Available online at: https://ijact.in

Date of Submission	04/04/2020
Date of Acceptance	05/05/2020
Date of Publication	30/06/2020
Page numbers	3714-3724 (11 Pages)

ISSN:2320-0790

An International Journal of Advanced Computer Technology

NOVEL HYBRID THREE SIGMA (3∑W.I.S.D.O.M) APPROACH: DEEP MULTIMODAL FUSION FOR SMART CITY APPLICATIONS TOWARD WISDOM

F.FATHI[1], M.OUZZIF[2], N.ABGHOUR[3]

[1]Fadwa FATHI RITM, Research Lab ESTC Hassan II University Casablanca, Morocco
[2]Mohammed OUZZIF, RITM Research Lab ESTC Hassan II University Casablanca, Morocco
[3]Norddine ABGHOUR, LIMSAD Research Lab Faculty of Sciences, Hassan II University of Casablanca, Morocco

Abstract: In big data era, data shows characters of large volume and velocity, especially variety that is also called heterogeneity, which is the generated datasets from various city domains. Recently, modeling heterogeneous data sources has gathered significant interest especially with the power of artificial intelligence. AI and big data, as it is the case with every tool, are used for good and bad whereas in reality we do not need just intelligence. We need wisdom to create a new powerful and complete image. Our concept is based on the inspiration from human being that every system is like a human being with five senses and the intuition or the sixth sense will be the result of the fusion of all other senses to pave the way to wisdom. In this paper we will showcase how diversity and heterogeneity are key to Fusion for better behaviour and Decision which is the area of the study of many domains like Healthcare, Self-driving cars, and smart Recruitment. Then, we will propose our novel hybrid three Sigma (3∑ w.i.s.d.o.m) approach, Deep Multimodal Fusion for smart city applications toward wisdom.

Keywords: Big data, Deep learning, Deep Multimodal Fusion, intelligence, wisdom component.

I. INTRODUCTION

Big data is poised to change the way we live and work. If structured data is big, then unstructured data is large and huge. The characteristics of big data, such as many data sources, diverse data structures, high data dimensions and fast growth make the traditional data storage and data mining technologies in need of technological innovations. Heterogeneity of data appears in structured data, semi-structured data and unstructured data including relational data, spatial data, temporal data, text data, image data, audio and video data. Due to the rich characteristics of natural phenomena and because the world is multimodal, it is rare that a single modality provides complete knowledge of the phenomenon of interest.

Data fusion is challenging because the data are generated by very complex systems and due to the augmented diversity, number, type, and scope of new research questions that can be posed is potentially very large. Traditional multi-source information fusion is a method of information processing for sensors or multi-source Systems[1]. It processes the measurement information obtained from multiple sources and uses methods such as information association, information integration and filtering to improve the estimation accuracy of the target state and other features, and ultimately to correct the situation, threats and their importance assessment. Traditional data fusion methods have limitations in multisource heterogeneous data fusion. A lot of research and many new fusion methods have appeared to solve this problem.

Deep learning is capable of the analysis and hence is the learning of massive amounts of unsupervised data. The deep learning could fusion the feature of multi-source heterogeneous data which is of considerable effect. However, there are still many problems to be solved in Multi-source heterogeneous data fusion such as how to solve heterogeneity problems among the data universally and how to achieve better integration, AI is a way to navigate and gather insights in the world of Big Data [2]. AI is a tool and like every new tool, it can be used for good and bad. In some cases, we don't need just intelligence. We have to control this artificial intelligence and that's why we need the wisdom.

We live in a world of increasing division because we've put too much faith in intelligence which can be artificial instead of wisdom. The arrival of AI raises a deeper question that needs to be addressed that, can we manage artificial intelligence effectively? How can we use the heterogeneity, multimodality, diversity of big data with the artificial intelligence to create a new powerful and complete image not just intelligence? We must acknowledge an underlying truth that is critical to a healthy society that our decisions must be made with the combined wisdom, not the presumed intelligence. If intelligence is knowing [3], then wisdom is knowing what to do.

