

**TabPFN:** 一個能在一秒內解決小型表格分類問題的 **Transformer**

## 總結重點

### 1. TabPFN 概述:

- 一個預訓練的 **Transformer** 模型，能夠在不到一秒內解決小型表格分類問題
- 無需超參數調整，性能與最先進 **AutoML** 系統相當
- 使用上下文學習(ICL)，通過單次前向傳遞產生預測

### 2. 技術創新:

- 基於先驗-數據擬合網絡(PFN)架構，近似貝葉斯推理
- 設計了新穎的先驗，結合結構因果模型(SCM)和貝葉斯神經網絡(BNN)
- 先驗納入簡單性和因果關係的思想，偏好參數較少的簡單結構

### 3. 性能優勢:

- 在 **OpenML-CC18** 基準的 **18** 個純數值數據集上明顯優於增強樹模型
- 與需要一小時訓練的複雜 **AutoML** 系統性能相當，但速度提高 **230-5700** 倍
- 特別適用於無缺失值的純數值特徵數據集

### 4. 實用特點:

- 單次前向傳遞完成訓練和預測，**CPU** 上 **1.3** 秒內完成，**GPU** 上僅需 **0.05** 秒
- 可與其他方法(如 **AutoGluon**)集成，進一步提高性能
- 能夠泛化到訓練中未見過的更大數據集尺寸

### 5. 局限性與未來方向:

- 僅適用於小型數據集( $\leq 1000$  訓練樣本， $\leq 100$  特徵， $\leq 10$  類)

- 對分類特徵和缺失值處理能力有限
- 未來工作包括擴展到更大數據集、改進特徵處理、整合到現有 AutoML 框架等

**TabPFN** 代表了表格數據分類方法的根本性變革，將傳統的從頭訓練模型方法替換為預訓練模型的單次前向傳遞，大大提高了效率同時保持了高性能。

## 摘要 (Abstract) 翻譯與重點歸納

論文介紹了 TabPFN，這是一個經過訓練的 Transformer，能夠在不到一秒的時間內完成小型表格數據集的監督分類任務，無需超參數調整，並且性能與最先進的分類方法相當。

TabPFN 執行上下文學習(In-context Learning, ICL)，它學習使用輸入中給定的標記樣本序列( $x, f(x)$ )進行預測，無需進一步更新參數。TabPFN 完全包含在網絡權重中，將訓練和測試樣本作為集合值輸入，並在單次前向傳遞中為整個測試集產生預測。

TabPFN 是一個先驗-數據擬合網絡(Prior-Data Fitted Network, PFN)，僅需離線訓練一次，以近似在從先驗中抽取的合成數據集上的貝葉斯推理。這個先驗包含因果推理的思想：它蘊含了大量的結構因果模型，並偏好簡單的結構。

在 OpenML-CC18 套件中包含的 18 個數據集上（這些數據集最多包含 1000 個訓練數據點，最多 100 個純數值特徵無缺失值，以及最多 10 個類別），論文顯示該方法明顯優於增強樹模型，並且性能與複雜的最先進 AutoML 系統相當，但速度提高了 230 倍。使用 GPU 時，速度可提高到 5700 倍。這些結果在 OpenML 的額外 67 個小型數值數據集上也得到了驗證。

## 1. 介紹 (Introduction) 翻譯與重點歸納

雖然表格數據是實際機器學習應用中最常見的數據類型，但長期以來被深度學習研究所忽視。在表格數據分類問題中，梯度提升決策樹(GBDT)仍然占主導地位，主要是因為它們訓練時間短且穩健性強。

作者提出了一種表格分類方法的根本性變革。不是從頭開始為新數據集訓練模型，而是使用一個大型的、預先訓練好的 Transformer 進行單次前向傳遞，該 Transformer 已經被訓練來解決從表格數據先驗中人工生成的分類任務。

