

Project ID:

R26-IT-026

1. Topic (12 words max)

SmartCost: AI-Driven Supply Chain Cost Analytics and Profit Prediction System for Micro-SMEs

2. Research group the project belongs to

TIM - Technology Integration and Management

3. Specialization of the project belongs to

Information Technology (IT)

4. If a continuation of a previous project:

Project ID	-
Year	-

5. Brief description of the research problem including references (300 words max) – references not included in word count.

- The Digital Gap in MSMEs Particularly in the food and beverage industry, Micro, Small, and Medium Enterprises (MSMEs) have incredibly narrow profit margins. Revenue tracking has been successfully digitised by the use of Point-of-Sale (POS) systems, but Cost of Goods Sold (COGS) management is still primarily done by hand and is not connected.[1]
- The Technical Barrier of Thermal Paper: Thermal paper receipts, the main source of expense data, pose a special difficulty. Chemically unstable, thermal paper is prone to fading quickly, friction damage, and "thermal noise." Important cost data is sometimes left in an unstructured, analogue state when standard Optical Character Recognition (OCR) systems are unable to process this deteriorated picture.[2]
- Data Isolation & Information Asymmetry: This results in a "Data Silo" where business owners have access to real-time sales data but only delayed, retrospective expense data. As of right now, there is no automatic method for calculating live profit margins by correlating real-time sales records with ingredient costs.
- Inadequacy of Current Tools: This particular gap is not addressed by current software solutions. Micro-enterprises cannot afford Enterprise Resource Planning (ERP) systems, and simple expense tracking programs lack the semantic intelligence needed to normalise complicated ingredient data (e.g., mapping "Sugr" to "Sugar" or converting units).[3]
- Reactive vs. Proactive Decision Making: MSMEs rely on reactive decision-making in the absence of digitised historical cost data. They are unable to predict supply chain volatility. This results in "profit erosion," whereby growing ingredient costs gradually reduce profit margins before entrepreneurs can modify their pricing tactics.[4]

References

1. 1] OECD (2021). The Digital Transformation of SMEs. OECD Studies on SMEs and Entrepreneurship.
2. [2] Smith, R. & Patel, K. (2019). Challenges in Optical Character Recognition for Degraded Historical and Thermal Documents. Journal of Document Analysis and Recognition, 12(4), 45-58.
3. [3] Wong, L. (2022). Financial Literacy and Technology Adoption in Micro-Enterprises. International Journal of Small Business Management, 15(2), 112-129.
5. [4] Kumar, A. (2020). Supply Chain Volatility and its Impact on SME Profitability. Journal of Business Economics, 8(1), 22-30

6. Brief description of Existing Research and Systems (300 words max)

Summary of Existing Products & Systems

- General Accounting Software: Platforms like QuickBooks and Xero are the industry standard for financial compliance. However, they record expenses as aggregate totals (e.g., "Supplier A: \$50") rather than granular line items. They lack the capability to perform recipe-level costing or automated inventory tracking for micro-enterprises.
- Receipt Scanning Applications: Tools like Expensify and Dext excel at digitizing clear office documents. However, they rely on general-purpose OCR APIs (like Google Vision), which struggle with the low contrast and "thermal noise" inherent in supermarket receipts. They also lack the domain logic to normalize units (e.g., converting "lbs" to "kg").
- Point-of-Sale (POS) Systems: Modern POS systems (e.g., Toast, Square) effectively digitize Revenue. However, they act as data silos. They cannot automatically ingest external purchase data, requiring manual data entry to update cost prices.

Summary of Existing Research

- OCR Limitations: Academic literature confirms that standard OCR engines suffer significant performance degradation when processing thermal paper due to its chemical instability, friction marks, and rapid fading [5]. Research highlights the need for specialized pre-processing algorithms (like adaptive thresholding) which are currently absent in consumer apps.
- SME Technology Adoption: Studies indicate a "Digital Divide" where predictive analytics and demand forecasting tools (standard in Enterprise ERPs like SAP) are too complex and expensive for micro-SMEs [6].

