

Model Introduction to Improve
the Coplanarity Prediction
Capability of New Semiconductor Products

半導體新產品平面度預測能力提升之模型導入

指導教授 施明昌 博士

學 生 陳筠霏



Contents

1. 研究動機

3. 研究架構

5. 文獻回顧

7. 實驗結果

2. 製程概述與平面度定義

4. 基礎理論

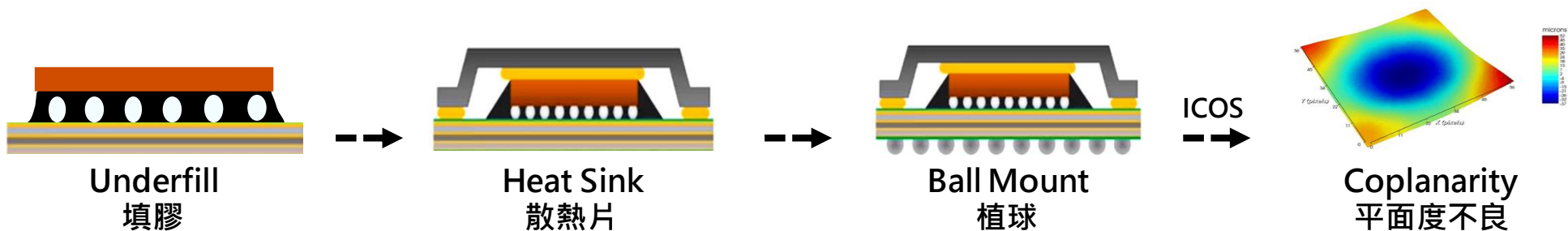
6. 模型設計

8. 結論與未來展望



研究動機

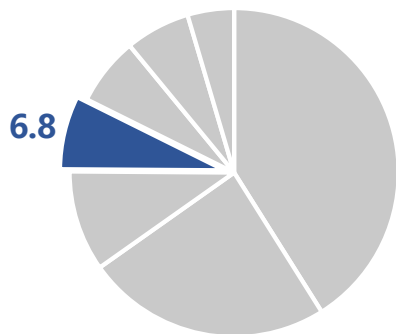
廠內產品在後段封裝過程中，晶片與基板不同材料之間的 **CTE Mismatch**、參數組合的差異，加上含膠的限制，為了釋放溫度所產生的內部應力，常藉由形狀改變來釋放這些內力，導致產品 **Coplanarity fail**，進而影響後續上板成功率、出貨時間。



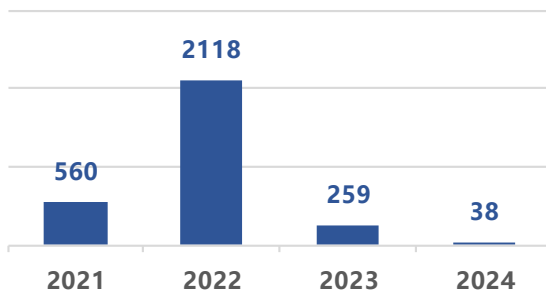
進而 **影響** 後續上板成功率、出貨時間。

符合 **預測準確率** 僅 11%

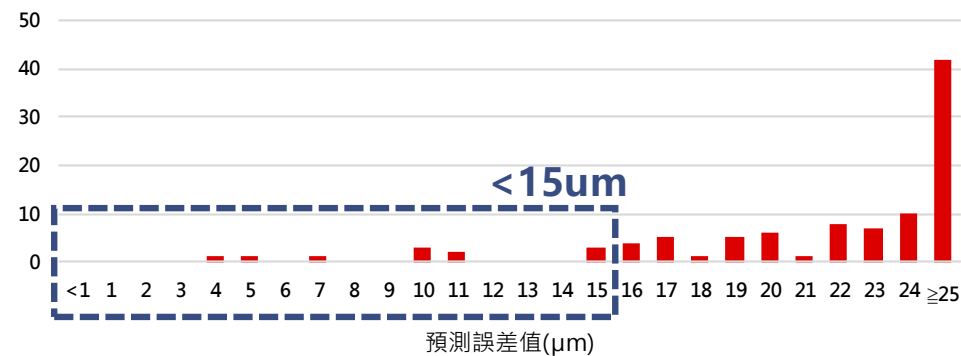
後段製程重工(%/year)



