic Classification	60.7%
Random Forest Classifier	75.0%
Neural Network	32.1%
Gradient Boosting Classifier (GBM)	64.2%
XGBoost	57.1%
Ensemble Random Forest Classifier	78.0%

Training method:

- Experiment result shows the benefit of increasing ET sites tested is insignificant. Using standard production sampling of 9 ET sites per wafer will be a good trade off between prediction accuracy and feasibility
- A first stage classification training separate wafers into high and low yield group first to resolve data imbalance issue
- Each of the high and low yield group data will be trained individually using regression method to give the final prediction




ET Full-map with all 66 reticles tested

ET partial map with 20 reticles tested

ET Production Setup with 9 reticles tested

Wafer	Predict(All sites average)	Actual	Delta %
1	92.30	92.30	2.45
2	92.36	92.36	2.28
3	92.47	92.47	2.33
4	92.62	92.62	2.13
5	92.78	92.78	1.94
6	92.80	92.80	2.17
7	92.96	92.96	2.07
8	93.18	93.18	2.36
9	93.37	93.37	1.15
10	93.39	93.39	1.38
		Average Delta	2.02
		Rsquare	0.65

Wafer	Predict(20 sites average)	Actual	Delta %
1	95.00	92.30	2.70
2	94.79	92.36	2.43
3	95.14	92.47	2.67
4	94.82	92.62	2.20
5	95.04	92.78	2.26
6	95.18	92.80	2.38
7	94.86	92.96	1.90
8	95.19	93.18	2.01
9	94.69	93.37	1.32
10	94.73	93.39	1.34
		Average Delta	2.12
		Rsquare	0.50

Wafer	Predict(9 sites average)	Actual	Delta %
1	95.03	92.30	2.71
2	94.82	92.36	2.68
3	95.02	92.47	2.46
4	94.92	92.62	2.57
5	94.99	92.78	2.09
6	95.10	92.80	2.39
7	94.86	92.96	1.95
8	95.08	93.18	1.98
9	94.73	93.37	1.40
10	94.68	93.39	1.60
		Average Delta	2.18
		Rsquare	0.46

Result and discussion

- Two products with > 10K wafers data were selected with 90% data used for training and 10% data retained for testing. Both LT and M2 ET data were attempted. Training time for both product is less than 3hrs

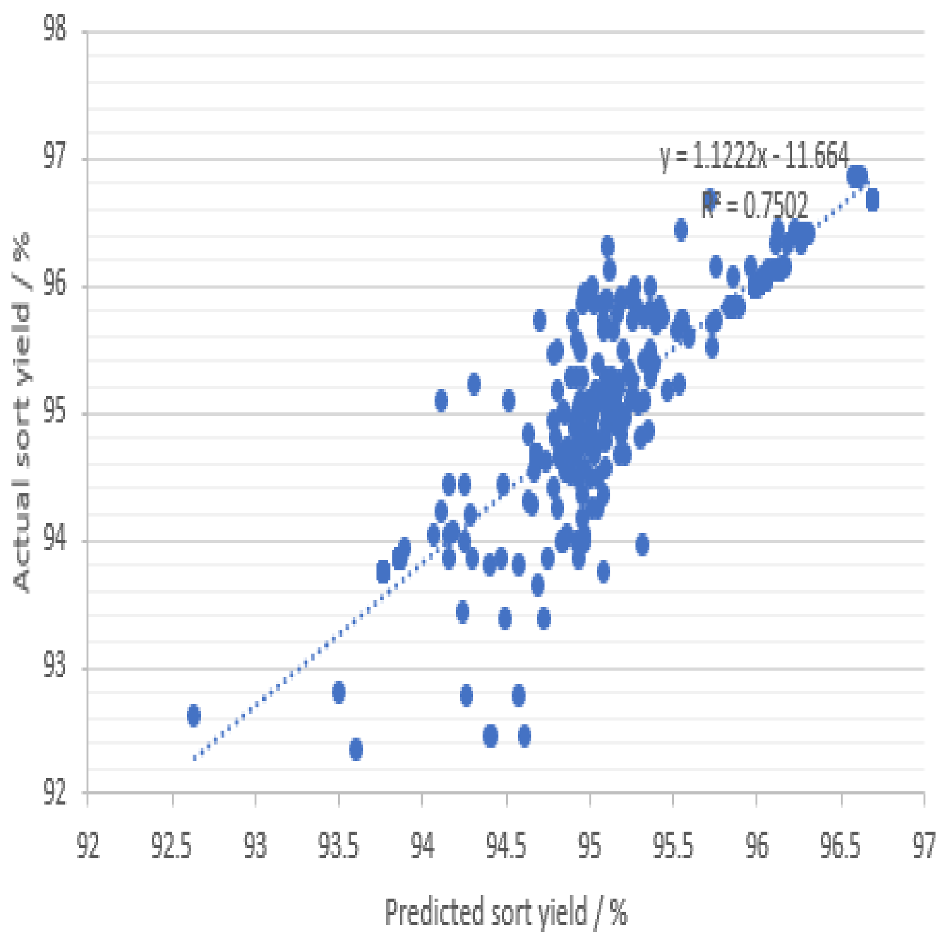
- Hold out test data shows good performance on LT ET data to sort yield prediction with < 0.5% delta on high yield wafers and < 7% delta on low yield wafer. Result is consistent for both product