Gaps & Limitations Identified

- The Integration Gap: There is no existing system that correlates real-time manual sales data with automated receipt-based cost data to calculate dynamic profit margins.
- The Predictive Gap: Current tools are reactive (recording past history) rather than proactive. There is a lack of "Lightweight AI" solutions capable of forecasting supply chain volatility for small businesses using limited datasets.

7. Brief description of the solution's nature, including a conceptual diagram (maximum 500 words).

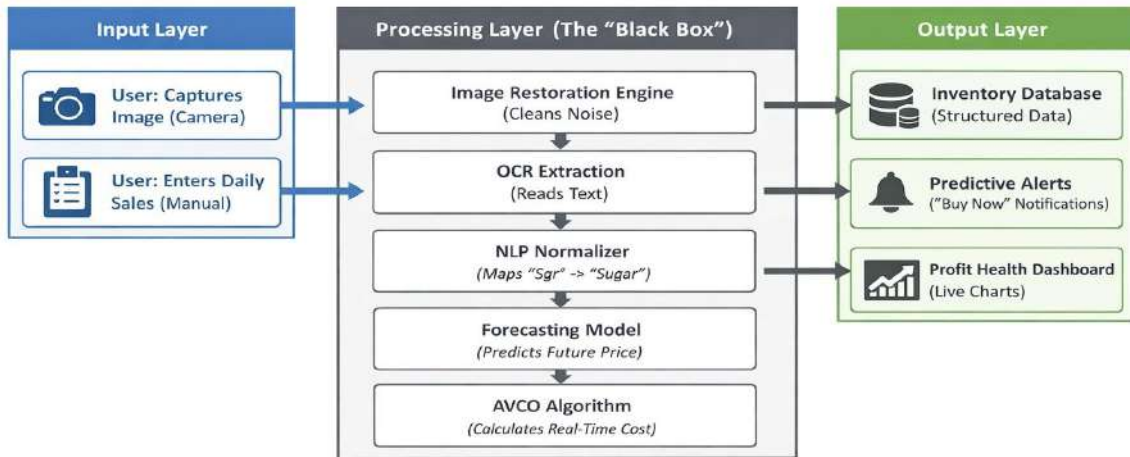
Addressing the Identified Gaps **SmartCost** is a mobile-first Intelligent Decision Support System (IDSS) designed to bridge the "Integration" and "Predictive" gaps identified in existing research. Unlike general accounting tools that passively record totals, this solution actively automates the extraction of granular line-item data and correlates it with sales. By introducing Adaptive Thermal Image Restoration, the system solves the technical failure of standard OCR on degraded paper. Furthermore, by democratizing Time-Series Forecasting, it provides micro-enterprises with the proactive "Future Insight" previously reserved for enterprise ERPs.

System Perspective: Input, Process, Output

The solution operates as a pipeline with three distinct stages:

1. **Input Layer (Data Acquisition):** The system accepts two primary data streams. First, the Visual Input, where the user captures images of degraded thermal receipts via the mobile camera. Second, the Manual Input, where the user enters daily revenue figures (from their POS Z-Report) and specific category budget limits.
2. **Processing Layer (The Intelligence Engine):** This "Black Box" consists of four interconnected research modules:
 - **Preprocessing:** The Restoration Engine (Novelty 1) applies adaptive thresholding and morphological operations to clean thermal noise and de-warp the image.
 - **Extraction & Normalization:** The NLP Engine (Novelty 2) extracts text and uses Fuzzy Logic to map unstructured abbreviations (e.g., "Flr") to standardized inventory items, normalizing units to a common standard (kg/L).
 - **Forecasting:** The Predictive Model (Novelty 3) analyzes historical price trends using ARIMA/Prophet algorithms to forecast future unit costs.
 - **Calculation:** The Dynamic Profit Engine (Novelty 4) utilizes a Weighted Average Cost (AVCO) algorithm to update inventory values in real-time.
3. **Output Layer (Actionable Intelligence):** The processed data is presented as actionable insights:
 - **Inventory Database:** A structured, searchable record of all ingredients and costs.
 - **Predictive Alerts:** Notifications warning of predicted price hikes (e.g., "Sugar prices rising next week").
 - **Financial Dashboard:** A real-time visualization of Gross Profit Margins and "Health Scores," alerting the user if profitability drops below safety thresholds.

