A Comparitive study on Pre-Fusion, Post-Fusion, and Hybrid Models (LLM) for Data analysis

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 refers to the process of integrating data from multiple modalities such as text, images, audio, and structured numerical data into a single decision making pipeline. The goal is to combine complementary strengths of each modality to achieve improved accuracy, robustness, and contextual understanding compared to relying on a single source.

This paper discusses the real-time usecases of applying Multi-modal fusion in data science for decision-making using Pre-Fusion or Post-fusion techniques. There is a hybrid model also possible to combine the advantages of both these techniques to create an amalgamation of multi-modal services.

1. Introduction

In large data handling sectors like Banking, Financial Services, Securities and Capital Markets and Insurance (shortly called as BFSI sector) usecases like fraud detection, Pre-fusion a.k.a Early Fusion[1] can identify subtle patterns by correlating transaction metadata, biometric verification data, and customer communication logs in a unified model. Conversely, in credit scoring and regulatory compliance, Late Fusion a.k.a. Post-fusion[2] can preserve modality-specific reasoning maintaining separate decision logic for structured credit history, unstructured customer narratives, and supplementary documentation thus enhancing transparency for audits and regulatory review.

In the BFSI sector, the stakes are high.

- **Regulatory Compliance**: Financial institutions are subject to stringent regulations such as KYC (Know Your Customer), AML (Anti-Money Laundering), and GDPR. These often require verifying customer identity and activity by cross-checking multiple data sources. Fusion enables systems to combine document scans, facial recognition results, and transaction logs for more reliable compliance checks.
- **Fraud Detection**: Fraudulent activities often leave subtle, multi-faceted traces. A suspicious transaction may look legitimate in isolation but reveal anomalies when correlated with unusual geolocation data, device fingerprint mismatches, or recent customer complaints.
- **Credit Scoring**: Traditional scoring models often rely solely on structured credit history. However, integrating qualitative data such as customer explanations, call transcripts, or alternative payment histories can yield a more complete and fair assessment.

2. Pre-fusion

The central principle of pre-fusion is the creation of a shared representation space in which all modalities coexist. This unified space allows the model to identify subtle and potentially critical correlations that might remain hidden if each data source were processed in isolation. For example, in a BFSI fraud detection context, a sudden change in a customer's transaction location becomes far more meaningful when considered alongside a change in login device type and an unusual surge in password reset requests. These disparate signals, when combined, may reveal a coordinated pattern indicative of fraudulent activity something that might be overlooked if each signal were analysed separately.

To achieve this integration, pre-fusion techniques often rely on merging features through methods such as concatenation, weighted combinations, or advanced mechanisms like attention based fusion. Once merged, the resulting feature representation is processed by a downstream model such as a neural network, gradient boosting machine, or a hybrid architecture that learns patterns from the integrated signal set to make predictions.

2.1 Advantages

One of the most significant strengths of the pre-fusion approach lies in its ability to support **rich feature interactions**. Because data from different modalities is combined into a single unified representation early in the processing pipeline, the model has the opportunity to identify and learn from complex, interdependent relationships between these inputs. For example, in a BFSI fraud detection setting, a model can jointly analyse the correlation between an unusual transaction location, the device's browser fingerprint, and sentiment patterns extracted from recent customer service interactions. This early exposure to multimodal relationships allows the system to capture dependencies that would be far less apparent if the modalities were processed separately and merged later in the workflow.

Another advantage is the potential for **higher predictive accuracy**. When modalities are fused at the feature level, the model can leverage subtle cross-modal cues that might otherwise be overlooked. Consider a credit scoring scenario where an applicant's income history, employment documents, and personal loan request narrative are all considered together during model training. A slight inconsistency between the applicant's self-reported job title and the designation stated in their scanned employment letter may be flagged because the model has learned to identify such patterns across modalities. By capturing these fine-grained interactions, early fusion can enhance the model's decision quality, particularly in complex or high-risk use cases.

A further benefit is **end-to-end optimization**. In pre-fusion systems, the training process covers the entire feature set in a single unified model, avoiding the need to train multiple modality-specific models and then combine them later. This unified approach allows joint learning, where adjustments to the model's parameters can improve performance across all modalities simultaneously. In operational terms, this not only streamlines the development process but also reduces the risk of error propagation between independent models. For BFSI institutions managing critical systems like compliance checks or loan risk scoring, such optimization can lead to faster deployment cycles and more consistent model performance across varied data environments.

