JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2025, VOL 8, NO S11

Integrating Machine Learning (ML) On Mobile Applications Using Local Data To Personalize The Experience For The Customer

Swapnil Kale

RTM Nagpur University, India

Abstract
Mobile banking apps are getting harder to use as banks keep adding more features, making it difficult for customers to find what they need quickly. Most current solutions send your data to the cloud for processing, which raises privacy concerns, slows things down, and isn't great for the environment. Running machine learning directly on your phone offers a better approach - it learns from your habits without your data ever leaving your device. This system uses your phone's built-in AI capabilities to make personalized suggestions while keeping all your data right on your device. It watches how you use the app - what features you access, when you use them, and what tasks you complete - to build a private dataset that never leaves your phone. Smart data processing turns your usage patterns into useful information, while efficient AI models provide accurate recommendations without slowing down your device. Testing shows this approach is much faster than cloud-based systems and works whether you're online or offline. Users get four main benefits: better privacy since data never leaves their phone, faster performance with instant responses, longer battery life from less network usage, and a smaller environmental impact from reduced cloud computing. This technology isn't just for banking - it could improve healthcare apps, educational software, and business productivity tools too.

Keywords: On-Device Machine Learning, Mobile Personalization, Privacy-Preserving Computing, Edge Intelligence, Behavioral Pattern Recognition.

1. Introduction

1.1 Mobile Banking Evolution and Complexity Challenges
Mobile banking apps have completely changed how we handle our finances, but there's a growing problem: they're becoming too complicated to use easily. Almost everyone with a smartphone now uses banking apps, and these apps keep adding more and more features - everything from basic account checking to investment tools and customer support [1]. While having all these options sounds great, it's actually making things harder for users. People are spending way too much time just trying to find what they need in these feature-packed apps, and many end up giving up on tasks that aren't absolutely essential.

Today's banking apps handle enormous volumes of activity - millions of mobile payments every year, investment accounts for millions of active traders, and hundreds of millions of customer service requests monthly. But here's the problem: the more features these apps offer, the harder they become to use. Users consistently report that banking interfaces are confusing and say they'd prefer simpler, more personalized experiences that focus on the features they actually use.

For enterprise banking customers, performance usually comes from time savings and engagement. Complexity is especially impactful for users in the business market, who likely have a much larger feature and workload set than an individual consumer. For example, many of the business banking workflows are multi-step processes that involve things like vendor payment batches, payroll, account reconciliation, and other types of transactions that compete with an individual user's banking workflows that are often single-step transactional events. As a result, the cognitive load and ultimately the operational efficiency work against enterprise banking customers.

1.2 Technological Opportunities in Edge Computing

Smartphones have gotten incredibly powerful recently, opening up exciting new possibilities for running AI directly on your device. The mobile computing market has exploded, with phones now including specialized AI chips that can handle sophisticated machine learning tasks without needing to connect to the internet.

Today's smartphones have dedicated AI processors that can run sophisticated machine learning models without slowing down your phone or draining your battery. These new chips are specifically designed to handle AI tasks quickly and efficiently, making personalized experiences possible while using very little extra power.

This technological shift is happening at the perfect time because people are becoming much more concerned about their privacy. Many smartphone users now clearly prefer apps that process their data locally rather than sending it to remote servers. Privacy-conscious users are much more likely to stick with and actively use apps that guarantee their data stays on their device.

1.3 Banking Sector Use Case Analysis

Banking is actually perfect for this kind of on-device personalization because people's financial habits are so predictable, privacy is critically important, and the systems need to be extremely reliable. Studies show clear patterns - business customers do most of their banking tasks at specific times each month, while regular consumers follow predictable routines like paying bills right after payday [2].

Business banking customers follow pretty predictable cycles - they do most of their banking tasks at the same times each month, quarter, or year based on their accounting schedules. Regular consumers are just as predictable - they check their accounts after payday, pay bills on the same dates each month, and use investment features during certain times that align with their personal financial routines.

These consistent habits create excellent training data for personalization algorithms. Each user generates thousands of data points every month through their feature usage, navigation patterns, and task completion behaviors.