In this paper, we will present our vision based on the inspiration from human being. We designed our architecture which is based on answering the following research questions: Q1: We as human beings when we make a decision, do we make it from what we have really seen and heard or it is from our intuition? Q2: In doing so we can develop a better understanding of not only how our intuitions work, but also why we have them on the first place? The goal of this paper is to provide some ideas, perspectives, and guidelines as how to approach data fusion. "Computers are incredibly fast, accurate and stupid whereas humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination." (Albert Einstein). That's the way [4] we incorporated the principles of human intelligence. Brain - body - environment, as a source of inspiration that allows us to put a new concept based on AI– big data –domain and pave the way for wisdom.

The main contributions of this paper are summarized as follows: We first give a brief background about our concept based on the human being with a brief review about big data and deep learning. Second, we introduce big data fusion and methods for heterogeneous data fusion which is the result of our study of different domains in smart recruitment, smart healthcare and self-driving. Our third contribution proposes a summary from intelligence to wisdom: toward a cognitive wisdom approach and platform which focus on the application of multimodal deep learning methods in multisource heterogeneous data fusion. The proposed framework introduces new functionalities to big data analytics frameworks represented in data model from heterogeneous multisource big data fusion to creates huge values that makes our systems wiser.

II. BACKGROUND; BIG DATA, AI, DOMAINS

A. *Diversity, multimodality and heterogeneity:*

Heterogeneous data's are any data with high variability of data types and formats[5]. They are possibly confusing and low quality-laden because of missing values, high data redundancy, and untruthfulness. Heterogeneous data are regularly produced from various sources, heterogeneity of big data also implies that it is an obligation to acquire and deal simultaneously with structured, semi-structured and even completely unstructured data. Big data relates large-volume and complex datasets with different independent sources. Multi-source heterogeneous is similar to multi-modal but contains more data types. In the field of information the modal can be understood as the existence of data formats, such as text, audio, image, video, and other formats.

Multimodality as a form of diversity; Diversity is the property that allows enhancing the uses, benefits, and insights in a way that cannot be achieved with a single modality. Multimodal data is heterogeneous, but all belong to unstructured data. Multi-source heterogeneous data contains structured, semi-structured and unstructured data, covering the multi-modal data types. Therefore, the difficulty of multi-source heterogeneous data fusion can't be underestimated. Furthermore, without Multimodal, all of these systems are known to exhibit limitations in terms of meeting robustness, accuracy, and overall performance requirements, which, in turn, enormously limit the usefulness of such systems in practical real-world applications.

a) Multisource heterogeneous data and there sources:
It includes identifying the diverse source frameworks and ordering them dependent on their nature and type and determines the type of data; structured, semi-structured or unstructured.

b) Ingestion and Acquisition:
It determines the frequency at which data would be ingested from each source in batch or real-time.

c) Storage:
Hadoop distributed file system is the most commonly used storage framework in Big Data world, others are the NoSQL data stores – Mongo DB, HBase, Cassandra etc. The two kinds of analytical requirements that storage can support: synchronous and asynchronous data

d) Processing:
The Processing methodology is driven by business requirements in Batch, real-time or Hybrid.

i. Batch Processing is collecting the input for a specified interval of time and running transformations on it in a scheduled way. Historical data load is a typical batch operation.
ii. Real-time Processing: it involves running transformations as and when data is acquired.

iii. Hybrid Processing – It is a combination of both batch and real-time processing needs.

In the following table, we have a Brief review about technologies:

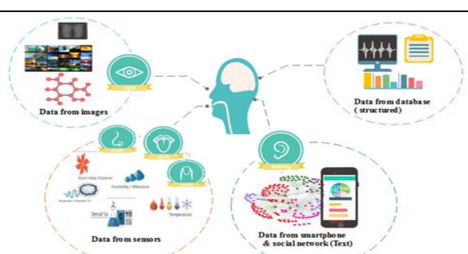
Table 1: Brief review about big data technologies

	Ingestion	Storage	Processing	Access
Description	Scalable , extensible to capture streaming and batch data	Depending in the requirement data	Is provide for both batch and near real time use cases	Data will be available to customer
Technology stack	Flume	Hdsfs	Hive	Qlikview
	Kafka	Hive Tables	Map reduce	Tableau
	Storm	Hbase/Map reduce	Spark	Apachkafka
	Sqoop	Elastic search	Storm	Rest APIs
	NfsGatway	No sql databases	Drill	Qlikview

B. From brain- body- environment to big data – Machine learning – Domain

In [4] We had used and learned from the human being as a reference and as a source of inspiration that allows us to put a new concept and pave the way for intelligent platform. The idea in [4] is that We as a human being with brain have the ability to perceive information that we get from the five senses and from memory to retain it as knowledge to be applied towards adaptive behaviors within an environment or context.