Coplanarity fail 導致的扣量



件數%



製程概述與平面度定義

晶圓表面製作元件結構

前段

- 氧化
- 光刻
- 蝕刻
- 薄膜沉積
- 離子植入

金屬互連與訊號通路的建立

中後段

- 沉積
- 蝕刻
- 填充
- 研磨
- 再分佈層
- 保護層覆蓋

電性連接與保護

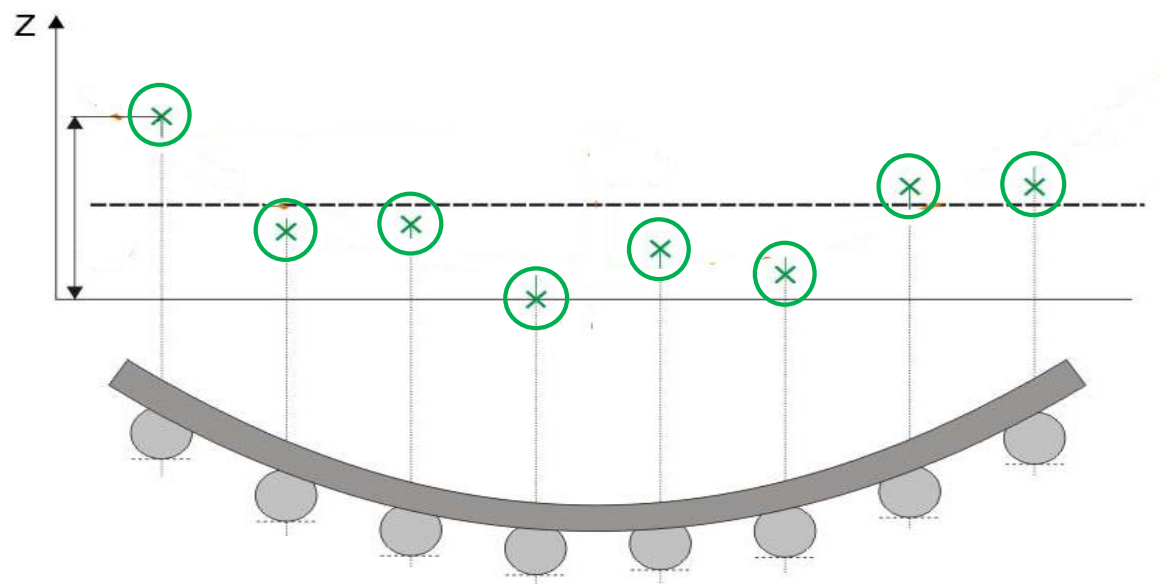
測試

製程概述與平面度定義

中後段

金屬互連與訊號通路的建立

- 沉積
- 蝕刻
- 填充
- 研磨
- 再分佈層
- 保護層覆蓋

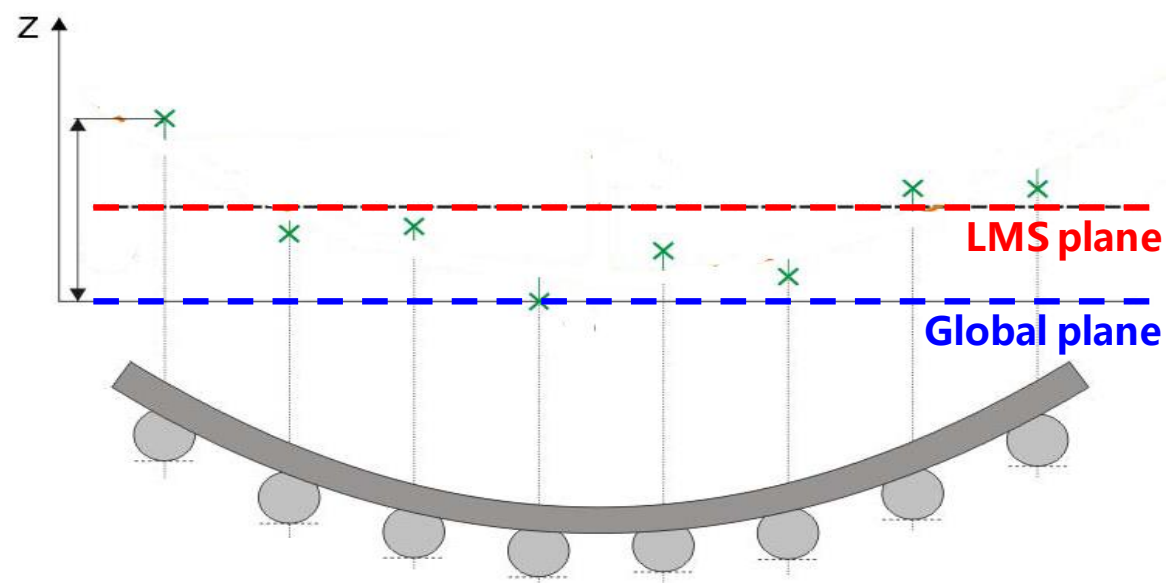