該方法基於先驗-數據擬合網絡(PFN)，這種網絡學習訓練和預測算法本身。PFN 直接近似給定任何可以從中抽樣的先驗的貝葉斯推理和後驗預測分佈(PPD)。傳統神經網絡和 GBDT 的歸納偏差依賴於它們實現的高效性（如 L2 正則化、dropout 或有限樹深度），而在 PFN 中，只需設計一個編碼所需先驗的數據生成算法即可，這從根本上改變了設計學習算法的方式。

作者設計了一個基於貝葉斯神經網絡(BNN)和結構因果模型(SCM)的先驗，來模擬複雜的特徵依賴關係和可能的表格數據底層因果機制。這個先驗也採納了奧卡姆剃刀的思想：更簡單的 SCM 和 BNN（參數更少）具有更高的似然性。通過參數分佈定義先驗，例如，數據生成 SCM 中節點平均數量的對數均勻分佈。結果的 PPD 隱含地模擬了所有可能的數據生成機制上的不確定性，根據給定數據的似然性和先驗概率對它們進行加權。因此，PPD 對應於無限大的數據生成機制的集合，即 SCM 和 BNN 的實例。作者學習在單次前向傳遞中近似這個複雜的 PPD，不需要交叉驗證或模型選擇。

論文的主要貢獻是引入 TabPFN，一個為上述新穎先驗（在第 4 節中詳細描述）在單次前向傳遞中近似概率推理的預訓練 Transformer，從而學會在不到一秒的時間內解決新的小型表格分類任務（ $\leq 1000$  訓練樣本， $\leq 100$  純數值特徵無缺失值， $\leq 10$  類別），同時達到最先進的性能。

## 2. 先驗-數據擬合網絡(PFN)的背景 翻譯與重點歸納

首先，本節概述了 PFN 的工作原理，詳細內容可參考 Müller 等人(2022)的論文。

### 監督學習的後驗預測分佈

在監督學習的貝葉斯框架中，先驗定義了一組關於輸入  $x$  到輸出標籤  $y$  關係的假設  $\Phi$ 。每個假設  $\phi \in \Phi$  可以看作是一個生成數據分佈的機制。例如，給定一個基於結構因果模型的先驗， $\Phi$  是結構因果模型的空間，假設  $\phi$  是一個特定的 SCM，而數據集包括通過這個 SCM 生成的樣本。

實際上，數據集包括具有觀察標籤的訓練數據和標籤缺失或保留以評估預測性能的測試數據。測試樣本  $x_{\text{test}}$  的後驗預測分佈(PPD)指定其標籤  $p(\cdot | x_{\text{test}}, D_{\text{train}})$  的分佈，這取決於訓練樣本集  $D_{\text{train}} := \{(x_1, y_1), \dots, (x_n, y_n)\}$ 。PPD 可以通過在假設空間  $\Phi$  上積分獲得，其中假設  $\phi \in \Phi$  的權重由其先驗概率  $p(\phi)$  和給定  $\phi$  的數據  $D$  的似然  $p(D | \phi)$  決定：

$$p(y|x, D) \propto \int \Phi p(y|x, \phi) p(D|\phi) p(\phi) d\phi$$

## 合成先驗擬合

先驗擬合是訓練 PFN 以近似 PPD 並進行貝葉斯預測。作者通過形式為  $p(D) = E_{\phi \sim p(\phi)}[p(D|\phi)]$  的先驗抽樣方案實現它，該方案首先以  $\phi \sim p(\phi)$  抽樣假設（生成機制），然後以  $D \sim p(D|\phi)$  抽樣合成數據集。作者反復抽樣這樣的合成數據集  $D := (x_i, y_i)_{i \in \{1, \dots, n\}}$ ，並優化 PFN 的參數  $\theta$ ，使其根據數據集的其餘部分  $D_{\text{train}} = D \setminus \text{I am running a few minutes late; my previous meeting is running over.}$   $D_{\text{test}}$  對  $D_{\text{test}} \subset D$  進行預測。因此，PFN 訓練的損失是合成數據集中保留樣本的交叉熵。對於單個測試點  $\{(x_{\text{test}}, y_{\text{test}})\} = D_{\text{test}}$ ，訓練損失可以寫為：

$$L_{\text{PFN}} = E_{\{(x_{\text{test}}, y_{\text{test}})\} \cup D_{\text{train}} \sim p(D)}[-\log q_{\theta}(y_{\text{test}}|x_{\text{test}}, D_{\text{train}})]$$