The idea is to classify all sources of data and compare it with five senses; for example humans have eyes to see and in our new platform we have images as a sight sense and the same thing for the other senses. [4] Represents the relationship between human intelligence system and our platform. The figure 1 describes the above idea: images for seeing data, Text to hear data and sensors to Taste, Smell and touch data. We focus on datasets from Images, sensors, and from Text: The main topics on relationship of this solution are: (1) Text analytics, (2) image processing, (3) Big data sensors.



a)

b)

Figure 1: [4] New design from principle of human five senses / Big data - machine learning - domain platform.

C. Text, images, sensors and big data

Text Analytics: Text is one of the most common form of stored information and includes social network feeds, e-mails, blogs, online forums, survey responses , communication, corporate documents, WebPages and call center logs are examples of textual data held by organizations [6]. Text analytics refers to techniques that extract information from textual data and involve statistical analysis, computational linguistics, and machine learning [7]. Today, natural language processing (NLP) technologies are widely used in a variety of applications like Sentiment analysis, Speech recognition, chat bots.

Image Processing: In recent years machine learning has revolutionized the fields of computer vision, currently, the most popular and successful ML model is convolution networks .CNNs have shown huge promise in automatic image and speech recognition. These include face detection and recognition systems, medical image analysis, image recognition, and full motion video analysis.

Sensor fusion: IoT represents one of the main markets of big data applications because of the high variety of objects. Smart city is a hot research area based on the application of IoT. A great quantity of noise may be collected during the acquisition and transmission of data in IoT.

D. When big data meet Deep learning

Big Data is usually collected from multi-modalities[8]. Each modality has a different kind of representation and correlation structure. For example, an image is usually represented by real values of pixel intensities but the text is usually represented as vectors of discrete sparse word count [9]. Deep learning systems are very large neural networks that are trained using considerable volumes of data and has gained wide spread attention as a reliable way to tackle difficult and computationally expensive problems. Deep learning is more appropriate for heterogeneous data integration due to its capability of learning variation factors of data and providing abstract representations for it.

Deep Learning cannot deal with confliction and fusion, which may normally exist due to the variation of data. However, there is little concern for how these systems were originally developed especially among newcomers in the field. Modern day deep learning systems are based on the Artificial Neural Network which is a system of computing that is loosely modeled on the structure of the brain. Deep Learning algorithms extract significant abstract representations of the raw data by using a hierarchical multi-level learning approach. The main driver behind this science-fiction-turned-reality phenomenon is the advancement of Deep Learning techniques, specifically, the Recurrent Neural Network (RNN) and Convolution Neural Network (CNN) architectures.

III. DIVERSITY AND HETEROGENEITY AS A KEY TO FUSION IN SMART DOMAIN: FROM FIVE TO SIXTH SENSES

A. Application in multiple domain like Self-driven cars, healthcare, recruitment

***Smart healthcare**:* The idea is to help people with disabilities and older people to regain their independence while preserving their privacy. An intelligent vocal and sensor based assistant able to learn from his environment, monitor, warn and recommend by analyzing situations and behaviors by Motion detection, falls, and Cognitive exchange system. It provides data analysis solution, based on the Data from sensors for connected object and motion detection, from images for fall detection, and data from text for the Voice exchange system like Google Home or Amazon's Echo. Digital service offerings that forces the current health system to promote access to the digital world, bringing an experience and a comfort of use, for them and their families and to improve support services by learning an AI, and then to help more autonomy, while preserving privacy.

Therefore, the goal of this solution platform is the study of membership behavior to ensure a cognitive exchange with patients and finally alert the right person.