Conceptual Diagram



8. Brief overview of the availability of necessary specialized domain expertise, knowledge, and data. (500 words max)

Acquiring Specialized Domain Expertise The project focuses on the niche domain of "Micro-SME Supply Chain Management" and "Managerial Accounting." To bridge the gap between technical development and business reality, the team will adopt a two-pronged approach.

- **Stakeholder Engagement:** The team will conduct semi-structured interviews with 3-5 local bakery and café owners. These sessions will focus on understanding their current purchasing behaviors, inventory "pain points," and manual accounting workflows. This qualitative data ensures the "Dynamic Profit Engine" (Novelty 4) aligns with real-world business logic.
- **Academic Study:** The team will consult literature on Cost Accounting Standards (CAS), specifically focusing on inventory valuation methods like Weighted Average Cost (AVCO). This ensures the financial formulas implemented in the application are mathematically accurate and compliant with standard business practices.

Acquiring Technical Knowledge and Skills Since the project involves advanced research components (Computer Vision, NLP, Forecasting), the team will rely on self-directed learning and academic resources:

- **Member 1 (Image Restoration):** Will utilize OpenCV documentation and research papers on "Document Image Binarization" (e.g., Sauvola's method) to master adaptive thresholding techniques.
- **Member 2 (NLP):** Will undertake coursework on Natural Language Processing via platforms like Coursera, specifically focusing on Named Entity Recognition (NER) and Fuzzy String Matching libraries (e.g., RapidFuzz).
- **Member 3 (Predictive Analytics):** Will study the documentation of Facebook Prophet and ARIMA models to understand time-series decomposition.
- **Member 4 (App Development):** Will leverage advanced React Native resources to master local database integration (WatermelonDB) and state management for real-time calculations.

Data Collection Strategy The project requires three distinct datasets:

- **Dataset A (Visual - For Novelty 1):** The team will physically collect 3000 - 5000 distinct thermal receipts from various local retailers (supermarkets, bakeries). These will be subjected to artificial degradation (crumpling, heating, fading) to create a robust "Noisy Dataset" for testing the restoration algorithm.
- **Dataset B (Semantic - For Novelty 2):** A "Master Ingredient Dictionary" will be curated by scraping open-source grocery datasets and manually enriching them with common bakery abbreviations (e.g., mapping "Ap Flr" to "All Purpose Flour").
- **Dataset C (Historical - For Novelty 3):** To tackle the "Cold Start" problem, the team will utilize publicly available historical price indexes for commodities (sugar, flour, oil) from government statistical portals or Kaggle to pre-train the forecasting model.

Ethical Clearance Yes, ethical clearance is required. Although the system does not track sensitive personal biometric data, it processes **Proprietary Financial Data** (receipts and sales figures) from business owners during the testing phase. The team will adhere to strict data privacy protocols:

- **Anonymization:** Any receipts used for system training will have the Business Name and Tax ID redacted.
- **Consent:** Informed consent forms will be obtained from all participants involved in the interview or beta-testing phases, clearly stating that their data is used solely for academic research purposes.