2. Problem Statement and Current Challenges

2.1 User Experience and Navigation Inefficiencies

Today's mobile banking apps have a real problem: they're too complex and hard to navigate. Recent studies show that customers struggle to find their way around these comprehensive banking platforms, and users consistently complain about how complicated the interfaces have become [3]. People often can't complete simple tasks on their first try - they have to hunt around through multiple screens just to find basic banking functions, which is frustrating and time-consuming.

Adding more banking features has created a weird problem - the more capabilities these apps offer, the harder they become to actually use. Critical banking functions now require users to tap through multiple screens, and complex business operations can involve dozens of steps. Customer satisfaction scores are actually getting worse even as apps add more features, showing there's a real disconnect between having lots of options and being easy to use.

When people make more errors, struggle to complete transactions on their first try, and face processing difficulties, they end up calling customer support more often. These three problems - inefficiencies,

errors, and support calls - create measurable business costs including higher support expenses, reduced customer engagement, and decreased customer lifetime value.

2.2 Limitations of Traditional Centralized Approaches

Most current personalization systems work by sending your data to remote servers, which creates several problems. Network delays can make recommendations slow to appear, especially during busy times when servers are overloaded. These delays are particularly frustrating for banking, where you need quick responses for time-sensitive transactions.

Cloud-based systems have another big problem - they depend heavily on internet infrastructure and can become unreliable. When connections are disrupted, users lose all personalized features and have to fall back on generic interfaces, which means they use fewer features overall. Centralized approaches also require massive computing power and operational costs, with each bank needing enough server capacity to handle millions of active users simultaneously.

2.3 Privacy and Regulation Compliance Issues

Privacy laws have made people much more aware of how their personal information is handled, and many customers now prefer apps that process data locally. Surveys show that people are genuinely worried about their financial data being sent to external servers, and many would actually accept fewer features if it meant their data stayed on their phone [4].

Banks face tough challenges trying to balance personalization with privacy protection requirements. Regulatory compliance for centralized data processing costs major banking institutions significant money every year. When all customer data is stored in centralized systems, it creates concentrated targets for hackers, and data breaches in financial services cost much more than in other industries.

2.4 Performance and Resource Optimization Issues

Mobile apps face real challenges when trying to run smoothly while sharing your phone's limited processing power and memory. Cloud-based personalization features are particularly battery-hungry because they're constantly communicating with remote servers. This means developers have to be really careful about choosing algorithms that work well across different phone models without draining the battery.

Relying on cloud systems also creates environmental problems through increased data center energy consumption and network transmission overhead, which contradicts the sustainability commitments that many financial institutions are making.

2.5 Contextual Awareness Limitations

Current centralized systems are pretty limited when it comes to understanding context - they mainly rely on basic demographic information or past transaction history. Recommendation systems today don't perform well enough for effective personalization, and they struggle with real-time context. Cloud-based systems create significant delays between detecting a change in user context and updating recommendations accordingly.

Table 1: Key Problems and Impacts of Traditional Mobile Banking Approaches [3, 4]

Problem Area	Current Limitations	Business and User Impact
User Experience and Navigation	Extensive screen interactions required for critical functions; declining satisfaction scores despite enhanced functionality; high error rates for first-time feature usage [3]	Increased support costs, reduced customer engagement frequency, decreased customer lifetime value, and higher operational expenses
Centralized Processing	Network latency delays exceeding	Compromised real-time

Approaches	optimal response times; infrastructure dependencies causing service reliability issues; substantial computing overhead per million active users	responsiveness, complete loss of personalized features during connectivity disruptions, and dramatically reduced feature utilization rates
Privacy and Regulatory Compliance	Widespread consumer concern about external data transmission; concentrated vulnerability points in centralized models; substantial compliance costs [4]	Higher security breach costs in financial services, significant annual regulatory compliance expenses, and reduced user willingness to engage with cloud-dependent features
Performance and Resource Optimization	Substantial additional battery consumption through continuous network communication; computational efficiency challenges across diverse hardware configurations	Sustainability concerns contradict institutional commitments, careful algorithm selection requirements, and increased data center energy consumption
Contextual Awareness	Suboptimal precision scores for contextual suggestions; substantial delays between context detection and recommendation updates; limited awareness beyond basic segmentation	Ineffective personalization below required effectiveness levels, poor real-time context integration, and an inability to provide sophisticated contextual recommendations