如 Müller 等人(2022)所示，最小化這個損失近似真實的貝葉斯後驗預測分佈。

## 實際世界推理

在推理過程中，將訓練好的模型應用於未見過的真實世界數據集。對於具有訓練樣本  $D_{\text{train}}$  和測試特徵  $x_{\text{test}}$  的新數據集，將  $\langle D_{\text{train}}, x_{\text{test}} \rangle$  作為輸入饋入到上述訓練的模型中，在單次前向傳遞中產生 PPD  $q_{\theta}(y|x_{\text{test}}, D_{\text{train}})$ 。然後使用 PPD 類別概率作為我們實際任務的預測。因此，PFN 在一步中執行訓練和預測（類似於高斯過程的預測），並且不在推理時對所見數據使用基於梯度的學習。這種學習能力可以稱為上下文內學習(ICL)，當為語言任務訓練大型神經序列模型時，這種能力被觀察到會驚人地出現。

## 架構

PFN 依賴於一個 Transformer，它將每個特徵向量和標籤編碼為一個 token，允許 token 表示相互關注。它們接受變長訓練集  $D_{\text{train}}$  的特徵和標籤向量（作為集合值輸入以利用排列不變性）以及變長查詢集的特徵向量  $x_{\text{test}} = \{x(\text{test}, 1), \dots, x(\text{test}, m)\}$ ，並返回每個查詢的 PPD 估計。

## 表格數據

Müller 等人(2022)在其他實驗中展示了 PFN 在有 30 個訓練樣本的微小、平衡二元分類表格數據集上的 PPD 近似。在這項工作中，作者大幅提高了性能，並擴展到最多 1000 個數據點和最多 10 個類別的小型數據集，包括不平衡情況。

## 3. TabPFN: 適用於表格數據的新先驗 PFN 翻譯與重點歸納

TabPFN 是一個先驗-數據擬合網絡(PFN)，它根據作者在第 4 節中介紹的表格數據新先驗進行擬合。作者對原始 PFN 架構(Müller 等人, 2022)進行了兩處修改：

1. 對注意力掩碼做了輕微修改，縮短了推理時間。
2. 通過零填充，使模型能夠處理不同特徵數量的數據集。

在先驗擬合階段，作者使用第 4 節中描述的先驗樣本一次性訓練 TabPFN。具體來說，作者在 8 個 GPU(Nvidia RTX 2080 Ti)的一台機器上，訓練了一個 12 層 Transformer，每批 512 個合成生成的數據集，共 18000 批次，總共耗時 20 小時。這產生了一個用於所有評估的單一網絡。雖然這個訓練步驟相當昂貴，但它是作為算法開發的一部分，預先離線執行的，而且只需為 TabPFN 執行一次。本文中所有實驗都使用相同的 TabPFN 模型。

在推理過程中，TabPFN 近似了我們數據集先驗的 PPD，即它近似了我們 SCM 和 BNN 空間（見第 4 節）的邊際預測，包括對數據的簡單和因果解釋的偏好。在實驗中，作者展示了 TabPFN 單次前向傳遞的預測，以及通過數據集的 32 次前向傳遞集成的預測，這些數據集通過幂變換（以 0.5 的概率應用）和旋轉特徵列和類標籤的索引進行修改。

