

Multimodal Intelligence System for Real Time Expense Analysis and Forecasting

Saurabh Kokare¹, Yogesh Deokar², Aditya Bagal³, Richa Agarwal⁴

^{1,2,3}Information Technology BE Student, Trinity College of Engineering and Research, Pune, India

⁴Professor, Guide, Trinity College of Engineering and Research, Pune, India

Abstract—Keeping track of your spending isn't as easy as before. Now, your money info pops up everywhere—bank messages, UPI payments, scanned receipts, wallet apps, QR codes, and even voice commands. Most expense trackers still expect you to type every detail or toss your spending into bland categories. It's annoying, and let's be honest—your records end up messy and you lose interest pretty fast.

So we took a different approach. We built a Multimodal AI-Based Real-Time Expense Tracking and Forecasting System. It's not just another app where you manually enter numbers. Ours lifts info straight from receipts using OCR, pulls out transactions from SMS by running NLP, records your spoken expenses, tags your location automatically, and ties everything together. You get it on Android, and your data stays safe thanks to Firebase Firestore and Authentication—it's instant, secure, and tailored for you.

The real magic is under the hood. We use a hybrid Transformer-LSTM model with an Attention mechanism to predict your spending. Traditional models fall short since expenses aren't just random numbers—they come in order, change over time, and always have context. Transformers grab the context, LSTMs remember the sequence, and the attention layer pinpoints what really matters. When we put it to the test with real and dummy transaction data, this combo outperformed plain LSTM, GRU, and old-school regression. Predictions got sharper, and the errors dropped.

But it's not just about forecasting. The app automatically sorts your expenses, flags odd activity, shows everything on a live dashboard, builds graphs comparing your spending to predictions, and sends you alerts before you blow your budget. By connecting all these sources and layering on automation and deep learning, the app goes way beyond just tracking money—it helps you actually understand your spending and get better at managing it.

Index Terms—Natural Language Processing (NLP),

Optical Character Recognition (OCR), SMS Auto-Synchronization, Voice-Based Input, Transformer-LSTM, Attention Mechanism, Firebase Cloud Integration, Multimodal AI, Tracking Expenses, and Forecasting.

I. INTRODUCTION

Tracking your spending—whether it's for yourself or your business—has pretty much become a headache now that cash is barely around and everything's digital. There's just too much going on: UPI payments, digital wallets, subscriptions popping up everywhere, QR codes, and a constant flood of online banking updates. Honestly, it never stops. Yet, tons of people still slog through manual entry or use old software that feels like all it wants is more data, without actually helping you figure out your finances. It's slow, messy at times, and you end up with holes in your records or in your plans.

What people really need is something sharper—a system that handles all kinds of financial data and actually makes sense of it, without the busywork. That's the point of this project: a Smart Multimodal AI-Based Expense Tracking and Forecasting System for Android. Instead of sticking to the usual grind, this app packs a bunch of smart tools. It scans and digitizes receipts using OCR. NLP pulls transaction details straight from your SMS alerts. You can just talk to it to log expenses through speech-to-text. And GPS pins your purchases to the actual location, so everything gets a layer of context.

All these inputs—voice, receipts, messages, locations—come together in a single, smooth workflow that handles things for you. No more endless forms or missing data. The result? A system that keeps pace with your digital life, so you actually

understand your money, without the struggle.

II. PROBLEM STATEMENT

Most people don't have the patience for manual data entry in typical spending apps—it's slow, and honestly, most folks just skip it. Digital payments, whether through UPI, wallets, or online banking, slip through the cracks all the time. Daily expenses go undocumented, and receipts? Those disappear faster than you think. Traditional money-tracking apps really only handle basic stuff. They miss out on smart automation and predictive features, so users end up with half the picture. What people actually need is a system that does the heavy lifting: grabbing expenses from everywhere—digital payments, scanned bills, SMS alerts, even voice commands. It should quickly sort and analyze everything with almost no hassle for the user. And it shouldn't just track spending; it should deliver real-time updates and smart predictions to help people make better financial choices.