製程概述與平面度定義

中後段

金屬互連與訊號通路的建立

- 沉積
- 蝕刻
- 填充
- 研磨
- 再分佈層
- 保護層覆蓋

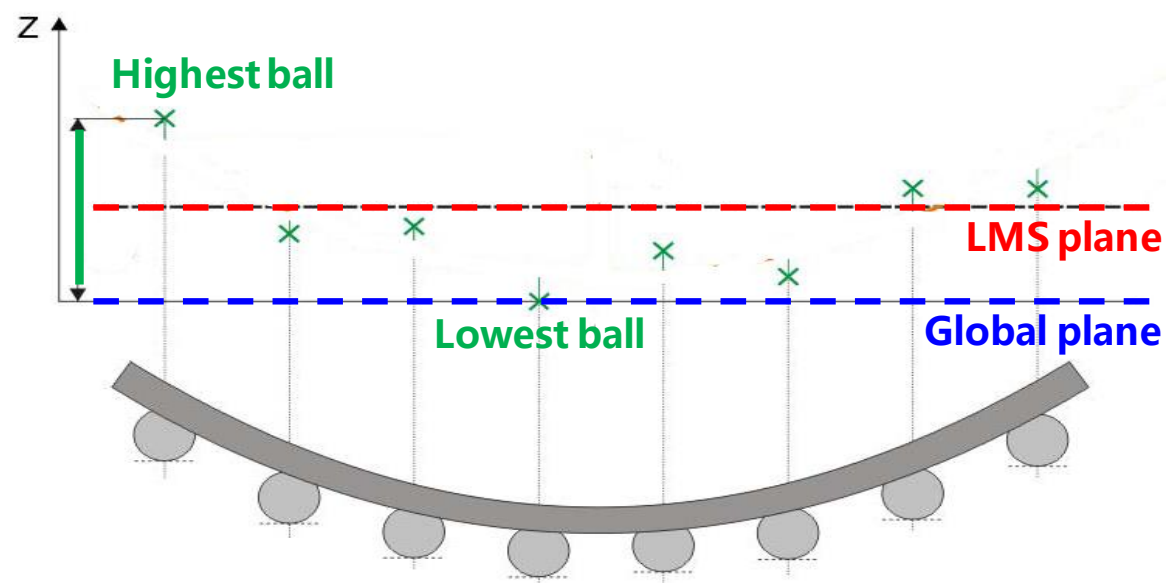


製程概述與平面度定義

中後段

金屬互連與訊號通路的建立

- 沉積
- 蝕刻
- 填充
- 研磨
- 再分佈層
- 保護層覆蓋



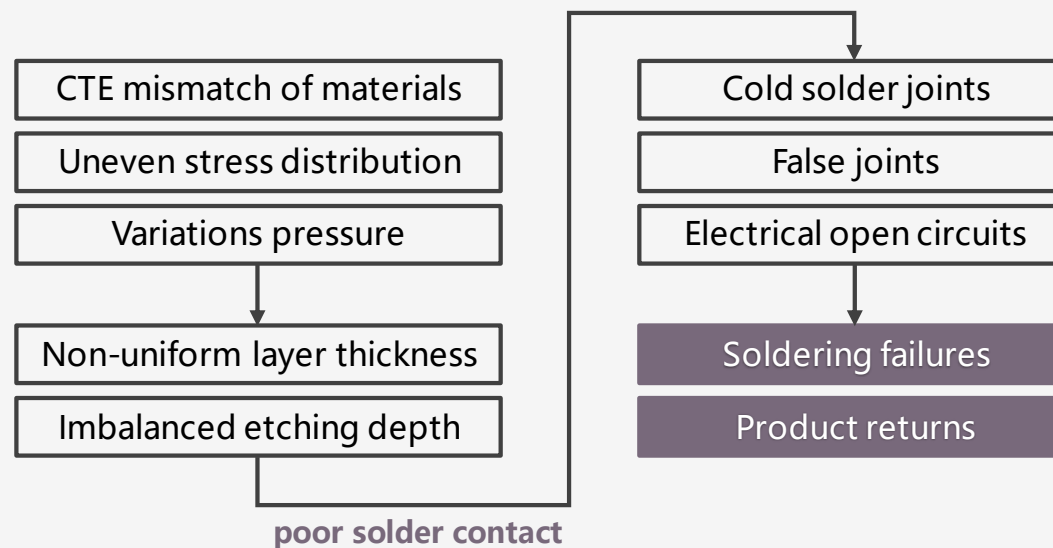
製程概述與平面度定義

中後段

金屬互連與訊號通路的建立

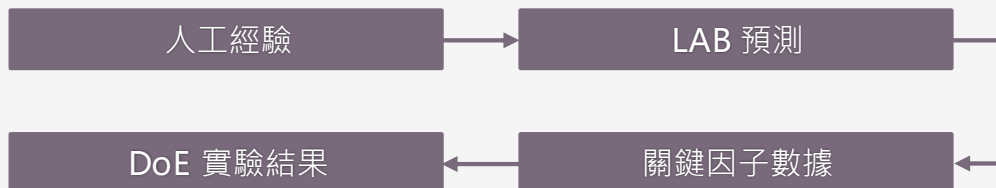