Autonomous cars are the future of transportation as the cars will generate a huge amount of data per vehicle and hence there are many challenges [10] (Tesla, Volvo, Google, Udacity). By applying our architectures, the self-driving pipeline will be composed of Computer vision from images, sensors fusion, and the text analytics.

The mobility depends on what the camera see hence one of the challenges is to predict the steering angle of the car which is based on a feed of images coming from a camera[11]. [12] Simulate a Self-Driving Car, for data generation by using 3 cameras (left camera, center camera, right camera) associated with steering angle, speed throttle. [13] Implemented a real-time traffic detection using twitter tweets analysis for event detection with particular reference to road traffic congestion and car accident. By Using text mining technique and natural language programming; Classify traffic related tweets, apply tokenization, stop word filtering, steaming and steam filtering. [14] Explained that a self-driving car sensor suite, includes cameras, radar, lasers, lidars and ultrasonic sensors. By combining all the sensors into a suite allows the vehicle to build a more intelligent system. A mathematical algorithm called a Kalman filter is often used to combine this merge.

Smart RH: Big Data impact on how we live, how we work and consequently how we work together. In this way Big Data ought to be on the motivation of the Human Resources Manager. The recruiter finds both active and passive candidates who fulfill a job description. A few issues emerge when the candidate needs to go to the office to conduct an interview with the recruiter. [15] provide a solution architecture based on an interaction Chatbot [15] , and user profiling on the video interview. The idea is to automate the candidate interview process using artificial intelligence technologies and big data for a virtual assistant who is capable of orchestrating the whole interview process and then send a report with the better decision to the manager. We use different machine learning techniques to build and train models.

Table 2. Data sources and utility- Self driving cars –Smart Healthcare –Smart recruitment

Data from	**Sources**	**Utility**
Self driving cars		
Sensors	mobile sensors such as mobile phones, GPS navigation systems, on-board diagnostics systems, radar...	• In automated driver assistance systems and localization ,
Images	Cameras , Panels/ signs	• Road traffic panels / signs recognition ,automated driver assistance systems contributing to the safety of vehicles
Texts	Data from social network, chat bots...	• Road traffic congestion and car accident.
Smart Healthcare		
Sensors	IR: Infrared sensor	• Motion Detection / Human Presence
Images	Camera IR	• Falls detection
Texts	Data from chat bots.	• Speech recognition for Alzheimer detection
Smart recruitment		
From web	Scan resumes and social public	• Data Tools can automatically scan resumes and social public data of applicants and give a score how well the applicant matches the requirements
Images	Videos	•The data obtained from a video can be a useful source of information for emotion and sentiment analysis.

Texts	Data from chat bots: virtual assistant	Speech recognition for Big Data online evaluations; They way how a candidate answers the questions, and so on can provide more information about the applicant’s behavior than the answers to the questions

B. Diversity and heterogeneity as a key to Fusion: from five to sixth sense

[16]The sixth sense is any supposed sense or means of perception such as intuition or clairvoyance other than the five senses of sight i.e. hearing, touching, tasting, and smelling. We as Human being, even instead of having all information (audio, images, texts) sometimes have to refer to the sixth sense or to the “intuition” for the decision making. In the case of self-driving cars the question arises Q1 how can the sixth sense improve our concept? And in what capacity can the intuition enhance our answer for better acting in the case of smart recruitment? Q2 When we settle on a choice we make it, is it truly from what we have seeing and hearing or on the other hand from intuition? We will present how the combination of all the components can play the role of the intuition or the sixth sense.

In Recruitment: As human beings we also rely on multimodal information more than unimodal. It is apparent that we get a better understanding of a speaker’s intention when we see his/her facial expressions while he/she is speaking. Together aural and visual mediums provide more information than they provide alone. This is often the case when the brain relies on several sources of heterogeneity inputs in validating events.

Purpose: Our approach is like image illusion with two meaning. In figure 2.a) we have two meaning: First: when we take each component: Houses, the man, trees as a nature picture. Second: If we take the all picture with the combination of all component this will give us a new vision of the picture and more integrated and consolidated; a wise person.