2.2 BFSI Use Cases

One of the most prominent applications of pre-fusion in the BFSI sector is **fraud detection through multimodal correlation**. In this scenario, data from device metadata, geolocation patterns, transaction histories, and customer service call logs are brought together into a single, unified feature representation. By merging these signals at the feature level, an early fusion model can uncover subtle, cross-modal patterns that would be invisible if each stream were analysed independently. For example, a single unusual login location might not be flagged as suspicious on its own, and a slightly altered device identifier could seem innocuous in isolation. However, when these indicators are considered together along with anomalies in transaction frequency and tone changes in recent customer calls they may reveal a coordinated fraud attempt that would otherwise go undetected.

Another critical area is **customer onboarding verification**, where financial institutions need to confirm identity with speed and accuracy while minimizing friction for legitimate customers. Pre-fusion enables the combination of data derived from optical character recognition (OCR) applied to submitted documents, biometric embeddings extracted from identity verification photos or videos, and structured Know Your Customer (KYC) form data. By evaluating these inputs simultaneously, the model can identify discrepancies between the name extracted from a document scan, the biometric match score against a reference image, and the demographic details entered in the onboarding form. This simultaneous assessment increases accuracy, helps prevent identity fraud, and allows genuine customers to complete onboarding more quickly.

A further example is in **loan approval processes**, where the decision-making model can integrate structured numerical data such as credit scores, visual or textual evidence from scanned income proof documents, and qualitative self-reported narratives about employment or business activities. The early fusion approach allows these different types of information to be processed in concert, enabling the model to detect inconsistencies between the income figures stated in financial statements and the claims made in accompanying text narratives. For instance, if an applicant declares steady employment in their written statement but the scanned payslips show irregular payment intervals, the model can flag this as a potential risk factor. By capturing these nuanced mismatches early, pre-fusion helps reduce approval errors and supports more responsible lending practices.

2.3 Challenges

While pre-fusion offers powerful advantages, it also presents notable challenges, particularly in the BFSI context where operational precision and regulatory compliance are paramount. One of the primary hurdles is **data alignment complexity**[3]. Because modalities are merged at the feature level, their inputs must be carefully synchronized to ensure they refer to the same entity, event, or time frame. In practice, this can be non-trivial—especially when dealing with real-time transaction data, asynchronous customer communications, and irregular document submission schedules. Any misalignment in timestamps, identifiers, or context can lead to noisy or misleading fused features, ultimately reducing model reliability.

Another challenge lies in **handling missing data**. In many real-world BFSI scenarios, it is common for one or more modalities to be incomplete or unavailable. For example, a fraud detection system might have transaction metadata and device information, but lack recent biometric verification data due to customer opt-out or technical failure. In a pre-fusion setup, the absence of this modality can make the fused feature vector incomplete, potentially degrading model performance unless robust techniques such as imputation, modality dropout, or learned masking strategies are applied. Designing these mechanisms adds further complexity to the engineering pipeline.

Pre-fusion can also lead to **computational overhead**. Because multiple high-dimensional feature sets are concatenated or projected into a joint embedding space, the resulting fused vectors can be extremely large. This increases the computational burden for both training and inference, leading to higher memory usage and potentially slower response times. In high-volume BFSI environments—such as real-time fraud monitoring or instant credit decisioning—this can become a critical operational constraint, especially when deploying on edge devices or resource-limited systems.

Finally, pre-fusion introduces **reduced interpretability**, a factor of particular concern in regulated sectors. When modalities are deeply entangled within a unified representation, it becomes much harder to determine which input contributed most to a specific prediction or alert. For instance, if a loan application is denied, the model's reasoning may involve subtle interactions between income documents, transaction histories, and customer narratives that cannot easily be disentangled. This opacity poses significant challenges for compliance teams who must provide clear, auditable justifications for automated decisions under regulations such as the EU's AI Act or the U.S. Fair Credit Reporting Act.

3. Post-fusion

In a BFSI post-fusion workflow, the process begins with modality separation. Rather than blending information at the outset, different data streams—such as structured transaction logs, unstructured text from customer support chats, and scanned documents—are deliberately kept independent during the early stages of processing. This separation ensures that each data source can be analysed with techniques best suited to its unique characteristics without interference from other modalities.