- On M2 ET data to sort yield prediction high yield wafers show < 1% delta and low yield wafers show < 7% delta. Slight reduction in accuracy likely due to some BEOL induced defects cannot be covered as M2 ET measurements were made at metal two layer

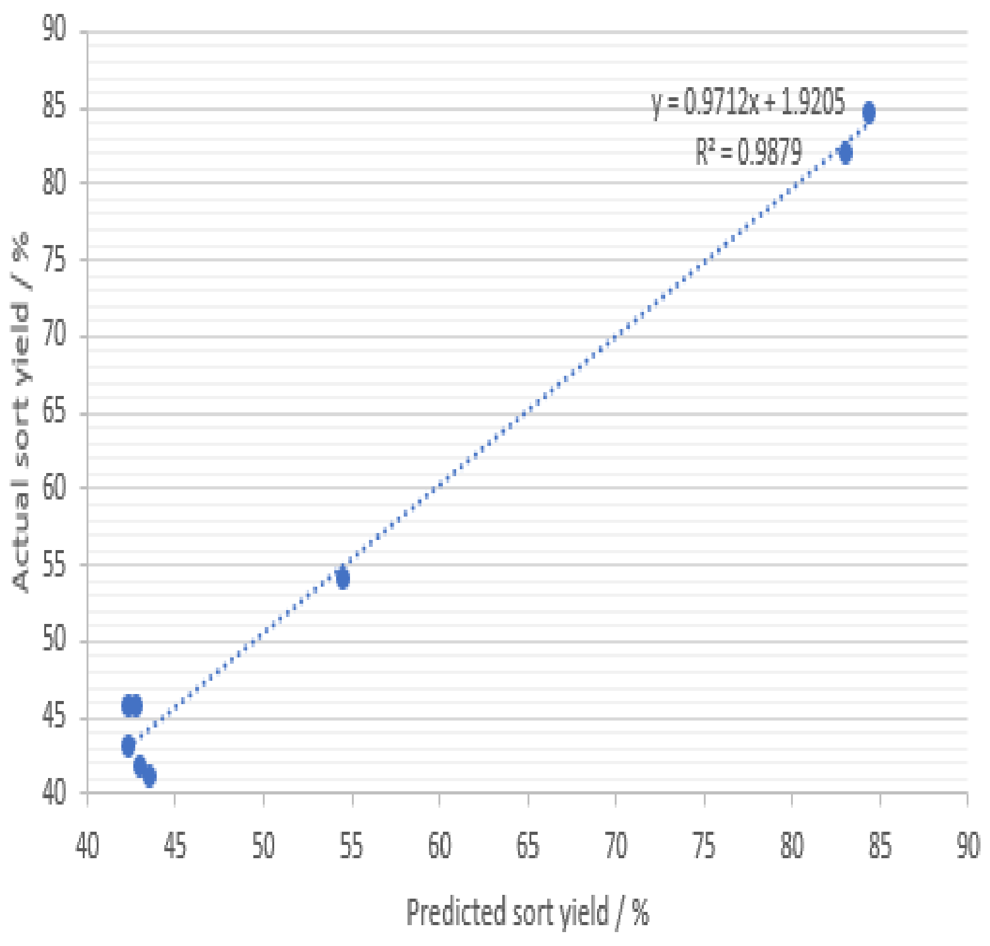
	Product A verification result			
	LT high yield model	LT low yield model	M2 high yield model	M2 low yield model
Verification count by wafer	250	7	1854	20
Average sort yield delta between actual and predicted	0.45%	6.93%	0.69%	4.52%
Rsquare	0.75	0.99	0.15	0.77