So as a result like in b) If data sources from texts, images, sensors present the five senses; The intuition and perception, or the sixth sense are the result of the combination. From a) (figure 1) and a) –b) (figure 2) we can say that when we use one technology we can have the intelligence but with the combination we can get something which is more than intelligence. We can say that is something like a sixth sense. Hence we don’t need just an intelligent system but by using each sense (seeing, hear, taste, smell and touch) we need the use of all these senses to understand and get a wise system, and When we combine all this technologies, we can get like a maestro who plays on piano, to get a harmony music.

a)

b)

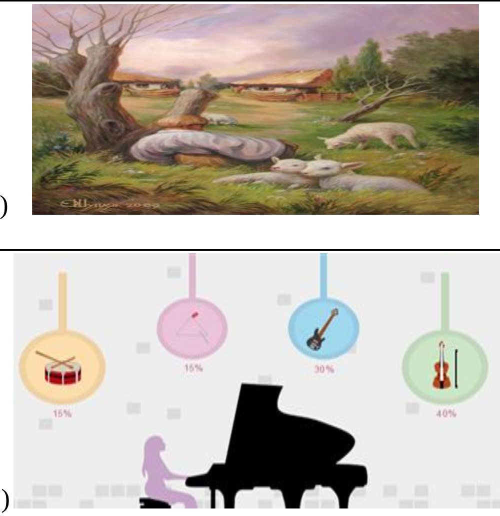


Figure 2 From five to 6 th sense

Q1: We as the human being really make a decision from what we have seen and hear or from our intuition? Q2: Can we measure this intuition? How can we calculate this intuition? Q3: In doing so, we can develop a better understanding of not only how our intuitions work, but also why we have them in the first place? .

We will present how the combination of all these senses plays the role of the sixth sense. Intuition is the result of combination of data from texts, images and sensors. If data sources from texts, images, sensors present the five senses; the intuition and perception or the sixth sense is the result of the combination, that influence the behavior and the action of our system is shown in figure 2 (a, b). Suggestion for question 2: how can we measure this intuition? Intuition is like a function, which is based on data from text, image and sensors. Suggestion for question 3: why intuition is at first placed when we take a decision. Sometimes we don’t need just one kind of data for the decision making. For example, during a car accident, a person may not see any flames, but the smell of burning rubber and heat proliferating through the dash would signal to the brain that a fire is kindling, thus demanding an immediate exit from the vehicle[17]. In this example, the information driving the brain’s reaction is greater than the aggregated dissimilar sensory inputs. Due to the rich characteristics of natural phenomena, it is rare that a single modality provides complete knowledge of the phenomenon of interest[18]. Data fusion is a challenging task for several reasons.

Result 1: Intuition is the result of combination of data from texts images and sensors. If data sources from texts, images, sensors present the five senses[4], the intuition and perception or the sixth sense are the result of the combination. Now question arises Q4 How can we represent this intuition? The solution is the FUSION.

C. *Multi source heterogeneous Data Fusion:*

Data fusion: is the investigation of several data sets such that different data sets can interact and inform each other. It is similar to multi-sensor data fusion which focuses on computation of structured and comparable IoT data to improve data quality to obtain appropriate decisions. Data fusion is generally divided into three levels i.e. [19] data layer fusion, feature layer fusion and decision layer fusion. Data layer fusion is the direct integration of data on the original data and analysis in a low-level integration. The feature-level fusion is a fusion of the middle level. The feature extraction of the original data is performed and then the feature data is analyzed and processed synthetically. It accomplishes extensive information compression and facilitates real-time processing. At the decision-making level the data are processed separately from various sources and the preliminary conclusions are acquired respectively. Then the decision-making fusion is made through the correlation processing and finally the joint inference result is obtained.

The integration of data layers requires that the data should be homogeneous and isomorphic while the structure of the multi-source data is heterogeneous. Therefore, Q5: How can we make these heterogeneous data work together to accomplish an objective? Such as identifying or retrieving tasks, through an effective fusion method is the primary problem to be solved by multisource heterogeneous data fusion.

Multi source Heterogeneous data fusion: Compared three levels multimodal information fusion approaches feature level fusion, decision level fusion and hybrid level fusion for cognitive load measurement. The multimodal fusion approaches outperformed individual modalities.

The feature level fusion: All features were input into the preprocess module, which standardized highlights and diminished their measurement with vital part analysis.

