# Analysis of Multimodal Data Using Deep Learning and Machine Learning

**Sai Srujan Gutlapalli**

Interior Architect/ Designer, RI Group, NY, **USA**

**ISSN: 2311-8636 (Print)**
**ISSN: 2312-2021 (Online)**

**DOI prefix: 10.18034**



**Source of Support:** Nil

**No Conflict of Interest:** Declared

**Email for correspondence:**
sgutlapalli07@gmail.com

## ABSTRACT

A modality is an event or experience. Life is multimodal, see, hear, smell, feel, and taste. Multimodal experiences involve some world modalities. Artificial intelligence must grasp multimodal views to understand our surroundings. Multimodal machine learning models interact and correlate input from several modalities. It's a multi-disciplinary field with great potential. In this study, we analyze emerging multimodal machine learning technologies and categorize them scientifically rather than focusing on specific multimodal applications. Multimodal machine learning offers more potential and problems than classifications. Most multimodal learning research collects quantitative data from polls and surveys. This research reviews a detailed library of observational studies on multimodal data (MMD) skills for human learning using artificial intelligence-powered approaches including Machine Learning and Deep Learning. This research also describes how MMD has improved learning and in what environments. This paper discusses multimodal learning and its ongoing improvements and approaches to improving learning. Finally, future researchers should carefully consider building a system that aligns multimodal aspects with the study and learning plan. These elements could enhance multimodal learning by facilitating theory and practice activities. This research lays the groundwork for multimodal data use in future learning technologies and development.

**Key Words:** Multimodal Analytics, Machine Learning, Deep Learning

## INTRODUCTION

Multimodal Learning involves collecting, measuring, analyzing, and representing data on students and the learning environment to improve learning and infrastructure. Traditional learning information is one-dimensional. A learning management system's log data is solely valuable for assessing learning. This data ignores student context.

Understanding pupils' learning methods requires context-based environment data. Uni-dimensional data provide incomplete learning system data (Eradze & Laanpere, 2017), making learning analysis difficult. Learning is unpredictable. We need multimodal data on learning behavior, student facial expressions, and physiological data to accurately analyze and show a learning process. Thus, a better, universal learning method can be found.

Our world is multimodal. Modality is the occurrence or experience of something. Most individuals associate modality with sensory modalities. To simplify, sensory modalities target our main communication and emotion pathways like vision and touch. Adding many modalities makes a research challenge or database multimodal. Multimodal machine learning helps construct models that relate and extract data from several modalities. To comprehend human learning, we discussed six multimodal research objectives.

Due to data heterogeneity, multimodal machine learning researchers have unique opportunities and constraints (Carmi et al., 2013). A top-to-bottom depiction of regular portents can be obtained by collecting data from multimodal sources and learning multiple methods. This study explores five mostly-specialized multimodal machine learning challenges. We also explore the opportunities multimodal learning provides once these problems are overcome since they are intrinsic to the multimodal environment and should be addressed to improve the subject.

## Learning Analytics Methodology

### Dependent Variables

Scientists examined research advancements. These changes often lead field specialists in the hypothetical direction. However, dependent variables revealed some regular classes. Learning, behavior, teamwork, expertise, affect, attention, presentation, and success appeared across articles. Analysts usually analyze each construct differently (Eradze & Laanpere, 2017). These use heuristics or human coding. Analysts learn a comparable development using multiple modalities. Some scholars considered impact using speech, while others used expressions.

### Collaboration

Twenty two percent deemed the groups essential for assessing the dependent variables. In analytics, 48 percent focused on individuals and 30 percent on groups.

### Tools and Methods

Some multimodal landscape data was created using smart approaches to explore data streams. Some research projects employed specific codes, while most used existing codes and instruments. Linguistic Inquiry Word Count, OpenEAR, and FACET are examples. Scientists created bespoke API- and SDK-based technologies. SVM, Decision Trees, Bayesian Networks, and others were used by several of them. Numerous research developed supervised learning systems using hand-comments.

## Current Situation in the Multimodal Learning Environment

Existing literature in multimodal learning landscape depends on non-specific papers. These documents were based on field specialists' main concepts. This section describes current MMLA pioneer categories.

**Mobility:** A few articles suggested using cell phones to track users. Experts can consider members' real-time location and easily collect geo-location, video, and other multimodal information flows. This information also helps analyze educator development in a learning environment and identify student-student, student-innovation, and student-teacher relationships.

**Models and Frameworks:** Traditional multimodal learning tools and models are supplemented by systems and models that improve generalization and relevance. These models can also help specialists organize their study goals and define standards for data analysis across environments. Finally, models and frameworks can improve framework formation and give authenticity to multimodal models and tools.