	Product B verification result			
	LT high yield model	LT low yield model	M2 high yield model	M2 low yield model
Verification count by wafer	288	8	995	10
Average sort yield delta between actual and predicted	0.29%	1.59%	0.66%	6.74%
Rsquare	0.77	0.99	0.21	0.68

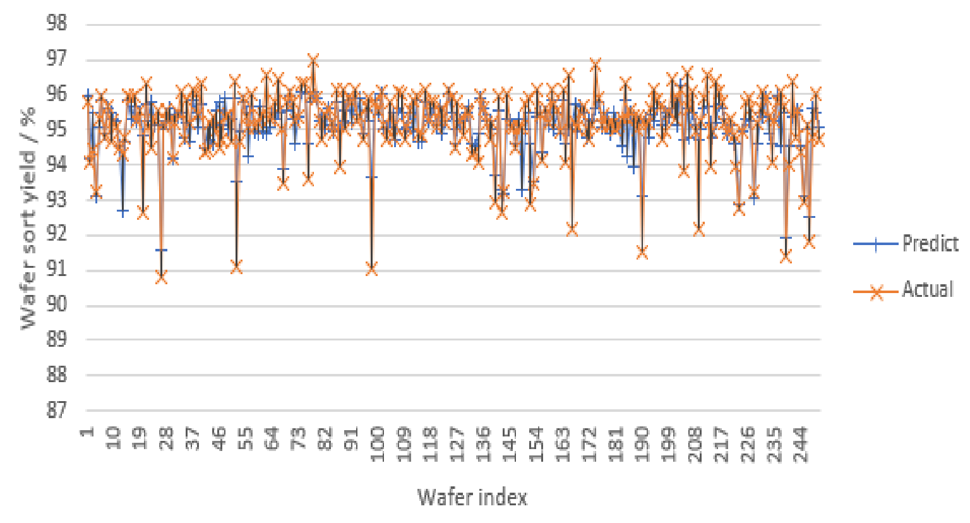
High yield model predicted vs actual wafer sort yield



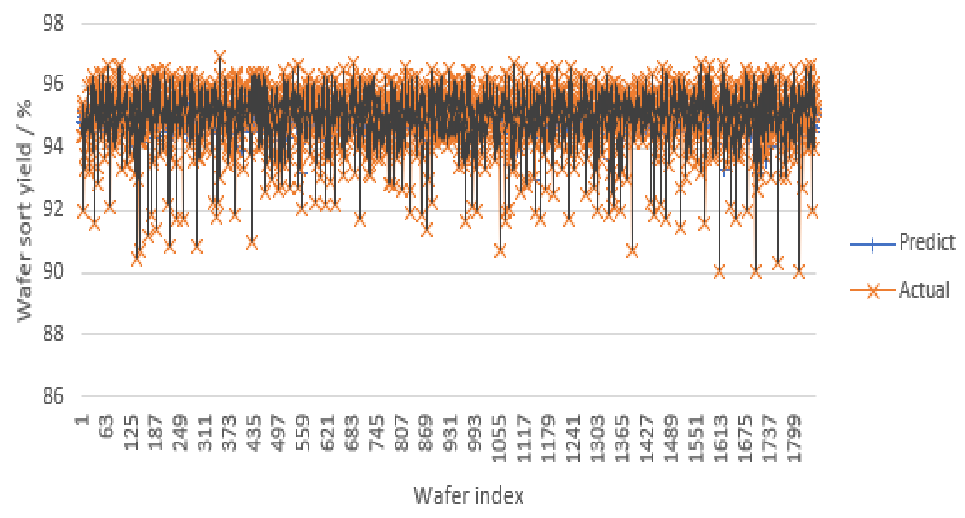
Low yield model predicted vs actual wafer sort yield



