Multi-Model Data Fusion in Machine Learning: A Study and Analysis

V. Haripriya[1]*, *L. Gnanaprasanambikai*[2]

Abstract

A technique that is frequently used for prediction is known as machine learning. Machine learning, a branch of artificial intelligence, is a highly developing area currently. Numerous libraries consist of a wide range of algorithms which can be used for prediction. The fundamental objective of machine learning is to create intelligent computers capable of thinking and acting like humans. The algorithms take a value as an input and predict an output. One aspect of the data is the building/training of a model using numerous algorithms on a large amount of data. The second step of applying machine learning in the real world involves using these models in the various applications. To make more accurate prediction and fastest decision, we need to analysis all information from different modalities like image, pattern, and text, audio and so on, which are related to the specific issue which is gone to be solved. So, in this paper is based on study of Multi modal Data Fusion in Machine learning is the technique of combining data from several sources.

Keywords — Machine Learning; training data; Prediction; algorithm, Multi-model, Modalities, and Fusion;

I. INTRODUCTION

Machine learning is a branch of computer science that uses statistical approaches to enable computer systems to learn and improve based on previous experiences and observations without being explicitly programmed. Instead of programming, users feed information to a generic algorithm, which generates logic according to the provided information. Tom M. Mitchell proposed a more formal explanation of the algorithms researched in the machine learning discipline, which is commonly cited: "a computer program is said to learn from experience E with respect to some class of tasks T and performance measures P if its performance at tasks in T, as measured by P, improves with experience E" [1].

Software offerings may become more accurate at formulating predictions without being explicitly programmed because of kind of algorithms in machine learning (ML). The fundamental idea underlying machine learning is to develop algorithms that are capable of taking input data and, simultaneously upgrading outcomes as new data becomes available, making use of statistical analysis in order to predict an output. The goal of machine learning is to create algorithms that can acquire data and utilize it to learn independently [2] [3]. The learning procedure begins with observational information, such as scenarios, first-hand experience, or suggestions for the purpose of identifying patterns in the information and planning future opinions more properly on the examples furnished. In another word, ML is Responding interrogatives with available information.

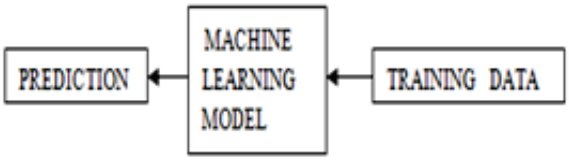

Fig. 1.1 Flow Representation

The machine gives an output from the available information or trained data, it is called prediction.

[1]Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore.
*Corresponding Author

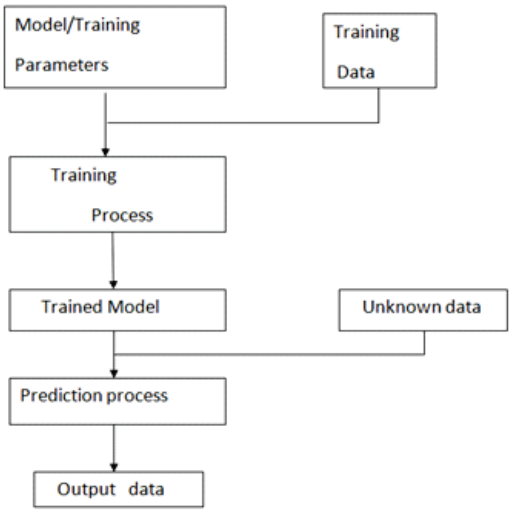

Fig.1.2 Architecture of ML

1.1 Multimodal Data Fusion

Information regarding the same phenomenon can be found in a variety of fields using different detectors, various circumstances, experiments, or subjects, among others. In order to define each of these frameworks for acquisition, we use a term called "modality." Natural phenomena include a wide range of properties; As a result, it is impossible for a single modality to provide comprehensive knowledge about the phenomena of interest. As additional modalities informing on a similar environment become available, greater levels of freedom emerge. These increased levels of freedom present problems beyond those pertinent to utilizing each modality independently.