- 沉積
- 蝕刻
- 填充
- 研磨
- 再分佈層
- 保護層覆蓋

Cause



Prediction method

產品結構 / 材料特徵 / 參數設定



研究架構

單顆產品平面度
模型導入之研究

研讀並摘要
相關文獻及理論

NPI 平面度
預測現況剖析

特徵因子
訂定

結論
探討

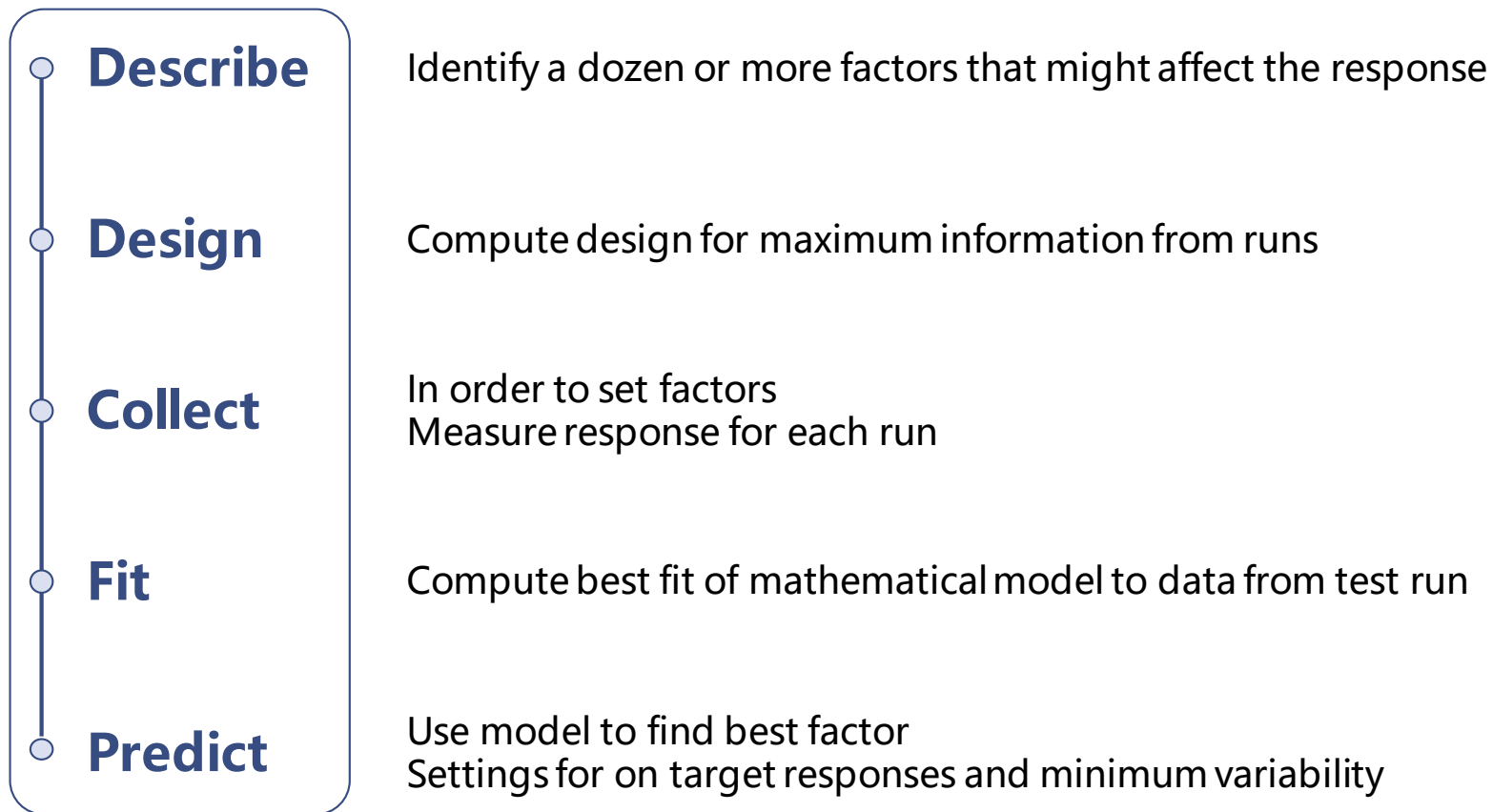
結果
分析與驗證

模型設計
執行

研究
方法

基礎理論

實驗設計方法的應用與挑戰 (Design of Experiments, DoE)