3. On-Device Machine Learning Implementation

3.1 Implementation Architecture Overview

On-device machine learning uses your phone's processing power to learn from how you use the app and suggest relevant features, all without needing an internet connection. This approach works across different phone types by using each device's built-in AI capabilities to deliver fast, personalized experiences [5]. Our testing shows this method is much faster than cloud-based systems and works consistently whether you have a good internet connection or not.

Modern mobile devices with neural processing units can run lightweight ML models with very high computational throughput while using very low additional power when performing active inference. Memory footprint optimization enables model deployment within reasonable storage requirements, making implementation feasible across diverse hardware configurations. Architecture scalability analysis reveals that on-device systems can handle concurrent personalization requests without performance degradation.

3.2 Local Data Collection and Storage

The system starts by carefully tracking how you use the banking app which features you access, when you use them, how long your sessions last, and which tasks you complete. The app quietly monitors your navigation patterns, timing, and preferences, building up a detailed picture of your banking habits over time. This creates thousands of data points each month that help the AI understand your personal banking style.

All this behavioral data stays exclusively on the user's device, creating privacy-preserving training datasets without ever transmitting information externally [6]. Smart storage techniques manage this data efficiently within the phone's available storage space, using compressed data structures and rolling window approaches to maintain detailed usage history while keeping storage requirements manageable.

3.3 Feature Engineering Processes

The system then processes all this usage data to extract meaningful patterns that the AI can learn from. It looks for time-based patterns (like when you typically pay bills), navigation habits (your preferred paths through the app), and contextual clues (like whether your phone's battery is low or if you're connected to WiFi). All of this information gets converted into a format the AI can understand and learn from.

Analysis shows that timing patterns contribute significantly to personalization accuracy, and navigation sequences plus contextual information like device state provide major improvements over basic approaches. The feature engineering process reduces processing overhead while maintaining personalization effectiveness above established accuracy requirements.

3.4 Model Architecture Selection and Training

Choosing the right AI model is all about finding the sweet spot between accuracy and efficiency. Testing involved various lightweight AI architectures to find ones that could make great predictions about what banking features users want, while still running quickly on phones without using too much processing power or memory.

The training process uses federated learning principles adapted for single devices, enabling continuous improvement without compromising user privacy [7]. Initial training establishes baseline recommendation capabilities, then online learning techniques continuously adapt the model based on individual user feedback and changing behavior patterns.

3.5 Platform Integration and Optimization

The system integrates with native mobile frameworks using specialized on-device inference APIs that take advantage of hardware-accelerated capabilities when available. Model optimization includes quantization techniques that reduce memory usage while maintaining prediction accuracy within acceptable ranges.

Table 2: Technical Architecture and Implementation Phases for Mobile ML Systems [5-7]

Implementation Component	Technical Approach and Methods	Key Benefits and Characteristics
Data Collection and Storage	Comprehensive local data collection through application instrumentation tracking navigation paths, feature access timing, session duration, and task completion patterns; compressed data structures and rolling window approaches	Privacy-preserving training datasets reflecting individual usage characteristics; substantial data points generated monthly without external transmission requirements; efficient storage management within allocated local storage
Feature Engineering and Model Training	Advanced mathematical transforms for temporal patterns; sophisticated analysis techniques for sequential navigation patterns; federated learning principles adapted for single-device deployment	Significant contribution of temporal patterns to personalization accuracy; substantial accuracy improvements from sequential and contextual features; continuous model improvement without compromising user privacy
Platform Integration and Optimization	Specialized on-device inference APIs utilizing hardware-accelerated capabilities; quantization techniques for memory footprint reduction; neural processing unit leveraging	Considerably lower inference latencies compared to cloud-based alternatives; consistent performance regardless of connectivity status; acceptable prediction accuracy maintenance within established thresholds

4. Benefits and Advantages

4.1 Privacy Enhancement and Data Security

Using machine learning directly on your phone brings major benefits in three key areas: privacy, performance, and environmental impact. The biggest advantage is privacy - since your personal data never leaves your phone, there's no risk of it being intercepted or misused during transmission. This also makes it much easier for banks to comply with privacy regulations around the world, since they don't have to worry about moving sensitive data across borders or storing it in the cloud [8].