## 4. 表格數據的先驗 翻譯與重點歸納

方法的性能很大程度上取決於合適先驗的設定，因為 PFN 近似這個先驗的 PPD。

### 4.1 基本概率模型

傳統模型擬合通常需要找到合適的超參數，如嵌入大小、層數和激活函數等。通常需要進行資源密集型搜索來找到合適的超參數，但結果只是超參數選擇的點估計。在多個架構和超參數設置上進行集成可以粗略近似這些超參數的分佈，並已被證明能提高性能，但成本與所考慮的選擇數量成線性增長。

相比之下，PFN 允許在先驗的超參數上完全貝葉斯化。通過在先驗中定義超參數空間（如 BNN 架構）的概率分佈，TabPFN 近似的 PPD 共同整合了這個空間和相應的模型權重。作者將這種方法擴展到不僅是超參數的混合，還包括不同的先驗：混合 BNN 和 SCM 先驗，每個先驗又包含架構和超參數的混合。

### 4.2 簡單性

作者的先驗基於簡單性原則，如奧卡姆剃刀或速度先驗。在考慮競爭假設時，

應優先選擇更簡單的假設。認知科學研究也發現人類思維中存在這種對簡單解釋的偏好。不過，任何簡單性概念都取決於選擇特定標準來定義簡單性。以下介紹的基於 SCM 和 BNN 的先驗中，作者將簡單性實現為具有較少節點和參數的圖。

### 4.3 SCM 先驗

因果知識可以促進各種機器學習任務，包括半監督學習、遷移學習和分佈外泛化。表格數據經常表現出列之間的因果關係，並且因果機制被證明是人類推理中的強先驗。因此，作者基於模型因果關係的 SCM 構建了 TabPFN 先驗。

SCM 由一系列結構賦值（稱為機制） $Z := (\{z_1, \dots, z_k\})$  組成： $z_i = f_i(z_{\text{PAG}(i)}, \epsilon_i)$ ，其中  $\text{PAG}(i)$  是基礎有向無環圖  $G$  中節點  $i$  的父節點集（直接原因）， $f_i$  是（可能非線性）確定性函數， $\epsilon_i$  是噪聲變量。 $G$  中的因果關係由指向從因到果的有向邊表示，每個機制  $z_i$  被分配給  $G$  中的一個節點。

### 基於因果推理的預測

以往的工作通過使用因果推斷（一種通過干預和觀察數據識別系統組件之間因果關係的方法）將因果推理應用於未見數據的預測。預測的因果表示然後用於對新樣本進行觀察預測或提供可解釋性。大多數現有工作集中在確定單一因果圖用於下游預測，這可能存在問題，因為在沒有干預數據的情況下，大多數 SCM 類型是不可識別的，並且由於 DAG 空間的組合性質，兼容 DAG 的數量爆炸性增長。最近，Ke 等人和 Lorch 等人使用 transformer 從觀察和干預數據中近似因果圖。作者跳過任何顯式的圖表示，直接近似 PPD。因此，作者不執行因果推斷，而是直接解決下游預測任務。這種假設 SCM 類似過程生成數據的隱含假設可以在 Pearl 的"因果梯階"中解釋，這是一個抽象的推斷類別層次，每個更高的階梯代表一個更複雜的推斷概念。最低階梯是關聯，包括大多數機器學習。第二階梯考慮預測干預的效果，即當我們直接影響特徵時會發生什麼。作者的工作可被視為"階梯 1.5"，類似於 Kyono 等人的工作：作者不執行因果推理，而是基於觀察數據進行關聯預測，假設 SCM 很好地模型化常見數據集。

### 基於因果模型定義先驗

為了創建基於 SCM 的 PFN 先驗，作者需要定義一個創建監督學習任務（即數據集）的抽樣過程。每個數據集基於一個隨機抽樣的 SCM（包括 DAG 結構和確定性函數  $f_i$ ）。給定一個 SCM，作者在因果圖  $G$  中抽樣一組節點  $z_X$ ，每個節