基礎理論

Prediction method

產品結構 / 材料特徵 / 參數設定



Improve in AI





文獻回顧

AI 與半導體



文獻回顧

機器學習 Machine Learning

逐步修正

模型殘留的誤差

建模能力強

非線性結構與變數交互關係

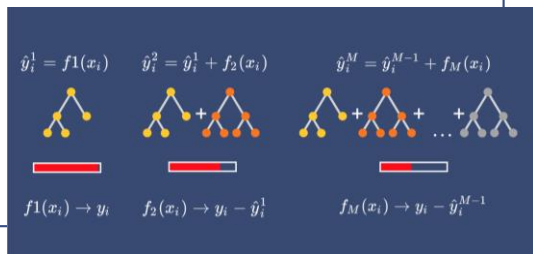
特徵選擇

自動學習變數重要性、模型解釋性

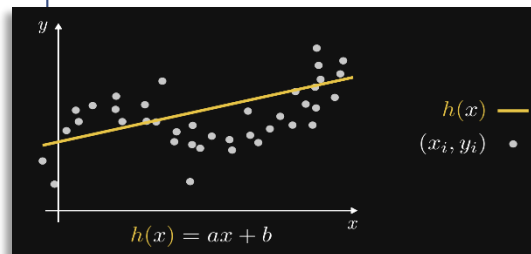
良好泛化能力

學習率與限制樹深度，降低過擬合

Gradient Boosting 梯度提升技術



Linear Regression 線性迴歸



自動變數

L1 正規化將部分不具貢獻參數壓縮

防止過擬合

正規化減少模型對訓練數據依賴

模型簡潔

去除冗餘資訊利於後續解釋

高維數據

適用製程變數多場景

穩健性

降低對個別樣本或噪聲敏感性

抗過擬合

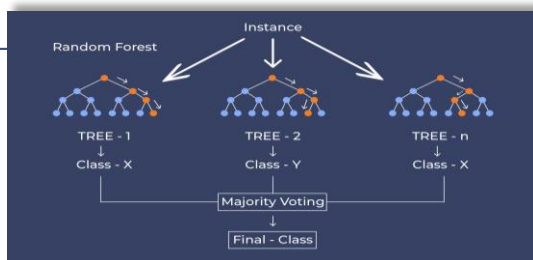
隨機選取樣本與特徵訓練子模型

大量特徵

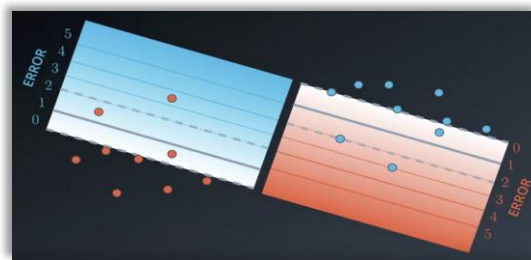
應對類別與數值特徵混合資料集

特徵重要性

了解模型依賴的關鍵參數



Random Forest 隨機森林



Support Vector Machine 支援向量機

分析邊界

最大化間隔，分類泛化能力

高維資料

適合小樣本但多特徵場景

核函數應用

將非線性問題轉線性可分

異常值韌性

極端值或雜訊影響有限

模型設計

目的

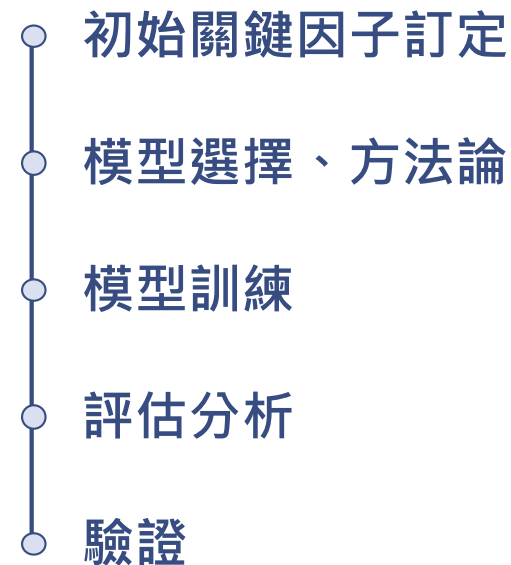
透過AI模型之導入，降低預測誤差值，提高準確率，為本研究目標，並縮減因子範圍，進而減少額外生成的成本及實驗結果時間，實現更高成本效益。

指標

符合預測規範準確率：從11% 提升至60%

間接

RA form 風險等級評估不再只有高風險和低風險兩個層級，根據不同風險情境更準確地識別和應對潛在風險

- 
- 初始關鍵因子訂定
 - 模型選擇、方法論
 - 模型訓練
 - 評估分析
 - 驗證

模型設計

初始關鍵因子訂定

模型選擇、方法論

模型訓練

評估分析

驗證

Device 資料集

產品結構大小分類蒐集

PKG size	Total
30x30mm~50x50mm	202
50x50mm~70x70mm	121
	323

Structure Factor	名詞解釋	Format
Die size(X, Y)	晶片水平和垂直長度	Numeric
Die thickness	晶片厚度	Class
HS thickness	散熱片厚度	Numeric
HS ring foot width	散熱片底座寬度	Class
PKG size(X, Y)	封裝水平和垂直長度	Numeric
PKG thickness	封裝厚度	Numeric
SBS thickness	基板厚度	Numeric
SBS layers	基板層數	Numeric
SBS core type	基板核心層類型	Class
SBS core thickness	基板核心層厚度	Numeric
SBS Cu thickness	基板銅箔厚度	Numeric
Hypotenuse	斜邊	Numeric
Area ratio (Die/SBS)	面積比	Numeric
Volume ratio (Die/SBS)	體積比	Numeric

Material Factor	名詞解釋	Format
Adhesive type	黏膠類型	Class
HS type	散熱片類型	Class
SBS core type	基板核心材料類型	Class
UF type	底部填充劑類型	Class