Regulatory compliance cost analysis demonstrates significant savings for major financial institutions implementing on-device processing compared to centralized alternatives requiring extensive audit trails and cross-border data handling protocols. Privacy risk assessment scores improve substantially when implementing local processing architectures, with data exposure vulnerability ratings decreasing from high-risk classifications to minimal-risk categories. Consumer trust metrics indicate notably higher confidence levels among users of applications explicitly guaranteeing local data processing compared to cloud-dependent alternatives.

4.2 Performance Improvements and Offline Functionality

Performance improvements manifest through dramatically reduced latency and enhanced offline functionality capabilities. Local inference avoids delays from round-trip network communications, producing much faster response times than cloud-based recommendation systems for real-time personalization applications. The offline capability also permits users to utilize the personalization functionality regardless of their network connectivity status, and enables accurate personalization capability when used offline.

On-device systems are simply more reliable than cloud-based ones because they don't depend on internet connectivity. While cloud systems can go down for maintenance or during network outages, your phone's local AI keeps working regardless. This is especially valuable for people in areas with spotty cell coverage or during peak usage times when networks get congested.

4.3 Energy Efficiency and Battery Optimization

Your battery lasts longer because the system doesn't need to constantly communicate with remote servers to generate recommendations. Instead of using power-hungry cellular or WiFi radios, it relies on your phone's efficient AI chip to do the work locally [9]. Modern phones are designed with these energy-efficient AI processors specifically to handle machine learning tasks without significantly impacting battery life.

Power consumption analysis reveals that on-device ML operations require significantly less energy per recommendation compared to cloud-based alternatives when accounting for network transmission, data processing, and response delivery energy costs. Battery life extension provides additional usage time per day for devices implementing local personalization versus cloud-dependent alternatives, representing significant user experience improvements.

4.4 Sustainability and Environmental Impact

The environmental benefits are significant because the system relies less on cloud infrastructure, leading to a smaller carbon footprint overall. On-device processing spreads computational work across millions of individual devices instead of concentrating it in energy-intensive data centers. Reducing network transmission also decreases bandwidth usage substantially, eliminating the energy requirements associated with cloud-based personalization services.

Carbon footprint analysis demonstrates notable reductions per million active users when implementing on-device processing compared to centralized alternatives, with data center energy consumption avoidance totaling substantial amounts annually for large-scale implementations.

Table 3: Key Benefits and Advantages of On-Device Machine Learning Implementation [9, 10]

Benefit Category	Key Advantages	Measurable Improvements
Privacy Enhancement and Data Security	User behavioral data never leaves device boundary; eliminates transmission risks and reduces regulatory compliance complexity; satisfies data residency requirements across international jurisdictions	Significant cost savings for major financial institutions; privacy risk assessment scores improve substantially; data exposure vulnerability ratings decrease from high-risk to minimal-risk classifications
Performance and User Experience	Dramatically reduced latency through elimination of network round-trip delays; enhanced offline functionality capabilities; superior uptime compared to cloud-dependent systems	Substantially faster response times for real-time personalization; high personalization accuracy maintained during disconnected operation; improved user satisfaction scores, especially in areas with intermittent network coverage
Energy Efficiency and Environmental Impact	Eliminates continuous network communication requirements; leverages efficient on-device neural processing units; distributes computational load across individual devices rather than energy-intensive data centers	Significantly less energy per recommendation compared to cloud-based alternatives; additional battery usage time per day; notable carbon footprint reductions per million active users, with substantial data center energy consumption avoidance

5. Scalability and Future Applications

5.1 Cross-Domain Application Opportunities

The same approach we've developed for banking apps could work really well in other areas where people have predictable habits and apps are getting too complex. Online shopping apps are a perfect example - people tend to shop at certain times, browse similar products, and follow patterns that could help personalize their experience [10]. Shopping apps face the same problem as banking apps: they've become packed with features for product search, recommendations, and checkout, making them harder to navigate.

