From Data to Decisions: Pre-Fusion, Post-Fusion, and Hybrid Models (LLM) in BFSI Enterprise Systems

Dr. Gopichand Agnihotram, DMTS - Senior Principal Member & Director, Wipro Limited
Venkatesh Balasubramaniam, Partner, Head of BFSI Consulting, Wipro Limited
Joydeep Sarkar, DMTS - Principal Member & Senior Architect, Wipro Limited
Dr. Magesh Kasthuri, Chief Architect & Distinguished Member of Technical Staff, Wipro Limited

Abstract

Multimodal fusion, the integration of heterogeneous data types such as text, images, audio, structured records, and sensor feed has become a foundational capability for advanced analytics and intelligent decision-making in the Banking, Financial Services, and Insurance (BFSI) sector. In an era where fraud tactics evolve daily, compliance regulations tighten, and customer expectations rise, the ability to unify signals from diverse modalities offers a decisive competitive and operational advantage.

Two primary paradigms dominate the design landscape, Pre-Fusion (Early Fusion) and Post-Fusion (Late Fusion**)**. Pre-Fusion merges multiple data modalities at the feature representation stage, enabling a single predictive model to learn complex cross-modal relationships before generating outputs. Post-Fusion, in contrast, processes each modality independently, generating separate outputs that are subsequently combined at a decision layer. Each strategy presents distinct trade-offs in terms of accuracy, robustness, computational efficiency, interpretability, and resilience factors that are especially critical in BFSI's high-stakes environment.

This paper provides an in depth examination of the conceptual foundations, operational mechanics, and business implications of both fusion paradigms. It analyses BFSI-specific scenarios, outlines key design considerations for deploying multimodal fusion in mission-critical systems, and explores emerging directions such as adaptive fusion strategies and agentic AI orchestration. By the conclusion, BFSI stakeholders including data scientists, compliance officers, risk managers, and technology leaders will gain a structured framework for selecting, implementing, and optimizing fusion strategies that align with both business goals and regulatory mandates.

Keywords : LLM, BFSI, Artificial Intelligence, Multi-model fusion, data science, data qualification

2. Introduction

In today's data driven BFSI ecosystem, decision making is increasingly dependent on synthesizing diverse streams of information. Customer verification may require matching a scanned government ID with a selfie, transaction history, and location metadata. Fraud detection models need to weigh real time behavioural analytics alongside historical spending patterns, call centre transcripts, and biometric cues. Credit scoring often benefits from both structured financial data and unstructured customer narratives, such as support interactions and social sentiment. This interplay between different data sources is at the heart of multimodal fusion[1].

Multimodal fusion is not a single method but rather a family of strategies, each with trade-offs in terms of accuracy, interpretability, computational demands, and adaptability. Two dominant paradigms have emerged:

1. **Pre-Fusion (Early Fusion)**: where modalities are combined into a unified representation before modelling.
2. **Post-Fusion (Late Fusion)**: where each modality is processed independently and combined at the decision stage.

The choice between these paradigms is not merely technical, it reflects deep operational priorities. For example, a fraud detection team seeking maximum sensitivity to complex patterns may lean toward pre-fusion, while a compliance team needing explainable, auditable decision trails may prefer post-fusion.

This paper will explore both paradigms in depth, analysing how they function, where they excel, where they fall short, and how BFSI organizations can strategically deploy them to meet both business and regulatory goals.

2. Pre-Fusion (Early Fusion)

2.1 Concept

Pre-fusion, often referred to as early fusion, is a multimodal integration approach in which data from multiple sources is combined into a single, unified representation before being passed into a predictive model. This means that textual, visual, numerical, or other modality-specific features are merged at the feature level, enabling the model to learn cross-modal relationships directly during the training process.

In enterprise scale BFSI systems, pre-fusion is particularly valuable when there is a need for the model to reason about relationships between diverse data streams in real time. Its ability to capture nuanced, cross modal interactions can lead to higher predictive accuracy in scenarios such as fraud prevention, anti-money laundering (AML) investigations, and risk scoring. However, the approach also requires significant preprocessing and alignment of modalities to ensure compatibility, which can add complexity to large-scale deployment.

2.2 Workflow

In BFSI contexts, the pre-fusion workflow typically begins with data acquisition, where information is collected from multiple sources and modalities[2]. This might include structured transaction logs, unstructured text such as customer support tickets or complaint narratives, and additional formats such as scanned documents or facial images from identity verification processes. Each of these data types captures a different perspective on the same operational or customer activity.