Recent technological developments have expanded the amount of data sets that pertain to similar phenomena in an increasing number of fields, which has raised interest in effectively applying them. Because many providers of multi-view, multi-relational, and multimodal data are affiliated with significant impact business premises and cultural, biomedical, environmental, and military purposes, their thirst to establish innovative and effective analytical approaches is robust and extends beyond the boundaries of exclusively academic interest.

There are actually several reasons behind combining data. Improving decision-making skill, conducting exploratory studies, responding to particular system inquiries, such as recognizing regular vs. distinctive aspects over modalities or time, and gathering basic information from data are all for various motives. Data fusion is the procedure for analysing multiple data sets so that various data sets can be combined [4].

Multimodal data fusion in machine learning refers to the procedure of combining data from various sources or modalities, such as text, images, audio, and more, to improve the overall understanding and performance of a given task. This approach is particularly useful when individual modalities provide complementary information that can enhance the accuracy and robustness of a modal. The architecture of multimodal data fusion can vary based on the specific task and modalities involved.

Typically, multimodal consists of multiple forms of unimodal. An audio-visual model, for instance, can be made up of two unimodal networks: one for audio data and one for visual data. Typically, this unimodal process their data independently. This method is known as encoding. The information retrieved from each model must then be combined after unimodal encoding. Many fusion methods, from basic concatenation to attention techniques, have been proposed. One of the most crucial success criteria is the multimodal data fusion technique. After the conclusion of the fusion, a final "decision" accepts the fused encoded data and is trained on the final task.

Multimodal architectures often include three parts in order to express it simply:

a) Different modalities are encoded by unimodal encoders. Typically, one for each input modality.

b) A fusion system that, during the encoding process, integrates the features extracted from every input modality.

c) A classifier accepts the fused data and predicts the output.

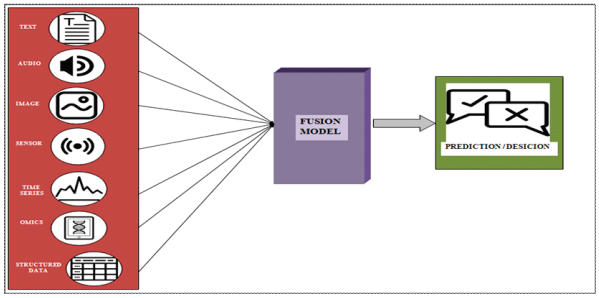

Fig. 1.3 Architecture

Multimodal Machine Learning's main objective is to establish a shared representation space that can efficiently gather complementary data from several modalities. Following that, many processes, like speech recognition, natural language processing, and picture captioning, may be carried out using this common representation.

Multiple neural networks, each specialized in analyzing a certain modality, are commonly used in multi modal machine learning models. Then, a variety of fusion techniques are used to merge the output of these networks.

II. STRATEGY OF FUSION

Multimodal data fusion is the process of fusing data from numerous sources for regression, classification, and other tasks [5].The architectures of machine learning models vary based on the fusion approach. Data from different modalities are combined using machine learning techniques or even more basic arithmetic operations like uncomplicated combination[4]. Fusion is frequently done at all three phases and might take place at various stages throughout the modeling phase.

2.1 Early Fusion

Early fusion involves fusing model characteristics at the model's input layer, usually by integrating many different kinds of data before applying a specific algorithm [5].It involves combining features or representations from different sources or modalities before passing them through a model. For example, if working with multi-modal data like images and text, it could concatenate image features and text embedding's before feeding them into a neural network. This allows the model to learn joint representations from the combined features. Early fusion can be effective when the relationships between different modalities are straightforward and can be learned together. It is also known as Feature-Level Fusion .

Data Pre-processing is the initial phase of Early Fusion; in the context of multi-modal data, each modality (e.g., pictures, text and audio) is processed individually to extract significant characteristics or representations. Visual characteristics derived from pictures using convolutional neural networks (CNNs), embedding's created from text using natural language processing techniques, or spectrogram features recovered from audio sources might all be examples. Second phase, after extracting features or representations from several modalities, they are integrated into a single feature vector. Combining feature vectors from several sources may be as easy as concatenating them. When working with pictures and text, for example, it may concatenate the image features and text embedding's to form a combined feature vector.