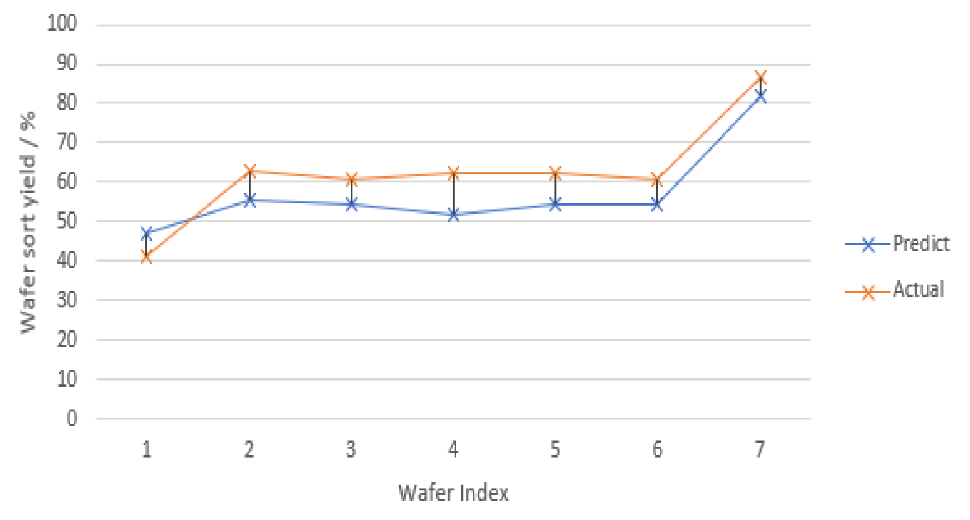
Chip A high yield wafers actual vs predicted sort yield based on LT ET data



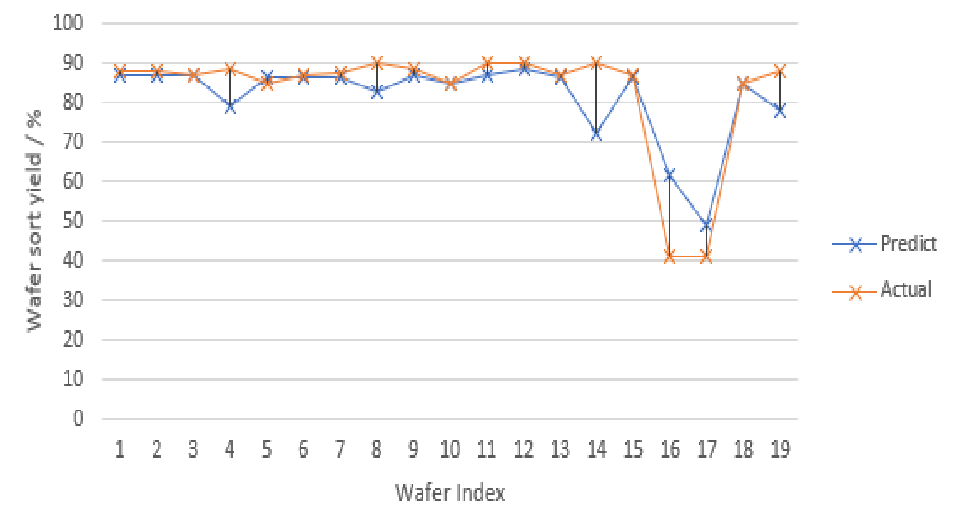
Chip A high yield wafers actual vs predicted sort yield based on M2 ET data