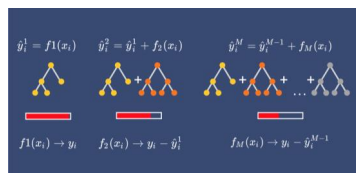
Process Parameters	名詞解釋	Format
BM reflow	焊膏回流焊接	Class
HS cure Temp.	散熱片固化溫度	Class
Snap cure(Y/N)	是否經熱壓合(快速固化)	Class
Snap cure Temp.	熱壓合(快速固化)溫度	Numeric
Snap cure Force	熱壓合(快速固化)力量	Numeric
UF cure Temp.	底部填充劑固化溫度	Class

模型設計

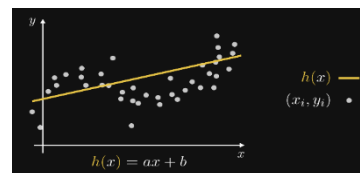
- 初始關鍵因子訂定
- 模型選擇、方法論
- 模型訓練
- 評估分析
- 驗證

演算法

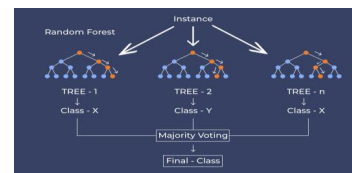
Gradient Boosting



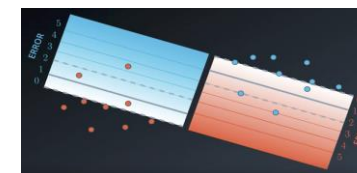
Linear Regression



Random Forest



SVM



方法論評估指標

平均絕對誤差 MAE

誤差的正負不會進行抵消，
能夠更準確地反映誤差的整體情況

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

均方根誤差 RMSE

識別出異常或離群值，
能夠顯示出是否有效控制大的預測誤差

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

決定係數 R²

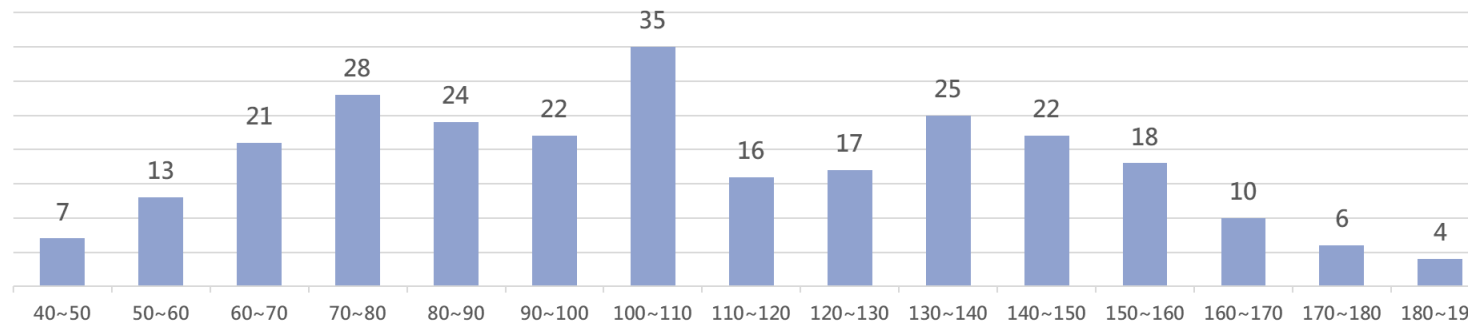
誤差的變異程度，
模型對資料擬合的好壞

$$R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$$

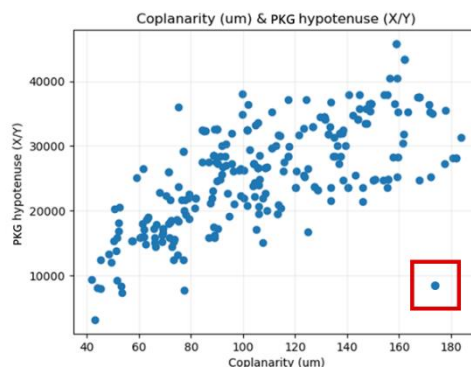
模型設計

模型訓練資料 : 323

PKG size 樣本分布



Model	MAE	R ²	RMSE	模型指標
Gradient Boosting	21.15	0.39	46.68	MAE <15(um)
Lasso Linear	25.34	0.05	36.81	
Random Forest	22.71	0.01	35.62	
Support Vector	27.45	0.11	33.84	



- 剔除離群值以降低誤差影響
- 移除系統中存在缺失值的因子筆數

資料筆數 → **293**

模型訓練資料 : 293

Model	MAE	R ²	RMSE	模型指標
Gradient Boosting	26.17	0.33	35.24	MAE <15(um)
Lasso Linear	22.38	0.45	31.39	
Random Forest	21.07	0.48	31.03	
Support Vector	33.44	0.08	41.21	