The next stage is independent model training. Here, specialized models are built for each modality, each optimized for the type of data it handles. For example, a gradient boosting model might be trained on numerical transaction features, a transformer-based language model could analyse textual communication patterns, and a convolutional neural network might process visual data from identity documents. This step ensures that the analytical approach is tailored to the strengths of each data type, thereby improving the accuracy of individual predictions.

Once the models are trained, they move into the prediction generation phase. Each modality-specific model produces an output, which could be a binary classification label, a probability score, or some other metric that indicates the likelihood of a particular event—such as fraudulent behaviour, credit risk, or compliance violations. These predictions represent the independent conclusions reached by each analysis pathway.

The decision fusion stage follows, in which these independent outputs are integrated to produce a unified decision. There are several approaches to this integration. Simple rule-based fusion might dictate, for example, that a case be flagged as potential fraud if two or more models exceed a certain risk threshold. Weighted averaging can be used to give greater importance to more reliable or historically accurate modalities. More advanced setups may employ stacked generalization, in which a separate meta-model is trained to learn the optimal way of combining the outputs based on past performance.

Finally, the process concludes with the final decision and explanation stage. The fused decision is presented as the ultimate system output, but equally important is the presentation of the modality-specific scores or assessments that contributed to it. This dual presentation supports transparency and interpretability, enabling analysts, compliance officers, and regulators to understand not just the final verdict but also the role each modality played in reaching it. This is particularly valuable in high-stakes BFSI contexts, where accountability and explainability are as critical as predictive accuracy.

4. Pre-Fusion vs Post-Fusion Trade-offs

Our analysis shows that neither approach is universally superior—instead, the optimal choice depends on use case priorities:

Factor	Pre-Fusion Strength	Post-Fusion Strength
Pattern Complexity	Captures subtle cross-modal correlations	Weaker, as patterns are learned per modality
Interpretability	Harder to explain decisions	Easier to trace per-modality contributions
Computation	Higher upfront processing	Parallel, potentially faster at inference
Fault Tolerance	Lower missing data can harm results	Higher modalities can be dropped
Adaptability	Stronger for tightly coupled data	Better for loosely related data sources

5. Comparative Analysis - Pre-Fusion vs Post-Fusion in BFSI

Selecting between pre-fusion and post-fusion approaches is not simply a technical preference, it's a **strategic decision** that can significantly impact accuracy, interpretability, compliance readiness, and operational costs. In BFSI contexts, these trade-offs are particularly critical given regulatory oversight, risk exposure, and customer trust considerations.

5.1 Analytical Framework for Comparison

Dimension	Pre-Fusion (Early Fusion)	Post-Fusion (Late Fusion)
Data Integration Stage	Features are combined before model training	Predictions are combined after model inference
Cross-Modal Learning	Strong, as interactions are modelled jointly	Weak, as modalities are modeled separately
Interpretability	Lower, as feature mixing can obscure modality-specific influence	Higher, as modality-specific scores remain separate
Robustness to Missing Data	Low—missing one modality affects all predictions	High—other modalities can still produce outputs

Model Complexity	Single unified model	Multiple specialized models plus fusion layer
Training Overhead	Potentially lower (one model) but requires complex preprocessing	Higher (train multiple models) but each is specialized
Adaptability	Harder to adapt if one modality changes significantly	Easier to swap or upgrade a single modality model
Latency Considerations	Potentially faster (one model pass)	May be slower unless parallelized
Best Fit Scenarios	Rich, well-aligned, consistently available multimodal data	Heterogeneous, intermittently available modalities with regulatory audit needs

5.2 BFSI Scenarios Where Each Excels

A. Pre-Fusion Wins

1. Real-time Fraud Detection in Mobile Banking[4]
    - Combines location metadata, transaction sequence embeddings, and device fingerprint in a unified feature space.
    - Captures complex correlations such as "High-value transaction + unusual device + abnormal location pattern".
2. High-Volume Credit Application Screening
    - Integrates structured financial history with parsed PDF application forms and voice tone sentiment from verification calls.
    - Single model optimizes for overall approval accuracy.

B. Post-Fusion Wins

1. Regulatory Compliance Reporting
    - Maintains distinct outputs from transaction monitoring, KYC document verification, and communication analysis models.
    - Allows compliance teams to isolate specific decision drivers during audits.
2. Insurance Claim Fraud Investigation
    - Combines results from image inspection, historical claim anomaly detection, and claimant behavioural pattern analysis.
    - Missing any single modality doesn't halt processing.