Once the raw data has been gathered, the next step is feature extraction. In this stage, each modality is transformed into a numerical format suitable for computational analysis. For example, textual data may be converted into word embeddings that capture semantic meaning, images may be processed through convolutional neural networks (CNNs) to extract visual features, and numerical transaction attributes may be normalized to ensure consistent scale and comparability.

Following extraction, a feature alignment phase ensures that the information from different modalities is synchronized both temporally and logically. This alignment is crucial so that, for instance, a customer's biometric verification image, the associated transaction data, and a relevant support ticket all correspond to the same individual and the same event window. Without this alignment, the fusion process could introduce noise or misleading associations.

The core fusion step then merges the aligned features from all modalities into a single, integrated representation. This unified dataset enables the downstream model to interpret patterns and relationships that emerge from the combined view rather than isolated data streams.

In the model training phase, this fused representation is fed into a predictive system such as a deep learning model or an ensemble method that can exploit cross-modal patterns to improve decision accuracy. By learning from integrated data, the model becomes capable of detecting complex scenarios, such as coordinated fraud attempts that involve simultaneous anomalies across multiple data types.

Finally, during inference and decision-making, the trained model processes incoming multimodal data in real time, generating actionable outputs. These outputs might include fraud alerts, credit risk scores, compliance flags, or other operational decisions. By following this structured workflow, BFSI organizations can harness the full power of pre-fusion to enhance accuracy, responsiveness, and resilience in their high-stakes decision-making processes.

3. Post-Fusion (Late Fusion)

3.1 Concept

Post-fusion, also known as late fusion or decision-level fusion, adopts a very different philosophy compared to pre-fusion. Instead of combining raw or intermediate features from multiple data sources at the outset, post-fusion treats each modality as an independent analysis pipeline. Each modality whether it is structured numerical data, unstructured text, images, or audio is processed separately using a dedicated model or sub-model that is best suited to that data type. These modality-specific models operate in parallel, each generating its own prediction, classification score, or probability estimate[3].

Once these independent predictions are obtained, they are brought together at a higher decision-making layer. This final fusion step can be implemented in a variety of ways, ranging from simple methods like averaging the scores or applying majority voting, to more sophisticated strategies like weighted combination based on model reliability, rule-based business logic, or even a secondary "meta-model" trained specifically to combine the outputs of the modality specific models.

In a BFSI fraud detection scenario, for example, a transactional analytics model might evaluate structured transaction logs and generate a fraud risk score based on behavioural anomalies in spending patterns. A separate text analytics model could process customer service chat logs to detect suspicious or deceptive language patterns. Meanwhile, a facial recognition system might verify biometric matches during an online account access attempt. Rather than merging these raw features earlier in the pipeline, post-fusion waits until each model has independently formed its own assessment. The decision layer then integrates these diverse signals to arrive at a single fraud probability or classification outcome.

This separation of processing pathways offers several advantages, including the flexibility to optimize each modality's analysis independently, the ability to handle missing modalities more gracefully, and the potential for greater transparency when explaining how each input contributed to the final decision. However, it also means that some subtle cross-modal relationships those that could have been discovered through joint feature learning in pre-fusion may be missed.

3.2 BFSI Use Cases

In the context of credit risk assessment, post-fusion enables financial institutions to maintain separate models for distinct data sources, such as credit bureau reports, customer income statements, and reputational signals derived from social media analysis. Each model produces its own independent score, reflecting its specialized evaluation of the customer's risk profile. These outputs are then combined to produce a final risk score, but importantly, the individual component scores are preserved. This approach ensures that, during regulatory audits, institutions can provide a transparent breakdown of the decision-making process, showing precisely how each data source contributed to the final assessment.

For regulatory compliance monitoring, post-fusion offers the advantage of integrating highly specialized detection mechanisms without diluting their individual strengths. An NLP model might be dedicated to analysing unstructured customer communications to detect suspicious language patterns, while a rules-based engine monitors transaction amounts and frequencies against compliance thresholds. In parallel, an AI-driven anomaly detection system can process time-series transaction data to uncover irregular patterns. The outputs from these independent systems are fused at the compliance dashboard level, providing compliance officers with a holistic view of potential risks while still allowing them to trace alerts back to their source models for detailed investigation.

In insurance claim validation, post-fusion supports a multi-pronged approach to assessing the legitimacy of claims. An image analysis model may evaluate damage inspection photographs to estimate the extent and authenticity of reported damage. Simultaneously, a text-based classifier can process the claim report narrative to detect inconsistencies or unusual language patterns, while a structured data risk model examines the claimant's historical behaviour, such as the frequency and nature of past claims. The decision fusion step then combines these perspectives, ensuring that no single modality dominates the approval or rejection decision. This approach reduces bias, increases fairness, and improves confidence in the final outcome.