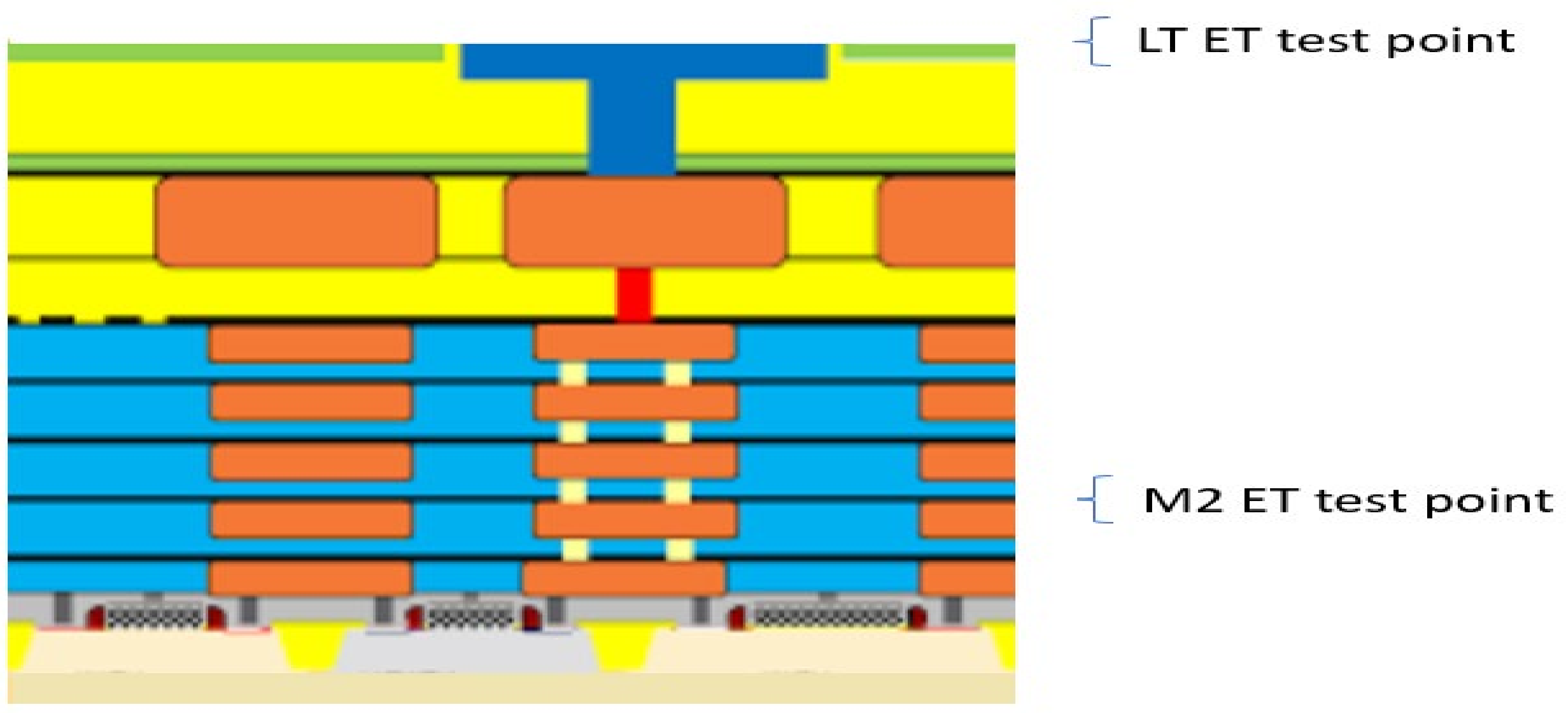


Chip A low yield wafers actual vs predicted sort yield based on LT ET data



Chip A low yield wafers actual vs predicted sort yield based on M2 ET data





Conclusion

- A machine learning based methodology was developed and demonstrated to predict sort yield with good accuracy and repeatability based on only wafer acceptance data

- The proposed two stage classification and regression model with optimized machine learning algorithm selection shows a prediction delta of < 7% for low yield wafer and < 1% for high yield wafer when using either LT or M2 ET data

- Future work involving ensemble of FDC and inline scan data for earlier manufacturing stage and more accurate prediction should be considered and examined

Contact information

- Chen Changqing (Changqing.CHEN@globalfoundries.com)
- Patrick Chan Yin Hong (yinhong.chan@globalfoundries.com)