**Data Display:** Field experts also want to solve data visualization, device development, and tool integration issues. While some basic research concepts and tools have been established to help with such activities, new and better data visualization tools are needed. This part also addresses data standardization and multimodal analysis. Scientists encourage leveraging existing APIs, however data standards, collection, and presentation are closely related.

**Computer-Human Analysis:** Directing research on human-computer collaboration is another interest. Experts are looking for ways to bootstrap human analysis with AI or use human engagement in the data analysis pipeline.

**Classroom Orchestration:** This classification considers current work to make multimodal learning frameworks more prominent. This activity collects data using student-teacher interfaces and intelligent frameworks to help arrange the best user learning experience.

**Cross MMLA:** The Cross MMLA is a multimodal learning community innovation. Experts are increasingly using multimodal learning frameworks to improve learning across advanced and digital domains. Data collection, interoperability, and normalization are difficult when leading a large research project. These classifications cannot cover all multimodal learning analytics research. However, they address some of the most innovative concepts that will be developed by several domain specialists.

## Multimodal Data, Learning Indicators, and Their Relationship

### Modal Data

Most recent research valued multimodal information. Few papers systematically organized multimodal data. This order structure included digital, physical, physiological, psychometric, and environmental domains. Online frameworks, virtual analysis frameworks, and STEAM frameworks were used to train the learning system in the digital world. Physical domain includes motion sensor and bodily movement data from various gadgets. With the rapid use of AI-enabled sensors, data on pupils' head movements and finger taps became increasingly precise, polished, and accurate. This physical knowledge helped the system realize vast insights and analyses.

Physiological domain demonstrated students' learning status by identifying internal physiological reactions. Psychometric domain, another typical learning data stream, focused on self-directed surveys that abstractly reflected student psychological status (Sankararaman et al., 2016). The environmental domain included weather and temperature data from the classroom. Research shows that environment greatly impacts learning. Multimodal learning requires environmental data's growing importance. IoT, cloud storage, and wearable devices make high-frequency and tiny data management easy and precise. In some courses, multimodal learning frameworks enabled by machine learning show students' actual learning state more effectively and precisely. These multimodal data streams are essential for analyzing learning processes because learners interact with content, peers, and teachers in many ways.

## Learning Indicators

Multimodal learning analytics measures conduct, deliberation, reasoning, metacognition, sentiment, teamwork, communication, commitment, and learning performance. Some can be reclassified. Learning conduct is divided into online learning, classroom learning, and informal learning. Consideration is individual and joint. Self-sufficient and communal learners are mentioned. Real-time and online teamwork are involved. Engagements are classroom dedication and face-to-face learning. Overall performance is measured by the assessment score, or learning score.

Several studies suggest performance measurement methods to improve learning indicator precision. Some judge this performance using critical thinking skills. Some researchers study coordinated effort, task execution, and learning. Clinical action and verbal presentation are included. Experts found that certain learning indicators indicate the complexity and importance of certain of the indicators listed here. Some specialists analyzed learning system conduct, intellectual engagement, and sentiment. However, some exams assessed the entire learning process (Gutlapalli, 2016).

Specialists measured engagement using linguistic, kinesic, and vocal modalities, depending on commitment. Teamwork can be assessed independently, and learning feelings can be measured jointly. Selecting learning indicators has recommendations. Independent learning requires focus, whereas collaborative learning requires teamwork. Real-time interaction with less remote engagement has learning signs. After assessing this multimodal learning system, learning markers may vary.

## Learning Indicators and Multimodal Data

Multimodal Learning Ecosystem creates a multi-dimensional research field to complicate data-learning indicator relationships. Studies have found three types of multimodal information-indicator relationships: 1. 1:1 2. Many-to-one 3. One-to-many.

First relationship implied that information may estimate only one indicator. The most popular multimodal machine learning kind. As technology advances, measurement estimation of all types of information is slowly being used. Thus, such relationships are rare. Online surveys and interviews are popular ways to assess thinking. The new logic technique evaluates reasoning using sound data. Since physiological estimation is available, EEG data is utilized to test reasoning. These methods are considered second-type relationships by experts. Second relationship showed how many data sources evaluate a learning indicator. ECG and EEG measure student engagement. Finally, the third sort of relationship: one type of information can measure several indications. For instance, eye movement indicates student attention.

We have so many corresponding relationships since substantial signs and data alter with specialized and hypothetical situations. A certain sort of information has a limited evaluation scope and defined benefits. Multiple indicators have meaningful measuring aspects. Online learning data is used to demonstrate learning behavior, eye movement to assess intellectual level, and data management according to learning content. Expressions affect emotions and commitment more. Expressions are a good portion of intense feelings in a learning atmosphere. Research specialists say indicators can be measured using one-dimensional or multi-dimensional data streams. However, such indicators should be estimated using existing data and data integration-relevant data combinations.