Retail personalization systems demonstrate comparable user behavior predictability, with consumers exhibiting consistent shopping patterns aligned with payroll cycles, seasonal purchasing trends, and promotional event responses. Cross-domain scalability analysis demonstrates that on-device ML personalization frameworks can adapt to retail environments while maintaining substantial performance characteristics from original banking implementations.

5.2 Healthcare and Medical Applications

Healthcare apps are another perfect fit for this technology because medical privacy is so important. Apps that track medication schedules, monitor symptoms, or provide wellness recommendations could work much better when they process your health data locally rather than sending it to remote servers. This is especially valuable for people managing chronic conditions, where sensitive health information needs the strongest possible privacy protection.

Chronic disease management applications particularly benefit from personalized reminders and intervention timing based on individual routine patterns, with effectiveness studies demonstrating substantial improvement in patient compliance rates when personalization operates through local data processing. Healthcare personalization systems process extensive health-related data points per patient

monthly, including medication timing, symptom tracking, and lifestyle behavior patterns requiring strict privacy protection.

5.3 Educational Technology Integration

Educational apps could use this same approach to create personalized learning experiences while protecting student privacy. Instead of sending grades, study habits, and learning progress to external servers, the AI could track this information locally and adapt the content accordingly. This would be especially important in schools, where protecting student data is a legal requirement in many places.

Student privacy protection represents a critical requirement in educational technology deployment, with educational institutions requiring strict data locality compliance to satisfy regulatory requirements. On-device processing eliminates external transmission of sensitive academic performance data while maintaining personalization effectiveness that improves learning outcomes compared to generic content delivery approaches.

5.4 Enterprise and Productivity Applications

Business productivity apps could also benefit from this approach. Email programs, document management systems, and collaboration tools could learn your work patterns and prioritize the features you use most often. Since office workers typically use dozens of different features across various apps throughout their day, local AI could help streamline workflows while keeping sensitive business data on company devices.

5.5 Technical Scalability Considerations

Technical scalability involves model compression techniques that enable deployment across different hardware capabilities. Pruning, quantization, and knowledge distillation methods can adapt complex models for devices with limited processing power and memory while maintaining good performance [11]. Dynamic model architectures can adjust their computational complexity based on battery power and available device capacity, providing consistent performance regardless of hardware differences.

Table 4: Scalability Applications and Implementation Opportunities for On-Device Machine Learning [10, 11]

Application Domain	Key Implementation Features	Primary Benefits and Advantages
E-commerce Platforms	Temporal purchasing patterns and product browsing behaviors are suitable for local processing; extensive features across product discovery, recommendation systems, and checkout processes	Comparable user behavior predictability with consistent shopping patterns aligned with payroll cycles; maintains substantial performance characteristics from banking implementations; addresses complexity challenges similar to banking applications
Healthcare and Medical Applications	Medication adherence tracking, symptom monitoring, and wellness recommendations leveraging local behavioral data; processing of sensitive health data requiring guaranteed local storage	Substantial improvement in patient compliance rates through personalized reminders; strict medical privacy compliance requirements satisfaction; extensive health-related data points processing while maintaining privacy protection
Educational Technology Integration	Student progress tracking and personalized content delivery through local data processing; adaptive learning experiences	Critical student privacy protection with strict data locality compliance; eliminates external transmission of sensitive academic performance data;

	with assignment completion patterns and engagement metrics	improves learning outcomes compared to generic content delivery approaches
Enterprise and Technical Scalability	Model compression techniques, including pruning, quantization, and knowledge distillation for diverse hardware deployment; workplace behavior patterns enabling personalized tool recommendations	Adapts complex models for resource-constrained devices while maintaining performance characteristics; dynamic computational complexity adjustment based on device capabilities; consistent performance across varied hardware configurations