點對應合成數據集中的一個特徵，以及一個節點  $z_y$ 。這些是觀察節點： $z_x$  的值將包含在特徵集中，而  $z_y$  的值將作為目標。對於每個 SCM 和節點列表  $z_x$  和  $z_y$ ，通過在 SCM 中抽樣所有噪聲變量  $n$  次，通過圖傳播這些噪聲，並檢索  $n$  個樣本在節點  $z_x$  和  $z_y$  的值，生成  $n$  個樣本。通過生成 DAG 結構，特徵通過前向和後向因果關係條件依賴，即目標可能是特徵的原因或結果。

在這項工作中，作者實例化了一個大型 DAG 子族和確定性函數  $f_i$  來構建 SCM。由於高效抽樣是唯一要求，因此實例化的子族非常通用，包括多種激活函數和噪聲分佈。

#### 4.4 BNN 先驗

作者還考慮了類似於 Müller 等人(2022)引入的 BNN 先驗，並將其與上述 SCM 先驗混合，在 PFN 訓練期間以相等概率從兩者中隨機抽樣數據集。要從 BNN 先驗抽樣數據集，作者首先抽樣 NN 架構及其權重。然後，對於待生成數據集中的每個數據點，抽樣輸入  $x$ ，將其通過具有抽樣噪聲變量的 BNN，並使用輸出  $y$  作為目標。附錄 B.4 中對這個 BNN 先驗相對於最終混合 SCM 和 BNN 先驗的實驗消融顯示了新穎 SCM 先驗的強大性。

#### 4.5 多類預測

到目前為止，描述的先驗返回標量標籤。為了為不平衡的多類數據集生成合成分類標籤，需要將標量標籤  $\hat{y}$  轉換為離散類標籤  $y$ ：

1. 抽樣類別數量  $N_c \sim p(N_c)$ ，其中  $p(N_c)$  是整數上的分佈。
2. 從連續目標  $\hat{y}$  的集合中隨機抽樣  $N_c - 1$  個類邊界  $B_i$ 。
3. 將每個標量標籤  $\hat{y}_i$  映射到包含它的唯一區間的索引： $y_i \leftarrow \sum_j [B_j < \hat{y}_i]$ ，其中  $[\cdot]$  是指示函數。

例如，使用  $N_c = 3$  類，邊界  $B_c = \{-0.1, 0.5\}$  會定義三個區間  $\{(-\infty, -0.1], (-0.1, 0.5], (0.5, \infty)\}$ 。任何  $\hat{y}_i$  如果小於  $-0.1$  將被映射到標籤 0，如果在  $(-0.1, 0.5]$  內則映射到 1，否則映射到 2。最後，作者打亂類標籤的順序，即移除類標籤相對於範圍的排序。

### 5. 實驗 翻譯與重點歸納

#### 5.1 玩具問題評估

作者首先在圖 4 中將 TabPFN 與標準分類器（無超參數調整）進行定性比較。第一行顯示有噪聲的月球數據集。TabPFN 準確建模了樣本之間的決策邊界；此外，類似於高斯過程，遠離觀察樣本的點的不確定性很大。第二行顯示有噪聲的圓圈數據集：TabPFN 準確建模了圓的形狀，在樣本混合區域之外的任何地方具有高確信度。第三行顯示了來自 iris 數據集的兩個類別和特徵，而第四行顯示了來自 wine 數據集的兩個類別和特徵（兩者都在 scikit-learn 中）；在這兩種情況下，TabPFN 做出直觀、校準良好的預測。