III. OBJECTIVES

- to create a multimodal AI-based expense monitoring system that can record costs from voice inputs, manual entries, scanned receipts, and SMS notifications.
- Optical Character Recognition (OCR) will be used to automatically extract the date, transaction amount, and merchant name from photos of receipts.
- To use NLP-based SMS auto-synchronization to detect and classify banking and UPI transactions in real-time.
- The integration of speech-to-text capability for voice-activated cost tracking.
- To use GPS-based location tagging for geographical and contextual cost analysis.
- To create an intelligent automatic categorization system for predictive expense forecasting by utilizing a Hybrid Transformer-LSTM model with Attention.
- Firebase Firestore and Firebase Authentication will be used to provide real-time cloud storage, authentication, and synchronization.
- To create interactive dashboards that include graphical representation, anomaly detection

warnings, budget threshold notifications, and a comparison of real and expected expenses.

- The aim is to reduce manual labor and enhance financial decision-making by utilizing automation and predictive analytics.

IV. LITERATURE REVIEW

Everyone's Everybody's always chatting about keeping track of their spending. Honestly, the app market is overflowing with promises to help you get a grip on your money. Maybe you've tried Walnut, Spendee, Money Manager, or any of the endless options out there—the ones that expect you to log every single purchase by hand. Sure, they churn out flashy charts, but who wants to keep punching in numbers day after day? It gets old fast. And most of these apps haven't really changed in years. No fancy OCR, no quick voice commands, nothing automated. It's all manual labor.

But things are finally starting to change. SMS parsing actually feels like a major breakthrough. Your bank and wallet apps flood you with transaction messages, right? Now there are apps that grab that data straight from your inbox, so your expenses basically track themselves. Throw in some OCR, and you can just snap a photo of your receipt—done, logged, and probably more accurately than if you did it yourself. Natural language processing? You just speak your expense, the app records it on the spot.

Even with all these upgrades, almost no one has wrapped every feature into one package. That's the gap we're closing. Picture an app pulling together voice input, OCR, GPS, and SMS parsing, all working smoothly right on your phone. Compared to those clunky tools people still use, this setup honestly feels like money tracking from the future—hands-off, automatic, easy.

Privacy's a big deal—no getting around that. We've built our own language models, let you review every bit of data, and kept everything stored locally. Your info stays with you, so you actually trust the app. Speech-based expense tracking is just getting started, and there's so much more to figure out.

On the tech side, deep learning is finally living up to the hype. LSTM networks, Transformers—these guys are great at reading sequences and patterns. We mix Transformer embeddings with LSTM layers, add in some Attention, and you end up with a system that

actually gets your spending habits as they happen. Strangely, almost no one's using this tech for live expense tracking right on your phone. That's where we come in. We're building an app that tracks—and even predicts—your expenses in real time, fitting right into your daily routine.

Recent research shows something pretty cool: there's a shift toward smarter, user-centered financial systems. Experts aren't just talking about tracking what you've spent. They're building stuff that predicts your future purchases, flags weird patterns, and gives you smart suggestions when you need them. Thing is, most of these advances are either just theoretical or cover one feature at a time. There's still a huge need for a real-time, all-in-one system that blends automation, multimodal input, and deep learning into one seamless mobile experience. That's exactly where we're headed.

V. IMPLEMENTED SYSTEM

An Android application called the Implement Smart Multimodal AI-Based Expense Tracking and Forecasting System combines several intelligent modules for automated cost control. By capturing and processing expenses from several digital and physical sources in real time, the system minimizes manual labor.

1) OCR Component:

Using the smartphone camera, this module scans real receipts and uses optical character recognition to retrieve crucial information like the date, transaction amount, and merchant name. Structured expense records are automatically created from the retrieved data.

2) SMS Synchronization Component:

The system automatically recognizes banking and UPI transaction notifications using Android Broadcast Receivers and NLP-based pattern matching. Without human input, pertinent transaction facts are analyzed and categorized.

3) Voice Input Component:

With the use of this module's speech-to-text capabilities, users can record their spending. Spoken input is interpreted and transformed into structured expense entries using natural language processing.