## Major Challenges in Multimodal Machine Learning Setting

In this section, we will go over some essential principles related to the five fundamental problems that are present in multimodal machine learning environments:

**Representation:** How to express multimodal information to show complementarity and synchronization is the first difficulty. Information variety makes connected and joint representations harder. Language is figurative, whereas music and graphics are indicative.

**Translation:** Deciphering information using one approach then another is the next issue. The data is varied and abstract. When vocally describing a picture, multiple depictions may be correct. Multimodal evaluation and translation may be abstract.

**Alignment:** Connecting components from at least two modalities is the next hurdle. How can we match specific signals to a person's voice and gestures? Long-term situations may align modalities, and the categorisation is often confusing (e.g., words or phrases).

**Fusion:** Fourth, incorporate data from at least two modalities to anticipate discrete or continuous events. To identify words, linguistic signals and lip movement graphics are used. Modal data may vary in power and noise. Such kinds require multimodal fusion.

**Co-learning:** Information transmission between modalities and representations is the fifth and final challenge. How might hypothetical grounding and zero-shot learning assist a machine learning framework learn a different method? This is especially challenging when one modality has limited learning resources (e.g., labeled datasets).

## Opportunities Presented by Multimodal Frameworks

The pace of adoption surrounding the incorporation of multimodal applications into mobile devices is continually increasing, and the following end-market verticals are most likely prepared. For the purpose of instantaneous inference and forecasting in the automobile industry, MMLA is being associated with ADAS (Advanced Driver Assistance Systems), HMI (In-Vehicle Human Machine Interface), and DMS (Driver Monitoring Systems). Retailers of advanced robots are aggressively incorporating MMLA frameworks into human-machine interfaces (HMIs) and automation in order to increase product quality and consumer appeal. This helps facilitate a more prominent cooperation between workers and AI-enabled robotics in the modern landscape.

Consumer-focused businesses in the telecommunications and smart home industries are engaged in a cutthroat competition to demonstrate the superior value of their product or service to that of their competitors'. Because MMLA components and improved frameworks are fundamental to the process of producing an impact with advertising, consumer hardware companies are great candidates for incorporating multimodal learning-enabled frameworks into the products that they sell. Among the developing applications are the verification of home security systems and the verification of installation, amongst many others (Mandapuram, 2016). Although hospitals and medical clinics are still in the early stages of researching and adopting MMLA tactics, there are already some promising use cases emerging in the field of medical imaging. In spite of the relatively low rate of adoption, the importance of MMLA for both patients and doctors will make it difficult for medical administrations to oppose this advice.

To further create content recommendation frameworks, tailored advertising, and automated compliance systems, entertainment companies are now using MMLA as a tool to aid with organizing their audio and video content into labeled metadata. This allows the companies to better categorize their content. Since the MMLA invention has only just become available for the industry, there has been a limit placed on the installation of labeling frameworks up until this point.

## Conclusion

Multimodal machine learning analytics is becoming more important as more learning approaches and processes become available. Multimodal learning may connect AI devices and enable commercial knowledge and big advancement. Multimodal Learning Analytics was surveyed in this work. It identified multimodal machine learning trends, applications, and promising future advances. The recent studies suggest that multimodal learning research has built a robust framework for ongoing studies over a long period of time.

Analysts are claiming techniques to obtain MMLA analytics data from students in different learning contexts to conduct individual and community study. The research topic appears poised to keep up with environmental phases and extract information across disciplines. People will also see more robust systems, data integration, and tools. Experts can advance this research topic by using and improving machine and deep learning. More importantly, the discipline thrives by exploring how multimodal machine learning might improve ease.

## References

Carmi, S., Palamara, P. F., Vacic, V., Lencz, T., Darvasi, A. and Pe'er, I. (2013). The variance of identity-by-descent sharing in the wright–fisher model. Genetics, 193(3):911–928.

Eradze, M. & Laanpere, M. (2017). Lesson Observation Data in Learning Analytics Datasets: Observata. In Proceedings of the 12th European Conference on Technology-Enhanced Learning (EC-TEL 2017), Tallinn, Estonia, pp. 504–508.

Gutlapalli, S. S. (2016). An Examination of Nanotechnology's Role as an Integral Part of Electronics. *ABC Research Alert*, *4*(3), 21–27. https://doi.org/10.18034/ra.v4i3.651

Mandapuram, M. (2016). Applications of Blockchain and Distributed Ledger Technology (DLT) in Commercial Settings. *Asian Accounting and Auditing Advancement*, *7*(1), 50–57. Retrieved from https://4ajournal.com/article/view/76

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Curr Biol. 26:1241–1247.

**--0--**