## 5.2 表格機器學習任務評估

現在，作者轉向對實際分類任務中方法的實證分析，將 TabPFN 與表格分類的最先進 ML 和 AutoML 方法進行比較。

### 數據集

作為測試數據集，作者使用了開源 OpenML-CC18 基準套件中所有包含最多 2000 個樣本（訓練拆分 1000 個）、100 個特徵和 10 個類別的數據集。結果集包含 30 個數據集。作者將這些數據集分為 18 個僅包含數值特徵且無缺失值的數據集，以及 12 個包含分類特徵和/或缺失值的其他數據集。在主論文的圖 5 和表 1 中，作者將分析限制在沒有缺失值的純數值數據集上，這是 TabPFN 先驗開發的重點。作者關注小型數據集是因為：(1)小型數據集在實際應用中經常遇到；(2)現有深度學習方法在這一領域最為有限；(3)對於更大的數據集，TabPFN 的訓練和評估成本會顯著增加。

### 基準

作者將 TabPFN 與五種標準 ML 方法和兩種最先進的表格數據 AutoML 系統進行比較。作為 ML 模型，作者考慮了兩個簡單而快速的基準，K 最近鄰(KNN)和邏輯回歸(LogReg)。此外，作者考慮了三種流行的基於樹的提升方法，XGBoost、LightGBM 和 CatBoost。對於每個 ML 模型，作者使用 5 折交叉驗證來評估隨機抽取的超參數配置，直到給定預算用盡或評估了 10000 個配置。然後選擇表現最佳的超參數配置（最大 ROC AUC OVO）並在整個訓練集上重新擬合。必要時，作者用均值填充缺失值，對分類輸入進行 one-hot 或序數編碼，標準化特徵並傳遞分類特徵指標。作為更複雜但強大的基準，作者選擇了兩個最先進的 AutoML 系統：AutoGluon，它將包括神經網絡和基於樹的模型在內的 ML 模型組合成堆疊集成；以及 Auto-sklearn 2.0，它使用貝葉斯優化並將評估的模型組合成加權集成。



## 評估協議

對於每個數據集和方法，作者評估了 5 次重複，每次使用不同的隨機種子和訓練-測試分割（50%訓練和 50%測試樣本；給定種子，所有方法使用相同的分割）。為了在數據集間彙總結果，作者報告了 ROC AUC（多類分類為 one-vs-one，OVO）平均值、排名和勝利次數，包括 95%置信區間，並與 {30,60,300,900,3600}秒預算的基準性能進行比較。作者的 TabPFN 使用 32 個數據排列進行集成，也評估了不使用排列的 TabPFN（在表 1 中標記為 "TabPFNN.e."）。

## 結果

圖 5 和表 1 顯示了在 18 個純數值數據集上的結果。圖 5 顯示 TabPFN 在準確性和訓練速度之間實現了比所有其他方法都大幅度改善的權衡：它在一個 GPU 上不到一秒內做出預測，其性能與最佳競爭者（AutoML 系統）訓練一小時後相當，並且優於調整過的 GBDT 方法。不出所料，簡單的基準（LogReg，KNN）已經能在小預算內產生結果，但總體表現最差。GBDTs（XGBoost，CatBoost，LightGBM）表現更好，但仍被 TabPFN 和最先進的 AutoML 系統（Auto-sklearn 2.0，Autogluon）超越。

TabPFN 比性能相當的方法快得多。對於訓練和預測（以及調整，如果適用）的總時間，TabPFNN.e.在 CPU 上平均需要 1.30 秒，在 GPU 上需要 0.05 秒，在 5 分鐘內表現與最強基準相當；因此，在 CPU 上加速 230 倍，使用 GPU 加速 5700 倍。

作者強調，這些結果是跨數據集的總體結果，包括 TabPFN 在內的任何分類方法都不會在所有單個數據集上表現最佳。確實，存在 TabPFN 甚至被默認基準超越的數據集。一般來說，當存在分類特徵或缺失值時，TabPFN 表現較弱。

## OpenML-AutoML 基準

此外，作者在 OpenML-AutoML 基準的小型（≤1000 訓練樣本，100 特徵，10 類別）數據集上評估了 TabPFN，這些數據集有外部驗證的性能數字。使用 OpenML-AutoML 基準提供的設置，即：度量標準、官方評估腳本和針對更廣泛 AutoML 基準預發布的評估。在單個 CPU 上平均每個數據集僅使用 4.4 秒，相比基準的 60 分鐘，TabPFN 在平均交叉熵、準確度和 OpenML 指標方面優於所有基準。