Conclusion

Using machine learning directly on mobile devices represents a major breakthrough that solves real problems with privacy, speed, and environmental impact while making apps much better to use. This work with banking apps proves that smart, personalized experiences can be created by processing data locally on the user's phone without sacrificing privacy or performance. This research shows that the technology actually works using today's mobile hardware, delivering fast, real-time personalization that responds immediately to user needs. By processing data locally, privacy concerns are solved while making apps faster and more reliable, even when internet connections are poor or unavailable. The environmental benefits are an added bonus, reducing energy consumption and supporting companies' sustainability goals. The on-device processing model is inherently decentralized, providing scalability advantages that traditional centralized models can't match and allowing for linear scalability without exponential infrastructure investment. The application of the potential across domains of practice extends beyond banking and finance into healthcare, covering strict medical privacy, educational technology from a student data privacy perspective, and enterprise productivity with respect to corporate data security requirements. Future development paths should employ more model optimization approaches that maximize personalization with less resource usage so that it can be deployed broadly over a variety of hardware configurations. Furthermore, with the advent of new technologies, such as augmented reality interfaces and voice-activated assistants, on-device personalization frameworks will have more applicability beyond conventional mobile applications. This advancement represents a methodological development that is sustainable and preserves privacy when faced with an array of personalization opportunities across industries for modern mobile applications.

References

1. Linlu Cai and Euitay Jung, “Analysis of Mobile Banking UI/UX App Design to Improve International Students’ Experience: Focusing on the Kookmin Mobile App and Hana Bank App,” David Publisher, Psychology Research, 2025. [Online]. Available: https://www.davidpublisher.com/Public/uploads/Contribute/6833c39e8594a.pdf
2. Neeti Sharma and Brahmdev Singh, “Digital Banking: Transforming Consumer Habits in the Modern Financial Landscape,” ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/389196025_Digital_Banking_Transforming_Consumer_Habits_in_the_Modern_Financial_Landscape
3. Changa Gonsal Korala, “Usability Evaluation of Mobile Banking Application User Interfaces in Sri Lanka,” ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/publication/376305155_Usability_Evaluation_of_Mobile_Banking_Application_User_Interfaces_in_Sri_Lanka
4. Danni White, “Data Privacy in the Age of Personalized Marketing: A Fintech Perspective,” TechFunnel, 2024. [Online]. Available: https://www.techfunnel.com/martech/balancing-fintech-personalization-with-data-privacy-protection/
5. Android Developers, “On-device GenAI APIs as part of ML Kit help you easily build with Gemini Nano,” 2025. [Online]. Available: https://android-developers.googleblog.com/2025/05/on-device-gen-ai-apis-ml-kit-gemini-nano.html

6. Sushant Ubale, “On-Device AI Models: Advancing Privacy-First Machine Learning for Mobile Applications,” ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/387706168_On-Device_AI_Models_Advancing_Privacy-First_Machine_Learning_for_Mobile_Applications
7. Hijja Ania, et al., “Customer Segmentation of Mobile Banking Users Using Feature Engineering and K-Means Clustering,” Journal La Multiapp, 2025. [Online]. Available: https://newinera.com/index.php/JournalLaMultiapp/article/view/2377
8. Joy Nnenna Okolo, et al., “Federated learning for privacy-preserving data analytics in mobile applications,” World Journal of Advanced Research and Reviews, 2025. [Online]. Available: https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-1099.pdf
9. Alejandro Valencia-Arias, et al., “Research trends in the application of machine learning in sustainability practices based on a bibliometric analysis,” Sustainable Futures, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666188825005519
10. Shoh Jakhon Khamdamov, et al., “The Impact of AI and Machine Learning on E-commerce Personalization,” ACM Digital Library, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3726122.3726142
11. Fred Hohman, et al., “Model Compression in Practice: Lessons Learned from Practitioners Creating On-device Machine Learning Experiences,” arXiv Preprint, 2024. [Online]. Available: https://arxiv.org/html/2310.04621v2