3.4 Advantages

One of the key advantages of post-fusion is its interpretability. By maintaining separate decision paths for each modality, the system allows auditors, regulators, and internal reviewers to clearly trace how each data source contributed to the final outcome. This transparency is particularly valuable in the BFSI sector, where regulatory frameworks demand that institutions explain not just the decision itself, but also the reasoning and evidence behind it.

Another strength is modularity. Because each modality is processed through its own dedicated model, individual components can be updated, refined, or replaced without retraining the entire system. For example, if a new fraud detection algorithm outperforms the current model for transaction monitoring, it can be swapped in without affecting the text analytics or biometric verification modules. This modular design significantly reduces system maintenance complexity and accelerates the adoption of emerging technologies.

Post-fusion is also inherently robust to missing data. In real-world BFSI operations, it is common for one or more data sources to be unavailable perhaps a biometric scan fails, a document is corrupted, or a customer chooses not to provide certain optional information. With post-fusion, the absence of one modality does not halt the decision-making process; the system can still combine the available modality outputs to produce a reasoned and actionable result.

Finally, post-fusion offers high flexibility in terms of algorithm selection. Because each modality is handled independently, data scientists can choose the most suitable algorithm for each type of input without forcing them into a shared feature representation. This means that image data can be processed with convolutional neural networks, text with transformer-based language models, and structured transaction data with gradient boosting all within the same integrated decision pipeline. This flexibility ensures that each data type is handled using the most effective modelling approach available.

3.5 Challenges

One limitation of post-fusion is its weaker ability to capture cross-modal interactions. Because each modality is processed in isolation, the system may fail to detect subtle correlations that span different data types. For example, in a fraud detection scenario, a minor anomaly in transaction history might not seem suspicious on its own, and an unusual phrase in a customer support chat could also appear benign in isolation. However, when these two signals occur together, they may indicate coordinated fraudulent activity—a connection that is easier to detect in pre-fusion approaches than in post-fusion pipelines[4].

Another challenge is the complexity of designing the fusion mechanism itself. Determining the optimal way to combine modality-specific predictions—whether through simple rules, weighted averaging, or meta-models—requires careful experimentation and tuning. This process becomes even more difficult when modalities vary greatly in predictive power, reliability, or data availability. The fusion logic must account for these differences without inadvertently over- or under-weighting critical signals.

Post-fusion systems can also introduce potential redundancy. Because each modality is modelled independently, different models may end up learning overlapping predictive patterns, which can waste computational resources. For example, both a transaction-based model and a behavioural biometrics model might indirectly capture the same "customer activity velocity" feature, duplicating effort and contributing little incremental value to the final decision.

Finally, post-fusion architectures may face latency concerns, especially in real-time BFSI applications. Running multiple full models in sequence can slow down inference times if the system is not optimized for parallel execution. This can be problematic in high-stakes scenarios such as instant fraud detection or rapid credit approval, where delays of even a few hundred milliseconds can degrade user experience or operational efficiency. Careful system engineering is needed to mitigate these performance issues without sacrificing decision quality.

4. Hybrid Possibilities

In practical BFSI implementations, one promising approach is **stage-wise fusion**, where pre-fusion is applied within closely related modality families, such as different forms of structured financial data, and post-fusion is then used to combine outputs across distinct modality families. For example, structured transaction records, account balances, and payment histories might be fused early to form a unified financial behavior profile, while outputs from this structured data model are later combined with independent models for text-based communications or document analysis at the decision stage. This layered method allows institutions to leverage the benefits of deep cross-modal integration without losing the interpretability and modularity of late fusion.

Another emerging practice is **adaptive fusion**, which involves dynamically selecting the fusion strategy based on real-time data availability and quality during inference. In this setup, if one modality arrives late, contains errors, or is partially missing, the system can automatically fall back to a late-fusion approach for that decision cycle. Conversely, when all modalities are present and clean, early fusion may be used to maximize the depth of cross-modal learning. This adaptability ensures decision-making continuity even in imperfect operational environments—critical for high-stakes domains like fraud detection or credit approvals.

A third innovation gaining traction is **weighted confidence fusion**, where real-time confidence scores and recent reliability metrics determine how much influence each modality should have in the final decision. For instance, if an OCR model for document analysis has recently encountered noisy inputs leading to lower confidence, its contribution to the decision score can be temporarily reduced while more reliable modalities take precedence. This weighting mechanism not only enhances resilience but also aligns decision-making with the fluctuating trustworthiness of individual data streams, creating a more robust and context-aware system.

