**1. Accuracy – Overall Prediction Correctness**

| **Observation** |  |
| --- | --- |
| **XGBoost** improved from **0.9714 → 0.9889** (best result overall) |  |
| **Hybrid Model** also improved (0.96 → 0.986), very close to XGBoost |  |
| **Random Forest** showed strong improvement (0.9851 → 0.9900) |  |
| **KNN** slightly dropped (0.9573 → 0.9404), possible overfitting or sensitivity to imbalance |  |
| ✅ All models except KNN,NN and SVM showed **accuracy gain** after tuning |  |
| ⚠️ **Baseline SVM** had poor accuracy (0.8329), tuning improved it drastically to 0.9700  **Conclusion:** ✅ XGBoost and RF achieved the **highest accuracy** after tuning.    **Precision (Fraud) – How Many Predicted Frauds Were Correct**   | **Observation** |  | | --- | --- | | **XGBoost** rose from **0.7269 → 0.934**, big leap — robust tuning impact |  | | **Random Forest** retained high precision (0.9849 → 0.9633) |  | | **SVM** improved from **0.2908 → 0.7489** — major improvement |  | | **KNN** dropped significantly (0.9783 → 0.5536) |  | | ✅ Hybrid model also saw a good rise (0.65 → 0.8423) |  |   **Conclusion:** ✅ **XGBoost and Hybrid** had the **highest precision** gains and are more reliable in reducing false positives.    **3. Recall (Fraud) – How Many Actual Frauds Were Detected**   | **Observation** |  | | --- | --- | | **Neural Network** improved recall from 0.7269 → **1.0** (but may risk precision) |  | | **Hybrid Model** retained very high recall (0.97 → 0.933) |  | | **XGBoost** went from 0.9528 → 0.9075 (slight tradeoff for better precision) |  | | **Random Forest** improved recall from 0.8019 → 0.8925 |  |   **Conclusion:** ✅ **NN and Hybrid** achieved **highest recall**, good for catching more fraud 📉 **XGBoost slightly sacrificed recall for improved precision**    **4. F1 Score (Fraud) – Balance of Precision & Recall**   | **Observation** |  | | --- | --- | | **XGBoost** improved from **0.8247 → 0.9206** — best overall F1 score |  | | **Random Forest**: 0.884 → 0.9265 — consistently strong |  | | **SVM** improved sharply (0.4451 → 0.8034) |  | | **NN** dropped drastically (0.8407 → 0.1323) — high recall but very low precision |  | | **KNN** improved slightly (0.5721 → 0.6580) |  |   **Conclusion:** ✅ **XGBoost, Random Forest, and Hybrid** are best balanced (high F1 = high usefulness) |  |

**5. AUC-ROC – Overall Fraud Detection Ability**

| **Observation** |  |
| --- | --- |
| **XGBoost** had top performance: **0.9935 → 0.994** |  |
| **Random Forest** improved slightly (0.9881 → 0.9908) |  |
| **Hybrid** saw solid improvement (0.9628 → 0.9874) |  |
| **KNN and SVM** improved as well, but still behind top models |  |

**Conclusion:**  
✅ **XGBoost wins** on this metric too — best at distinguishing fraud vs. non-fraud  
📈 Hybrid is also strong, and worth keeping as your secondary model

A graph of different colored bars

AI-generated content may be incorrect.

**Final Summary: Fair Comparative Analysis**

| **Model** | **Result Summary** |
| --- | --- |
| **XGBoost** | ✅ Consistently top performer across all metrics after tuning. Best **overall model**. |
| **Hybrid (XGBoost + Anomaly)** | ✅ Strong backup, especially for high recall and stable across metrics. |
| **Random Forest** | ✅ Excellent improvement. High F1 and ROC. Great classic ensemble model. |
| **SVM** | 🔁 Huge leap after tuning. Now competitive, though slightly behind XGBoost in precision. |
| **Neural Network** | ⚠️ Very high recall but sacrificed precision → F1 collapsed. May require threshold tuning. |
| **KNN** | 📉 Performance dropped in multiple areas. Less robust for fraud detection. |
| **Decision Tree** | 🔁 Improved with tuning, but not as strong as ensemble models or XGBoost. |

**✅ Final Recommendation**

| **Use Case** | **Recommended Model** |
| --- | --- |
| **Primary research model** | ⭐ **XGBoost** |
| **Secondary model** | ✅ **Hybrid (XGBoost + LOF + IF)** |
| **Balanced, interpretable alternative** | ✅ **Random Forest** |
| **If real-time speed is critical** | 🔁 SVM (post-tuning) |
| **Experimental / future work** | ⚠️ Neural Network (with threshold tuning / ensemble help) |

**Key Concepts from Confusion Matrix:**

| **Metric** | **Meaning** |
| --- | --- |
| **True Negatives (TN)** | Correctly predicted non-fraud |
| **False Positives (FP)** | Non-fraud wrongly predicted as fraud (costs effort, annoying) |
| **False Negatives (FN)** | **Fraud missed** (very risky!) |
| **True Positives (TP)** | Fraud correctly identified |

**Comparative Confusion Matrix Insights Table**

| **Model** | **TN (↓FP) Change** | **TP (↑Recall) Change** | **FP (False Alarms)** | **FN (Missed Frauds)** | **Insights Summary** |
| --- | --- | --- | --- | --- | --- |
| **Neural Net** | **27875 → 76** ⛔ Huge drop | **1541 → 2120** ✅ Max recall | 5 → 27804 ⛔ Skyrocketed | 579 → 0 ✅ None missed | Overfit to fraud class; **precision collapsed** |
| **SVM** | **22978 → 27264** ✅ Big gain | **2010 → 1837** ⬇ Slight drop | 4902 → 616 ✅ Major reduction | 110 → 283 ⬆ More frauds missed | **Much better precision**, mild recall trade-off |
| **KNN** | **27861 → 26494** ⬇ Declined | **857 → 1719** ✅ Strong gain | 19 → 1386 ⬆ Alarming rise | 1263 → 401 ✅ Huge drop | Better recall, but **false positives spiked** |
| **Decision Tree** | **25949 → 27326** ✅ Improved | **1974 → 1950** ⬇ Slight drop | 1931 → 554 ✅ Better precision | 146 → 170 ⬆ Slight increase | Balanced model, now more conservative |
| **Random Forest** | **27854 → 27808** ≈ Same | **1700 → 1892** ✅ Better fraud catch | 26 → 72 ⬆ Slight increase | 420 → 228 ✅ Lower FN | **Excellent trade-off**, stable performer |
| **Hybrid Model** | **26765 → 27510** ✅ Big gain | **2047 → 1978** ⬇ Slight dip | 1115 → 370 ✅ Great precision boost | 73 → 142 ⬆ | **Very strong balance**, reliable choice |
| **XGBoost** | **27121 → 27744** ✅ Big gain | **2020 → 1924** ⬇ Minor drop | 759 → 136 ✅ Great improvement | 100 → 196 ⬆ | **Best precision**, slight recall loss – still top overall |

A screenshot of a graph

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A blue and white squares with black text

AI-generated content may be incorrect.