The decision level fusion: Preprocessed the features from each modality independently and after that input them into various classifiers. Each classifier output a cognitive load sub-decision. The fusion part summed all sub-decisions with weights for the final cognitive load (Df) as shown in Eq;

$$D_f = w_1 D_1 + w_2 D_2 + w_3 D_3$$

Equation 1.[20]Decision level fusion

The hybrid level fusion combined the feature level fusion and decision level fusion. Features from more than one modality were preprocessed to make one sub-decision; while other modalities features were preprocessed separately to assess other sub-decisions. The final decision summed all sub-decisions with weights.

In [20]Hybrid fusion had the best result of 84.66 % compared to other fusion methods. Fusion method pertains to the method used to fuse modalities[21]. Possible options include data-level, decision-level, score-level, hybrid, and model-level fusion. In data level fusion, individual data streams are fused prior to feature engineering (e.g., fusing video data from two cameras). Feature-level fusion consists of independently computing features from each modality and then fusing the features prior to classification. In decision-level fusion, classification is first performed on the individual features and the outputs (decisions) are fused via one of several voting rules. Score-level fusion is related to decision-level fusion which affect likelihoods (or probabilities) computed by classifiers operating on independent modalities are fuse only a small number of systems relied on score-level fusion so these were coded as decision-level fusion due to the similarity between these two methods. Hybrid fusion combines both feature and decision-level fusion for example by combining independent decisions of individual UM classifiers with the decisions of a feature level fused MM classifier. Finally, model-level fusion takes advantage of the interdependencies among the various modalities during the fusion process. When multiple fusion techniques were implemented and compared in a single study, the fusion method that yielded the highest accuracy was analyzed. [22]divided heterogeneous data fusion methods into three types: stage-based data fusion method, feature-level data fusion method, and semantic meaning-based data fusion method.

[23] presents traditional and the new techniques of Multimodal Classification using Feature Level Fusion. Late fusion-combination of results gotten by various classifiers, fusion is done at the decision level. Early fusion; information from various modalities are combined at the feature level, and classification is done on the combined representations.

[24] refered to multimodal fusion as the combination of information provided by different media, under the form of their associated features or the intermediate decisions. More formally, if M1 and M2 denote the two media and D1 and D2 the decisions inferred respectively by M1 and M2, the goal is to make a better prediction D1, 2 using both M1 and M2. More than 2 modalities can be used. This paper addresses the case of classification tasks, but any other task, e.g. regression, can be addressed in the same way. One recurrent question with multimodal fusion is where the fusion has to be done: close to the data (early fusion), at the decision level (late fusion) or in between. In the case of neural networks, the fusion can be done at any level between the input and the output of the different unimodal networks.

Fusion	**Formulate**	**Comment**
Feature-level or early fusion	p= h([$v_1, \ldots, v_m$])	•methods create a joint representation of input features from multiple modalities

Late fusion	p = F(h_1 (v_1), ..., h_m (v_m))	• uses unimodal decision values and fuses them with a fusion mechanism. F (such as averaging, voting, or a learned model.) • It is easier to handle a missing modality as the predictions are made separately
Multimodal Deep Learning	u=∑ mf_m (v_m) or u = [f_1 (v_1), .., f_1 (v_m)],	•The dotted boxes are representations of single and combined modality features. We call them additive combinations because their critical step is to add modality hidden vectors

Table 3 .Comparison of heterogeneous data fusion[25].

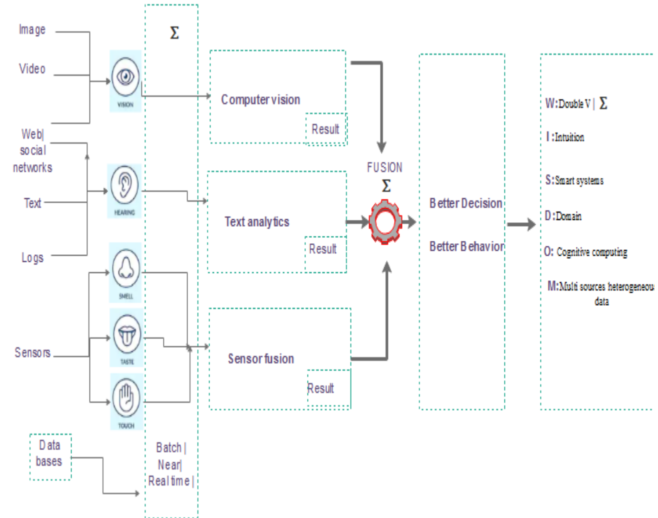
IV. THE (3∑) W.I.S.D.O.M PLATFORM AND DISCUSSION