After that, the merged feature vector is fed into a machine-learning algorithm. Early fusion allows the model to learn interactions and dependencies between different modalities right from the start. It can help capture complex relationships that might not be apparent when processing each modality independently and it can lead to better feature representation by jointly learning from different sources of data.

However, early fusion might face challenges if the modalities have very different scales or if one modality is noisier than the other. Proper normalization and feature scaling are important in such cases.

2.2 Late Fusion

Model architecture in late fusion obtains predictions at the decision level and is hence frequently referred to as decision fusion. Generally, a number of algorithmic models are

trained in late fusion (generally one per data type). Parallels may be drawn between combining these many models into an integrated evaluation and ensemble learning because this is basically what is occurring, but instead of several independent models trained on similar data. Late (or decision-level) fusion generates unimodal result outcomes that are then combined to produce the most appropriate decision. Since late fusion ignores some low-level interconnections across modalities, it enables for easier training with better flexibility and predictability when one or more modalities are not present. [6]. This approach is often used to improve the overall performance, reliability, and robustness of a system by leveraging the diversity and complementary strengths of different models or sources.

Each source or model involved in the fusion strategy makes predictions or decisions independently. These models could be different machine learning algorithms, classifiers, or even experts providing their opinions. The predictions or decisions from different sources are aggregated or combined to arrive at a final prediction. These consist of weighted, average, and majority voting. All models must be aiming at predicting the same outcome for late fusion to work.

Late fusion can take advantage of the diversity of models or sources, potentially leading to improved overall performance. It can help in scenarios where individual models or sources have complementary strengths and weaknesses [7]. Late fusion can increase the robustness of the system by reducing the impact of errors from individual sources. However, late fusion might require additional computational resources and can introduce some complexity, as it involves combining outputs from multiple models.

2.3 Hybrid Fusion

Hybrid fusion in machine learning refers to a combination of different fusion strategies or techniques to take advantage of their respective strengths and address specific challenges. Hybrid fusion aims to achieve better performance and robustness by leveraging the benefits of multiple fusion approaches [8]. This can involve combining early fusion, late fusion, and other fusion methods in creative ways. The exact workings of hybrid fusion can vary depending on the specific problem and data at hand.

Various fusion techniques frequently collect different aspects of the data. Combining them allows you to gather complementary information that a single strategy may not catch. This may lead to a more complete knowledge of the data's underlying trends. Hybrid fusion allows you to use a combination of models with different architectures, hyper parameters, or training data. This diversity can help mitigate over fitting and improve the generalization capability of your model.

The success of hybrid fusion depends on careful design, experimentation, and evaluation. It's critical to understand the problem, the data, and the strengths and limitations of each fusion approach before merging them. By leveraging the benefits of different approaches, hybrid fusion can lead to improved performance and robustness in various machine learning applications.

III. APPLICATIONS OF MULTIMODAL DATA FUSION

By merging data from numerous sources or modalities, multimodal data fusion plays a significant role in many applications by providing a more thorough and accurate understanding of complicated events.

3.1 Health care Diagnosis

Multi-modal medical data fusion based on machine learning has the potential to efficiently extract and integrate distinctive information of various modes, increase clinical applicability in diagnostic and medical assessment, and offer quantitative analysis, real-time monitoring, and treatment planning [7]. Medical providers may get an in-depth understanding of a patient's health with the help of multimodal fusion. Medical professionals may improve diagnosis and personalize therapy by combining data from imaging modalities, patient records, genetic data, and wearable sensors [8].

While implementing Multimodal Data Fusion, there are various challenges in the medical field. Formats, resolutions, and scales for various data modalities vary, and ensuring consistency, correctness, and dependability across many data sources. Consequently, it might be challenging to ensure data quality. Processing high-dimensional data and selecting appropriate features from each modality is hard. This makes it impossible to use a single model to solve every problem, thus it's important to choose the right fusion strategies based on the particular problem. Ensuring that clinicians can still comprehend the combined information.