5. Future Directions in Multimodal Fusion for BFSI

The evolution of multimodal fusion in the BFSI sector will be shaped by advances in AI, regulatory trends, and market pressures. Over the next five years, we expect three core themes to define the future: adaptivity, explainability, and agent-based orchestration.

5.1 Self-Adaptive Fusion Architectures

Traditional fusion designs whether pre- or post-fusion are typically static. The modality weighting, feature extraction techniques, and decision aggregation logic remain fixed until retrained.
Emerging research in **self-adaptive fusion** seeks to create systems that:

- Dynamically reweight modalities based on their **real-time reliability**.
- Automatically switch between pre-fusion and post-fusion modes depending on **data completeness and time constraints**.
- Leverage reinforcement learning (RL) to optimize fusion strategies for evolving fraud patterns or compliance requirements.

In BFSI, this could mean that if **OCR quality drops** for scanned documents due to poor lighting, the system temporarily reduces the weight of document-based features and increases reliance on behavioural data.

5.2 Causal-Aware Fusion

Next-generation models are beginning to integrate **causal inference techniques** into fusion pipelines. This allows systems to distinguish correlation from causation across modalities. For instance, a spike in mobile banking app logins (modality 1) and a rise in suspicious transactions (modality 2) may be correlated, but causal modelling could reveal whether the former drives the latter or if both stem from a third factor (e.g., a marketing campaign).

5.3 Explainable Multimodal Fusion (XMF)

As regulatory oversight becomes more stringent, the ability to explain model outputs will be just as important as achieving high accuracy. One key innovation in explainable multimodal fusion is the ability to trace decision pathways through each modality. This means that every data source—whether structured transactions, textual communications, or biometric inputs—can have its influence explicitly documented, enabling both internal teams and regulators to understand exactly how each contributed to the final outcome.

Another important advancement is the provision of per-modality confidence scores. These scores quantify the reliability of each modality's contribution, offering transparency to both customers and oversight bodies. For example, a credit approval decision could be accompanied by separate confidence levels for financial history, document verification, and behavioural analytics, making the decision-making process more accountable.

A further development is the introduction of interactive dashboards for compliance officers. These tools allow users to explore "what-if" scenarios, such as assessing how an outcome might change if biometric data were excluded or given less weight. By providing real-time, scenario-based insights, these dashboards empower risk and compliance teams to evaluate the fairness, robustness, and regulatory soundness of multimodal decision systems.

5.4 Agent-Based Fusion Orchestration

With the rise of **agentic AI architectures**, fusion strategies can become **context-aware**:

- A **Data Quality Agent** could assess modality reliability before fusion.
- A **Compliance Agent** could veto the use of certain modalities for jurisdictions with stricter privacy rules.
- A **Performance Optimization Agent** could reroute workloads to faster pre-fusion models during peak transaction loads.

This approach aligns with BFSI's increasing interest in **multi-agent governance frameworks** for AI.

7.5 Edge-Enabled Fusion for Real-Time Decisioning

Mobile banking, contactless payments, and ATM authentication are pushing fusion models closer to the edge:

- Running lightweight pre-fusion models directly on ATMs for sub-second identity verification.
- Using on-device biometric fusion for mobile payments, with **privacy-preserving techniques** such as federated learning.

6. Conclusion

Multimodal fusion is no longer a niche innovation, it is becoming a core enabler of advanced decision-making in BFSI. Whether for fraud detection, credit scoring, customer onboarding, or compliance auditing, the ability to merge signals from text, images, structured data, and behavioural patterns is rapidly differentiating market leaders from laggards.

7. References

[1] Vaccaro, A., et al. "An adaptive framework based on multi-model data fusion for one-day-ahead wind power forecasting." *Electric Power Systems Research* 81.3 (2011): 775-782.
[2] Zamani, Mohammad G., et al. "A multi-model data fusion methodology for reservoir water quality based on machine learning algorithms and bayesian maximum entropy." *Journal of Cleaner Production* 416 (2023): 137885.
[3] See, Linda, and Robert J. Abrahart. "Multi-model data fusion for hydrological forecasting." *Computers & geosciences* 27.8 (2001): 987-994.
[4] See, Linda, and Robert J. Abrahart. "Multi-model data fusion for hydrological forecasting." *Computers & geosciences* 27.8 (2001): 987-994.