9. Objectives and Novelty

Main Objective To develop a mobile-first Intelligent Decision Support System (IDSS) for micro-enterprises that automates the digitization of degraded thermal receipts, normalizes unstructured ingredient data, predicts supply chain price volatility, and calculates real-time dynamic profit margins to enable proactive financial decision-making.			
Member Name with Registration No	Sub Objective	Tasks	Novelty
Kodithuwakku K.A.C.H IT22544872	To develop an adaptive image pre-processing engine capable of restoring and digitizing highly degraded, noisy, and distorted thermal paper receipts.	1. Implement adaptive thresholding algorithms to separate text from "thermal noise" (black spots). 2. Develop morphological operations to repair broken character strokes caused by fading. 3. Implement geometric de-warping to flatten images of crumpled receipts. 4. Integrate Tesseract OCR for final text extraction.	Unlike standard OCR tools (e.g., Google Lens) which assume high-quality input, this research focuses specifically on reconstructing degraded thermal paper. The novelty lies in the custom pre-processing pipeline that removes specific "thermal noise" artifacts before extraction.

Perera N.K.M IT22151810	To build a semantic normalization engine that converts unstructured, inconsistent ingredient text into a standardized, structured inventory database.	1. Develop a domain-specific "Bakery Dictionary" for Named Entity Recognition (NER). 2. Implement Fuzzy Logic algorithms (Levenshtein Distance) to map abbreviations (e.g., "Flr") to master items. 3. Create a unit conversion module to normalize diverse weights (g, kg, lbs) into a standard metric. 4. Design the database schema for structured inventory storage.	Existing apps store receipt text as raw strings. This component contributes a Context-Aware Normalization Engine that uses domain-specific fuzzy logic to "understand" and standardize bakery ingredients, enabling accurate inventory tracking from messy data.
Jayasingha D.D.M IT22162250	To develop a predictive analytics model that forecasts future ingredient prices and identifies supply chain volatility trends using historical purchase data.	1. Collect and preprocess historical price datasets for training. 2. Implement Time-Series Forecasting models (ARIMA / Facebook Prophet) to decompose trends and seasonality. 3. Develop an API endpoint to serve price predictions to the mobile app. 4. Implement a "Confidence Interval" system to generate volatility alerts.	While large enterprises use expensive ERPs for forecasting, this research democratizes Time-Series Forecasting for micro-businesses. The novelty is applying lightweight statistical models to small, noisy datasets to provide "Future Insight" on ingredient costs.

Dasanayaka M.P.J.R IT22152664	To design and implement the mobile application interface and a dynamic profit engine that correlates real-time costs with sales data to calculate live margins.	1. Develop the React Native mobile interface (Dashboard, Camera, Charts). 2. Implement the Weighted Average Cost (AVCO) algorithm to calculate real-time COGS. 3. Build the "Profit Health" logic to trigger alerts based on margin thresholds. 4. Integrate all backend API services (Vision, NLP, Prediction) into the mobile app.	Most POS systems track sales, and accounting apps track expenses, but they remain disconnected. The novelty here is the Real-Time Integration of both using the AVCO algorithm, allowing the system to recalculate profit margins instantly with every receipt scanned, rather than at the end of the month.
-------------------------------------	---	---	---

**10. This part is to be filled by the
Supervisor and the Co-supervisor of the Project.**

- a) This research topic possesses a comprehensive scope suitable for a final-year project.

Yes	☒	No	☐
-----	-------------------------------------	----	--------------------------

- b) The proposed topic exhibits novelty.

Yes	☒	No	☐
-----	-------------------------------------	----	--------------------------

- c) The student group can successfully execute the proposed project.

Yes	☒	No	☐
-----	-------------------------------------	----	--------------------------

- d) The proposed sub-objectives reflect the students' areas of specialization.

Yes	☒	No	☐
-----	-------------------------------------	----	--------------------------

- e) Any other comments:

*Ethical clearance should be obtained if customer data is collected to create the dataset for model training.
Good to proceed with.*

	Title	First Name	Last Name	Signature
Supervisor	Mr.	Deemantha	Sinwardana	
Co-Supervisor	Ms.	Ayesha	Wijesooriya	
External Supervisor				
Summary of external supervisor's (if any) experience and expertise				