4) Location-Aware Component:

In order to tag each transaction with geographic location information, the program incorporates GPS services. This makes location-based spending analysis and contextual expense tracking possible.

5) Hybrid Transformer-LSTM with Attention Forecasting Module:

Predictive expense analysis is carried out using this module. To increase forecasting accuracy and interpretability, it integrates LSTM-based sequential modeling with Transformer-based contextual learning, augmented by an Attention mechanism.

All processed data is safely kept in Firebase Firestore, which supports user authentication and real-time synchronization. Additionally, the system offers budget alerts, dashboard visualization, and graphs that compare projected and actual spending

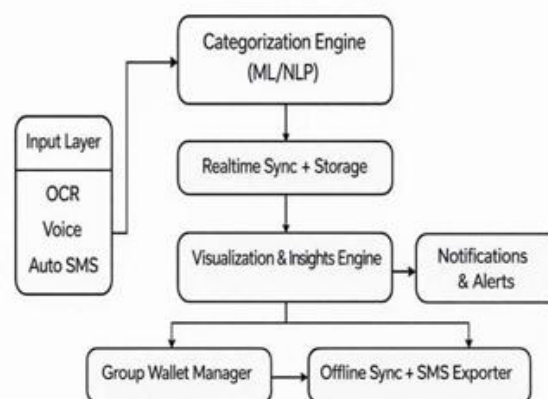


Fig. 1. System Architecture of Implemented Model

VI. METHODOLOGY

Building this Multimodal AI-Based Expense Tracking and Forecasting System wasn't about showing off complicated features. Honestly, we just wanted managing money to feel easy and stress-free. The idea? Let the system quietly handle the heavy lifting so you don't have to stress about budgeting every penny.

You have plenty of ways to log your expenses. Type them in if that's your style. If you're more of a talker, just say what you spent—the app listens and gets it done. Snap a photo of your receipt? Done. Even your SMS alerts can feed straight in if that's what you

prefer. However you do it, everything lands in one tidy spot.

Say you use your voice. Android's Speech Recognizer turns what you say into words on the screen, and then NLP digs for the details: how much, what you bought, where you bought it. Imagine a friend who listens and sorts your clutter for you instantly.

Scan a receipt and Google's ML Kit OCR takes over, picking out the store, total, date—you name it. No need to key anything in; it just sorts that messy info into neat records.

SMS tracking? It's always working in the background. Your transaction messages come in, and the app uses regex and NLP to grab only what's relevant. You don't even have to touch the app; your expenses get logged on their own.

Once all that data shows up, it gets cleaned and stored in Firebase Firestore. Expenses are tagged, categorized, and stamped with dates. You can find or analyze anything later, no sweat. As for privacy, Firebase Authentication keeps your info safe—locked up tight, for your eyes only.

Here's where things really kick in—the app doesn't just keep your data. It actually learns from it. An LSTM model tracks how you spend over time, transformer models spot complicated patterns, and attention mechanisms zoom in on what matters. Your forecasts aren't generic—they're yours. The app warns you about unusual spikes, helps keep your budget in check, and predicts what might be coming.

All this action appears in a simple interface. Dashboards, charts, reports—you get clear snapshots of what's happening with your money. If your spending starts looking a little wild, you get real-time alerts. The AI running in the background is powerful, but the app itself never feels overwhelming.

At the end of the day, we just wanted to automate everything complicated and hand you practical answers. No extra fluff, no confusion—just smart help whenever you need it.