3.2 Autonomous Vehicles

Multimodal data fusion is critical in the development and operation of autonomous vehicles. Autonomous vehicles rely on various sensors and data sources to perceive and understand their environment, make decisions, and navigate safely. Multimodal data fusion involves integrating information from different sensors and sources to produce a comprehensive and precise representation of the vehicle's environment. This improves the vehicle's perception capabilities, decision-making processes, and overall safety [9].

CCTV, LiDAR (Light Detection and Ranging), radar, ultrasonic sensors, GPS, IMUs (Inertial Measurement Units), as well as additional sensors are used in autonomous vehicles. Each sensor has advantages and disadvantages [9]. Data from various sensors are combined through multimodal sensor fusion to produce a more accurate and comprehensive representation of the surroundings. LiDAR, for instance, may give precise distance measurements, whilst cameras can provide information on colour and texture [10]. Adding these inputs together improves object localization, tracking, and detection. Multimodal data fusion is used in autonomous cars to improve perception, decision-making, safety, and flexibility. By addressing the problems of individual sensors, it makes autonomous driving more reliable and accurate[11].

3.3 Social Media Analysis

Multimodal data fusion is also applicable in the realm of social media analysis, where it can offer valuable insights into user behaviour, sentiment, trends, and more. Social media platforms generate vast amounts of diverse data types, including text, images, videos, audio, and user interactions[12]. Integrating and analyzing these multimodal data sources can provide a deeper and more nuanced understanding of online conversations, user preferences, and overall social dynamics [13].

Combining textual data with visual information (such as photographs and videos) might improve sentiment analysis. Images can offer extra indications about the emotions, responses, and sentiments conveyed by users, but text analysis alone could not capture the complete context or emotional tone of a post. Platforms may learn about user preferences, interests, and interactions by combining information from a variety of sources, such as posts, comments, likes, and shares, with visual material. User engagement may be increased and content recommendations can be personalized using this data.

3.4 Security and Surveillance

Multimodal data Fusion is essential for improving security and surveillance systems. Security professionals may increase the accuracy of threat detection, situational awareness, and decision-making by merging data from numerous sources and sensors. Guards may more precisely identify possible threats by integrating data from video cameras, audio sensors, and other sources [14]. For instance, security systems can identify abnormalities like glass shattering, gunshots, or violent behaviour by fusing camera data with audio analysis[15].

The examination of behavioural patterns that can point to questionable or illegal activity is made possible through multimodal fusion [16,17]. Unusual behaviour in crowded spaces or restricted zones can be detected using the combination of data from video cameras, motion sensors, and facial recognition systems.

The integration of data from multiple sources enhances decision-making[18,19], increases accuracy, and unlocks insights that wouldn't be possible using a single data source. Multimodal data fusion addresses the limitations of individual modalities and provides a more holistic understanding of complex systems[20,21].

IV. CONCLUSION

Multimodal data fusion encompasses a wide array of techniques, ranging from early fusion methods that combine modalities at the feature level, to late fusion methods that combine outputs of separate models, as well as hybrid approaches that blend both strategies. The choice of fusion technique should be driven by the specific application's requirements and the nature of the modalities being fused. With continuing research investigating innovative fusion approaches, refining our knowledge of cross-modal interactions, and expanding applicability to emerging disciplines like augmented reality, human-computer interaction, and beyond, the multimodal data fusion environment is dynamic. Further developments will be fuelled by continued cooperation between researchers, practitioners, and subject matter experts.

The importance of multimodal data fusion as a driver for developing machine learning applications is shown by this survey and research. The combination of many data sources will continue to be a crucial tool for creating increasingly complex, precise, and flexible models across a wide range of disciplines as technology advances and new problems are encountered.

REFERENCE

[1]Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 3 (3): 210– 229. CiteSeerX 10.1.1.368.2254. doi:10.1147/rd.33.0210

[2] Batta Mahesh," Machine Learning Algorithms - A Review", International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426.