## TabPFN 預測的深入分析

作者以多種方式評估了模型預測，以提供額外見解，如方法的歸納偏差。作者確認 TabPFN 學會做出偏向簡單因果解釋的預測，而 GBDT 方法不共享這種歸納偏差。作者還評估了方法對特徵旋轉的不變性和對無信息特徵的穩健性。圖 6 顯示，與基準相比，當數據集不包含分類特徵或缺失值時，TabPFN 尤其強大。

## 集成

作者觀察到 TabPFN 在不同於基準的數據集上表現最佳，即 TabPFN 與強基準之間的每個數據集標準化 ROC AUC 分數相關性低於強基準之間的相關性。這可能是由於其方法的新穎歸納偏差，導致不同的預測。當考慮的方法產生較少相關錯誤時，集成預測更有效，這鼓勵使用更多樣化的策略，使 TabPFN 成為與基準方法集成的理想候選者。此外，與基準的長運行時間相比，TabPFN 的評估速度極快，預測幾乎是免費的。為了展示其集成潛力，作者在表 1 中包含了一個"TabPFN + AutoGluon"項，通過平均 TabPFN 和 AutoGluon 的預測生成；這明顯優於所有其他方法。

## 模型泛化

PFN 架構接受任何長度的數據集作為輸入。然而，在合成先驗擬合階段訓練模型時，作者將合成數據限制在最大尺寸 1024，因為先驗擬合隨著數據集變大而變得更昂貴。然而，作者好奇：TabPFN 是否會泛化到訓練期間從未見過的更大訓練集尺寸？為了測試這一點，作者使用了 OpenML-AutoML 基準中的 18 個數據集集合，從中各選取 10000 個樣本。作者評估了 TabPFN 在最多 5000 個訓練樣本上的表現，使用 5 個隨機數據分割和 5000 個測試樣本。令人驚訝的是，模型能夠泛化超出訓練中見到的樣本尺寸，如附錄 B.1 中的圖 10 所示。

## 6. 結論與未來工作 (續) 翻譯與重點歸納

作者展示了一個單一的 Transformer，TabPFN，可以被訓練來完成一個完整 AutoML 框架對表格數據的工作，並且能在 0.4 秒內產生預測，這些預測與最佳可用 AutoML 框架在 5 到 60 分鐘內達到的性能相當。這大大降低了 AutoML 的計算成本，使其變得經濟實惠、環保。

TabPFN 仍然存在一些重要限制：底層 Transformer 架構只能擴展到小型數據集(如附錄 A 所述)；作者的評估集中在只有最多 1000 個訓練樣本、100 個純數值

特徵(無缺失值)和 10 個類別的分類數據集上，這促使未來工作(1)擴展到更大的數據集。作者對 TabPFN 歸納偏差的深入分析(見附錄 B.3)指向(2)改進處理分類特徵、(3)處理缺失值和(4)對不重要特徵的穩健性的擴展。

此外，作者的工作激發了一系列令人興奮的後續工作，包括(5)將 TabPFN 整合到現有 AutoML 框架中；(6)在給予更多時間的情況下繼續通過集成進行改進；(7)數據集相關的先驗選擇；(8)推廣到非表格數據和(9)回歸任務，以及關於可信 AI 維度(例如，(10)分佈外穩健性；(11)算法公平性，(12)對抗樣本穩健性，和(13)可解釋性)的 TabPFN 研究。

TabPFN 幾乎即時的最先進預測也可能會產生(14)新穎的探索性數據分析方法，(15)新穎的特徵工程方法和(16)新穎的主動學習方法。最後，作者在因果推理方面的進展值得後續研究(17)考慮 SCM 分佈近似干預和反事實的效果。