重要參數溫度做資料型態調整

- Snap cure Temp / Force 從 Numeric 調整成 Class
- 並非線性變化，而是具有特定的影響區間

Process Parameters	名詞解釋	Format
BM reflow	焊膏回流焊接	Class
HS cure Temp.	散熱片固化溫度	Class
Snap cure(Y/N)	是否經熱壓合(快速固化)	Class
Snap cure Temp.	熱壓合(快速固化)溫度	Class
Snap cure Force	熱壓合(快速固化)力量	Class
UF cure Temp.	底部填充劑固化溫度	Class

初始關鍵因子訂定

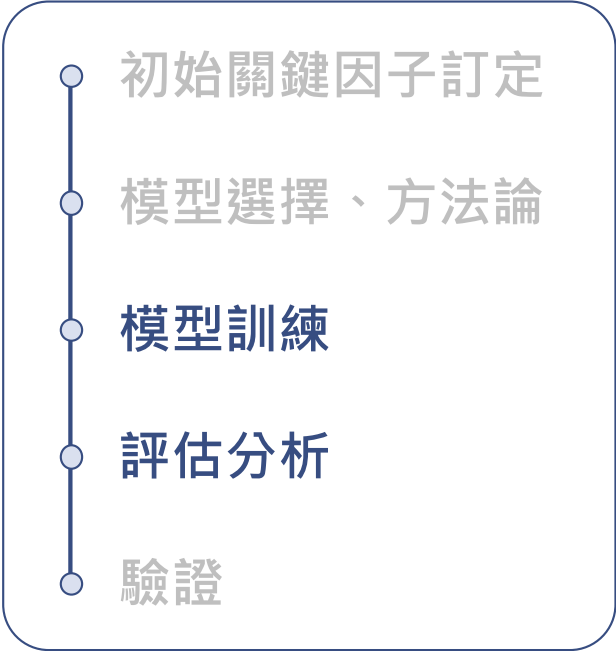
模型選擇、方法論

模型訓練

評估分析

驗證

模型訓練資料 : 263



Model	MAE	R ²	RMSE	模型指標
Gradient Boosting	22.04	0.44	25.55	MAE <15(um)
Lasso Linear	19.07	0.27	25.55	
Random Forest	16.77	0.44	22.39	
Support Vector	25.09	0.06	29.14	

↓ 取最佳演算法

Model	MAE	R ²	RMSE	模型驗證
Random Forest	16.77	0.44	22.39	MAE <15(um)

提高模型的泛化能力

- 數據呈現偏態分佈，使其偏離大多數樣本的典型水準
- 將數據的代表值從平均值改為中位數

Y	名詞解釋	Format
Coplanarity 中位數	平面度	Numeric

模型設計

模型訓練資料 : 210 / 模型驗證資料 : 53

Model	MAE	R ²	RMSE	驗證 MAE
Random Forest	14.79	0.77	20.14	15.69

Accuracy (MAE < 15um)

57% (30/53)

調整超參數

- 估算器數量 (n_estimators)、最大深度 (max_depth)、學習率 (learning_rate)
- 第一組參數能有效降低 MAE

n_estimators	max_depth	learning_rate	MAE	驗證 MAE
100	30	0.1	13.26	13.06
100	20	0.1	14.05	15.34
200	30	0.1	14.26	18.73