6. Design Considerations for BFSI Deployment

The deployment of multimodal fusion systems in BFSI is not purely a data science problem, it is equally a matter of governance, compliance, infrastructure, and operational resilience. The design phase must anticipate not only technical performance but also regulatory scrutiny, customer trust, and long-term maintainability[5].

6.1 Regulatory and Compliance Alignment

BFSI organizations operate under tight regulatory frameworks such as **Basel III**, **GDPR**, **PSD2**, and region-specific KYC/AML rules. Fusion design choices can directly affect compliance:

- **Auditability**: Post-fusion approaches tend to provide clearer audit trails, which can simplify compliance reporting.
- **Data Residency & Sovereignty**: Multimodal pipelines must ensure that all fused data processing respects jurisdictional boundaries.
- **Explainable AI (XAI)**: Regulators increasingly expect transparency. If a model denies a credit application, the system must show which modality's input had the greatest influence.

6.2 Data Quality and Availability

When data from all modalities is consistently available and well-aligned, **pre-fusion** becomes the preferred strategy, as it can extract richer and more complex patterns through deep cross-modal interactions. In such environments, the unified feature space allows the model to uncover subtle dependencies that might be invisible if each modality were processed in isolation, leading to more precise and contextually aware predictions.

In contrast, when data availability is intermittent, noisy, or prone to delays, **post-fusion** offers greater resilience. By allowing each modality to be processed independently, the system can still generate reliable decisions even if certain inputs fail, degrade in quality, or lag during real-time processing. This separation ensures operational continuity in live BFSI environments, where data inconsistencies are not only possible but expected.

6.3 Infrastructure and Latency Constraints

- **Pre-fusion**: Typically involves a single inference pass, which can be advantageous in real-time fraud detection scenarios with sub-second SLAs.
- **Post-fusion**: May require parallelized model execution to meet performance requirements; infrastructure should support concurrent GPU/CPU allocation.
- **Hybrid Approaches**: Require orchestration layers capable of dynamically routing and weighting outputs.

6.4 Security and Privacy

- Multimodal pipelines may aggregate highly sensitive data (biometrics, transaction logs, scanned documents). Encryption at rest and in transit is essential.
- Secure fusion layers should implement role-based access controls to prevent unauthorized access to full fused datasets.

7. Conclusion

In a **credit scoring platform**, the organization selected a post-fusion approach to maintain flexibility and modularity. By keeping the document parsing models and behavioural scoring models separate, the system allows each component to be updated independently without requiring a full retraining of the entire pipeline. This not only speeds up deployment of improvements but also reduces operational risk, ensuring that changes in one modality do not inadvertently affect the performance of another.

For **anti-money laundering (AML) transaction monitoring**, the choice was to implement a pre-fusion strategy. By combining transaction pattern features with network graph embeddings at the feature level, the model can capture deep correlations that would otherwise be missed. This unified representation enhances anomaly detection sensitivity, making it more effective at uncovering hidden relationships between seemingly unrelated transactions, a critical capability in identifying sophisticated laundering schemes.

In **retail banking mobile app fraud detection**, a hybrid fusion model was adopted to balance accuracy with robustness. Structured transaction data from multiple sources was merged through pre-fusion to exploit fine-grained cross-feature relationships, while biometric verification scores were incorporated via post-fusion. This ensured that fraud detection benefited from rich transactional insights while retaining the flexibility to handle biometric data streams that may occasionally be delayed or unavailable.

8. References

[1] Botwey, Ransford Henry, et al. "Multi-model data fusion to improve an early warning system for hypo-/hyperglycemic events." *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014.
[2] Abrahart, Robert J., and Linda See. "Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments." *Hydrology and Earth System Sciences* 6.4 (2002): 655-670.
[3] Azmi, M., S. Araghinezhad, and M. Kholghi. "Multi model data fusion for hydrological forecasting using K-nearest neighbour method." (2010): 81-92.
[4] Cao, Kang, et al. "Application research of multi-source data fusion and multi-model ensemble methods in aircraft approach state prediction." *IEEE Sensors Journal* (2024).
[5] Exbrayat, J-F., et al. "Multi-model data fusion as a tool for PUB: example in a Swedish mesoscale catchment." *Advances in Geosciences* 29 (2011): 43-50.