[3] Susmita Ray, "A Quick Review of Machine Learning Algorithms", IEEE, 10.1109/COMITCon.2019.8862451, 11 October 2019.

[4] Dana Lahat, Tu¨layAdali, and Christian Jutten," Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects", Vol. 103, No. 9, Proceedings of the IEEE, September 2015

[5] SaeedAmal, LidaSafarnejad, Jesutofunmi A. Omiye, IliesGhanzouri, John Hanson Cabot, Elsie Gyang Ross, "Use of Multi-Modal Data and Machine Learning to Improve Cardiovascular Disease Care",Volume 9, Front. Cardiovasc. Med., 27 April 2022.

[6] Yagya Raj Pandeya&Joonwhoan Lee," Deep learning-based late fusion of multimodal information for emotion classification of music video",Springer, 17 September 2020.

[7] Xiangdong Pei • Ke Zuo • Yuan Li • Zhengbin Pang, "A Review of the Application of Multi modal Deep Learning in Medicine: Bibliometrics and Future Directions", International Journal of Computational Intelligence Systems, Springer, 16 March 2023.

[8] QiongCai, Hao Wang, Zhenmin Li, And Xiao Liu," A survey on multimodal data-driven smart healthcare systems: approaches and applications", Volume 4, IEEE Access, 2019
.

[9] Dana Lahat, TulayAdali, Christian Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects", Vol. 103, No. 9, September 2015.

[10] Mauro Dalla Mura; Saurabh Prasad; Fabio Pacifici; Paulo Gamba; Jocelyn Chanussot; JónAtliBene," Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing", Proceedings of the IEEE ,Volume: 103,

Issue: 9,10.1109/JPROC.2015.2462751 September 2015.

[11] Yong Zhang, Ming Sheng, Xingyue Liu, Ruoyu Wang, Weihang Lin, PengRen, Xia Wang, Enlai Zhao and Wenchao Song," A heterogeneous multi-modal medical data fusion framework supporting hybrid data exploration", Health Information Science and Systems, https://doi.org/10.1007/s13755-022-00183-x, 2022.

[12] B.Rajalingam, R.Priya, R.Bhavani, "Hybrid Multimodal Medical Image Fusion Using Combination of Transform Techniques for Disease Analysis", Published by Elsevier Ltd, Procedia Computer Science 152 (2019) 150–157.

[13] Ganesh Chandrasekaran, Tu N. Nguyen, Jude Hemanth D," Multimodal sentimental analysis for social media applications: A comprehensive review", https://doi.org/10.1002/widm.1415, 31 May 2021.

[14] Ching-Tang Fan; Yuan-Kai Wang; Cai-Ren Huang, "Heterogeneous Information Fusion and Visualization for a Large-Scale Intelligent Video Surveillance System", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Volume: 47, Issue: 4, April 2017.

[15] Keli Huang, Botian Shi, Xiang Li, Xin Li, Siyuan Huang, Yikang Li, "Multi-modal Sensor Fusion for Auto Driving Perception: A Survey", arXiv:2202.02703, last revised 27 Feb 2022.

[16] Yue Zhang; Bin Song; Xiaojiang Du; Mohsen Guizani," Vehicle Tracking Using Surveillance With Multimodal Data Fusion", IEEE Transactions on Intelligent Transportation Systems (Volume: 19, Issue: 7, July 2018).

[17]Yassine Himeur, Abdullah Alsalemi, Ayman Al-Kababji, Faycal Bensaali, Abbes Amira,"Datafusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations", Volume 64, December 2020.

[18] Said YacineBoulahia, AbdenourAmamra, Mohamed RidhaMadi & Said Daikh ," Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition", volume 32, Article number: 191,Springer,30 September 2021.

[19] Zhu, Y., Chen, W., Guo, G.: Fusing multiple features for depth-based action recognition. ACM Trans. Intell. Syst. Technol. (TIST) 6(2), 1–20 (2015).