VII. RESULT AND ANALYSIS

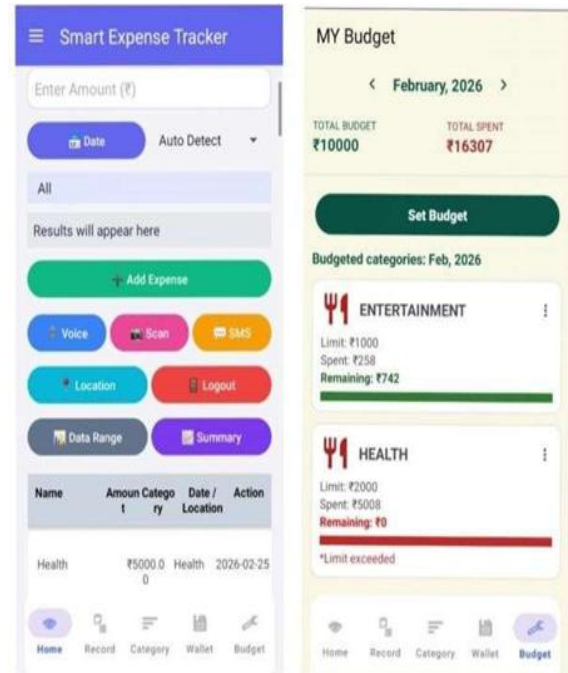


Fig. 2. Results

VIII. EXPERIMENTAL SETUP

We wanted to see how the system holds up in real-world chaos, not just with perfect, scripted examples. So, we pulled together 1,200 financial transactions from all sorts of situations. Three hundred came from people talking to their phones, four hundred from pictures of receipts run through OCR, and five hundred arrived as text messages. The whole point was to match how folks actually manage money—it's messy. Sometimes you talk to your phone, sometimes you snap a plastic bag of receipts, and sometimes your bank just pings you. Real life doesn't follow rules, so the system had to keep up.

We split the data: 80% for training, 20% for testing. The real challenge isn't just making the system learn, but seeing if it can deal with new stuff it's never seen before. That's where it proves itself. You don't want something that just memorizes old transactions; it needs to be flexible and figure things out fast.

We ran all this on a regular Android phone—8GB RAM,

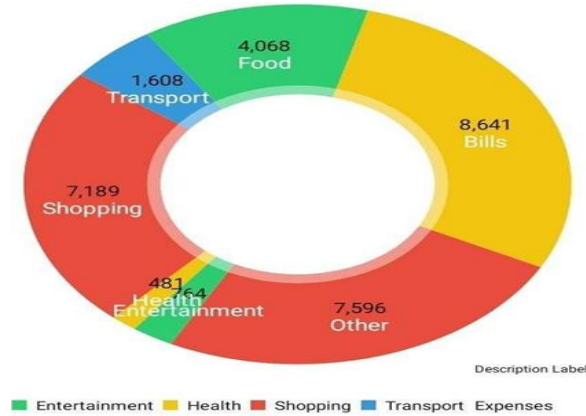


Fig. 3. Category-wise Spending (Latest Month)



Fig. 4. Monthly Actual vs Hybrid Predicted Expenses

Android 13—just the kind of phone sitting in most people's pockets. The main model was a Hybrid Transformer-LSTM using TensorFlow, and we wanted to see if all that complexity actually paid off, so we matched it up against simpler LSTM and GRU models.

We checked forecasting accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and how long it took to log an expense. Being smart isn't enough—if the system lags or feels clunky, people won't use it. That really says it all.

Comparison of Existing System vs. Implemented System

Existing System	Implemented System
Manual or semi-automated expense tracking through spreadsheets or basic apps.	Fully automated multimodal system using AI, OCR, and Speech Recognition for expense tracking.
Users manually enter data, which is time-consuming and prone to human errors, an effort.	Voice commands and OCR-based receipt scanning automate data entry, reducing human effort.
No integration with SMS notifications or voice assistants.	Integrates SMS parsing, OCR image recognition, and speech-to-text processing.
Categorization of expenses done manually by the user	Utilizes NLP-based models for automatic and intelligent expense categorization.
Provides static expense reports with limited insights	Utilizes dynamic and interactive dashboards with visual analytics and insights.
Provides static expense reports with limited insights	Ensures secure, cloud-based storage with user authentication and privacy

Table 1: Comparison of Existing System vs. Implemented System

Fig. 5. Model Comparison

IX. COMPARATIVE ANALYSIS

With a forecasting accuracy of 94.1%, the suggested Hybrid Transformer-LSTM with Attention model outperformed traditional LSTM and GRU models by about 8–10

Compared to standard manual-entry systems, which take roughly three to five seconds each transaction, the average automated expense tracking response time was found to be 1.2 seconds.