初始關鍵因子訂定
模型選擇、方法論
模型訓練
評估分析
驗證

實驗結果

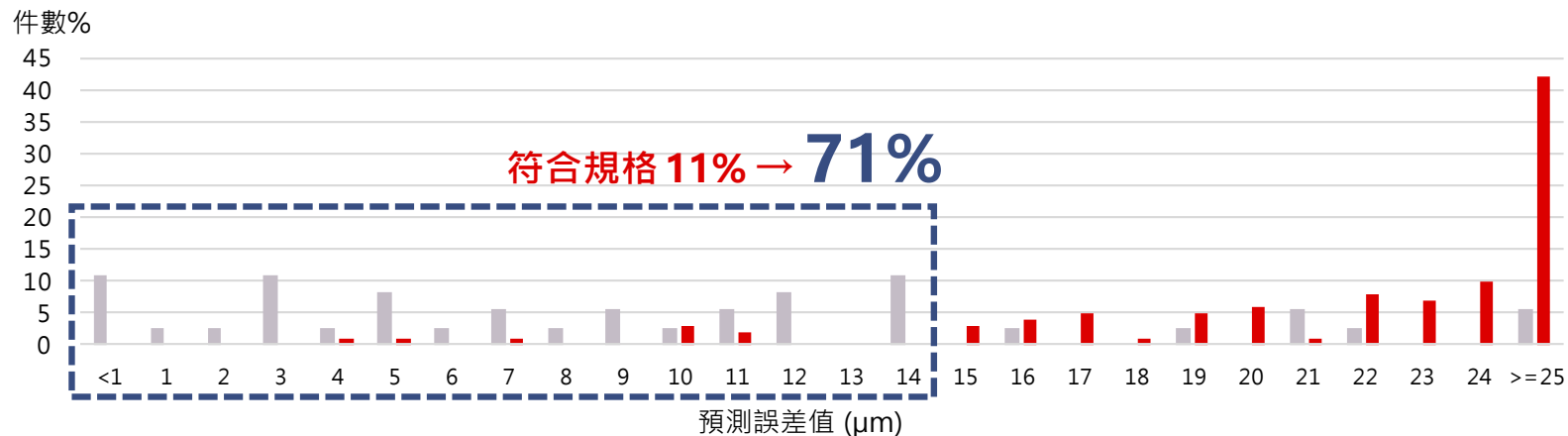
模型訓練資料 : 210 / 模型驗證資料 : 53

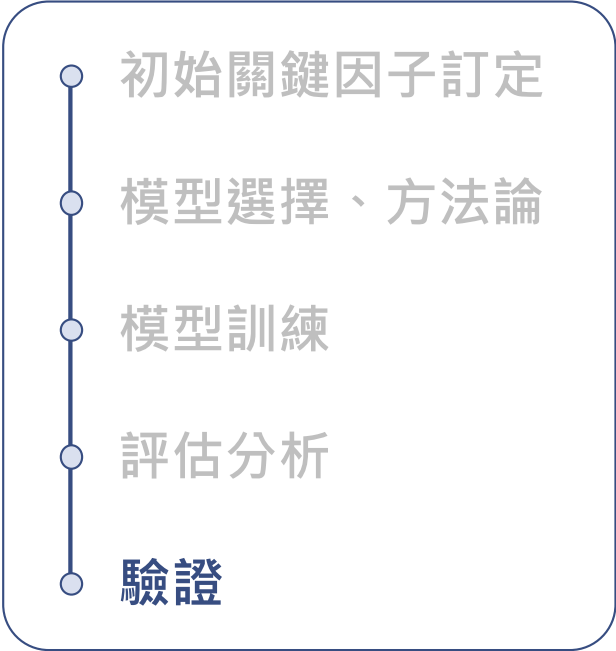
Model	MAE	50% Error	80% Error	90% Error	Max Error
Random Forest	13.06	9.1	14.94	19.04	27.77

Accuracy (MAE < 15 μ m)

71% (38/53)

- 初始關鍵因子訂定
- 模型選擇、方法論
- 模型訓練
- 評估分析
- 驗證





模型導入

不同產品結構進行測試

Device	PKG size (mm)	Die Size (um)	HS type	PKG thickness
m	65 x 65	26488 x 22427	CAVITY	3896
n	37.5 x 37.5	14857 x 18504	HAT	2660
o	31 x 31	12000 x 10425	RING	3140

↓ 結果

Device	MAE	預測值 (um)	實際值 (um)	預測誤差 (um)
m	13.11	24	37	13
n	12.27	43	47	4
o	12.68	97	106	3

客戶要求最大允收誤差 15μm

- 預測誤差控制在 2~5μm 之間，具備極高的穩定性與泛化能力
- 機器學習模型大幅提升預測精度

結論與未來展望

本研究聚焦於半導體製程中關鍵品質指標，平面度的預測問題，並透過機器學習技術的導入，探討其對預測準確性與製程優化的提升潛力。隨著製程結構的高度複雜化與產品設計的不斷創新，傳統依賴人工經驗與實驗設計 (DoE) 方法已難以有效掌握多維參數間的非線性關聯性。儘管 DoE 能提供初步變因篩選與參數設計的指引，但在時間成本、實驗複雜度及彈性適應性方面仍具明顯侷限。因此，本研究將機器學習視為一種更具擴展性與效率的預測手段，進行模型導入與比較分析。

模型特性與表現

- 傳統線性迴歸模型雖具解釋力，但對於變數之間的非線性關係掌握不足，整體預測效能不如集成式學習模型，難以應用於高複雜度製程條件下的平面度預測。
- 在本研究比較的四種模型中，以**Random Forest**表現最佳，能在有限資料下維持穩定的預測精度，並準確識別影響平面度的主要參數。
- 透過因子與目標的調整，使模型提高其泛化能力，本研究中顯示準確率有效提高，並在實際導入結果驗證其可行性。

研究成果

- 集成式演算法在預測準確率、殘差控制、模型穩定性表現均優於其他方法。
- 特徵重要性分析結果顯示，材料性質、製程條件，如溫度、壓力與結構參數是關鍵影響因子，有助於後續製程調整與新產品導入的設計參考。
- 證實 **Gradient Boosting** 與 **Random Forest** 具優異預測能力與穩定性，並能在中等樣本規模下取得良好表現，顯著優於傳統 **DoE** 方法，可提供未來做預測模型的適用性。



結論與未來展望

未來研究方向

- 建模品質受限於輸入數據的完整性與一致性，後續強化資料前處理作業，如異常值排除、缺值補齊，以確保模型穩定性。
- 整合製程知識圖譜，將製程知識與機器學習模型結合，提高模型可解釋性與決策透明度。
- 評估模型實際部署效益，針對模型在實際產線導入後的預測效能、運行成本與維護需求進行迭代優化。
- 引入不同製程站點資料以擴大模型適用範圍，**AI**在不同預測因子下提升準確率的可行性。

Thank You