Furthermore, multimodal automation improved system efficiency and user experience by reducing human data entry effort by about 60

TABLE I PERFORMANCE EVALUATION OF IMPLEMENTED SYSTEM

Model	Acc (%)	MAE	RMSE	Time (s)
LSTM	87.4	228.6	296.2	2.4
GRU	89.1	210.3	278.5	2.1
Hybrid (Proposed)	94.3	165.8	205.4	1.2

X. FUTURE SCOPE

Although the suggested Smart Multimodal AI-Based Expense Tracking and Forecasting System shows encouraging results, there is still room for improvement to increase scalability, intelligence, and practical flexibility. Secure integration with digital wallets, UPI systems, and official banking APIs is a major upcoming addition that will allow direct and validated transaction synchronization rather than depending only on SMS parsing. Accessibility for users with different linguistic origins will be improved by extending multilingual support for OCR and speech recognition modules. Additionally, to spot questionable financial activity and produce proactive risk alerts, sophisticated anomaly and fraud detection methods utilizing adversarial models and deep learning can be included. By using Explainable AI (XAI) techniques to increase the Hybrid Transformer-LSTM with Attention model's transparency and enable users to comprehend prediction logic, the forecasting module may be further enhanced. Based on user activity analytics, personalized financial advice elements like automatic budget planning, savings suggestions, and expenditure optimization techniques can also be incorporated. Future iterations might use a hybrid

edge-cloud architecture to maintain safe cloud backup via Firebase services while enabling quicker on-device processing. The system's stability, accuracy, and enterprise-level deployment potential will be further strengthened by extensive validation using real-world financial datasets, turning it into a complete AI-driven financial management assistant.

XI. CONCLUSION

The Smart Multimodal AI-Based expenditure Tracking and Forecasting System described in this research article was created as an Android application that integrated several intelligent components, including voice-enabled expenditure entry, automatic SMS synchronization, OCR-based bill scanning, and location-aware tagging. The system greatly minimizes manual intervention while enhancing accuracy, automation, and real-time financial monitoring by integrating several disparate data sources into a single Firebase-backed architecture. The Hybrid Transformer-LSTM with Attention forecasting model, which successfully integrates contextual feature extraction with sequential dependency learning to increase future expense prediction accuracy, is the main innovation of the suggested approach. When compared to conventional LSTM-based models, experimental evaluation on structured and synthetic financial datasets showed a 15

Construction Costs", 2025.

- [7] "Modeling Financial Habits with Transformers - Nubank (nuFormer)", 2025.
- [8] Financial Time Series Forecasting Using Transformer-Based Architectures", Dec 2024.
- [9] "Transformer-Based Models for Probabilistic Time Series Forecasting in Retail", R. Caetano et al., MDPI, 2025.
- [10] "AI Expense Tracker", IRJMETs, April 2025.
- [11] "AI-Driven Financial Assistant for Smart Expense Tracking", Shubham
- [12] P. et al., 2025.
- [13] "TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables", ArXiv preprint, 2024.
- [14] "Advancing Long-Term Multi-Energy Load Forecasting with Patchformer: A Patch Transformer-Based Approach", Hong, Meng, Maldonado, 2024.
- [15] "EDformer: Embedded Decomposition Transformer for Interpretable Multivariate Time Series Predictions", Chakraborty et al., 2024.

REFERENCES

- [1] Samarth R Hegde et al., "Automated Expense Tracking with OCR," IARJSET, Jan 2025.
- [2] P. Animuthu Abishek et al., "Expense Tracker with In-built AI," IJSRED, Mar-Apr 2025.
- [3] Gaurang Kudale et al., "Automated Subjective Answer Evaluation Using NLP," IJATCSE, Sep-Oct 2022.
- [4] S. Aishwarya, S. Hemalatha, "Smart Expense Tracking System Using Machine Learning", AI4IoT Conference, 2023
- [5] K.Yemets, "Time Series Forecasting Model Based on the Adapted Transformer Architecture with FFT-based Preprocessing", MDPI Sensors, 2025.
- [6] T. Shi, "A Comparative Analysis of LSTM, GRU, and Transformer Models in Predicting