A. The (3∑) W.I.S.D.O.M approach: From intelligence to wisdom

If we want to survive with the Big Data, then we must allow it to be autonomous and self-managed. The work presented in this paper is different in the following ways: The idea is that every system is like a human being with five senses (from big data) and the intuition or the sixth sense will be the result of the combination of all other senses to pave the way to wisdom. If intelligence knows, then wisdom knows what to do. As we develop more systems based on artificial intelligence, we can only hope we will have the wisdom to know what good we should do with this awesome new power. When reason becomes supple enough, when the emotional being is widened enough with empathy or when intuition reveals knowledge, all these paths lead to wisdom.

We propose a hybrid framework based and inspired from the human being, which outperforms an ensemble and fusion method toward wisdom. As we have seen in the previous examples, texts analytics alone or computers vision or sensors fusion are not enough. That is why the solution is the combination and the fusion of all results from different models and technologies to complete the image toward wisdom.

The proposed approach welcomes 3∑ layers; the first ∑ for all vs. of big data so it represents the big data layer. The second ∑ for Fusion models, the last ∑ for Wisdom which look like W (horizontal).These processes make a better choice for real-time big data analysis. The algorithms of AI proposed in the article for each unit and subunits are used to analyze datasets, which helps for better behavior and decision. 3-∑ architecture for big data analytics based on the fusion model.



W: In French,is called "double V", or $\sum Vs$ which refers to all Vs. of big data. Volume, Velocity,Variety. **I**: Intuition for sixth senses and Fusion. **S**: Smart systems self-learning. **D** : Domain and deep learning . **O**:CC Cognitive Computing. Cognitive computing involves self-learning systems; automate decision-making and problem solving and embodying computational intelligence by cognitive and autonomous systems mimicking the mechanisms of the brain. **M**: Multimodal deep learning

Figure 3.Three Sigma (3∑) new concept toward W.I.S.D.O.M with a new definition

B. Fusion for Better Behavior and Decision in some smart cities application or Domains.

According to the relations among sources [26] information fusion can be classified as follows; Complementary , Redundant, Cooperative:

- Complementary; when the information provided by sources represents different portions of a broader scene, information fusion can be applied to obtain a broad piece of information, like in the case of self-driving cars figure (4.b) for better behavior. Signals from different modalities often carry complementary information about different aspects of an object, event, or activity of interest.
- Redundant; If two or more independent sources provide the same piece of information, these pieces can be fused to increase the associated confidence as is the case with smart recruitment figure (4.c); is confident or not to the CV for better decision .
- Cooperative; Two independent sources are cooperative when the information they provide is fused into new information (usually more complex than the original data), which, from an application perspective, better represents reality. In the case of Healthcare figure (4.a); with the study of membership behavior to ensure a cognitive exchange with patients and finally to alert the right person; for better decision and behavior.

Depending on the previous method information, fusion can be processed with several goals, such as inference (to reach conclusion) and estimation. Fusion can be classified into three categories: rule-based, classification-based and estimation-based methods. The categorization is based on the basic nature of the methods and the space problem, as out lined. Combining complementary information from multiple modalities is intuitively appealing to improve performance of learning-based approaches[25]. It is challenging to fully leverage different modalities due to practical challenges such as varying levels of noise and conflicts between modalities.

Therefore, learning-based methods that combine information from multiple modalities are, capable of more robust inference. A natural generalization of this idea is to aggregate signals from all available modalities and build learning models on top of the aggregated information[25], ideally allowing the learning technique to figure out the relative emphases to be placed on different modalities for a specific task. This idea is ubiquitous in existing multimodal techniques including early and late fusion, hybrid fusion, model ensemble, and more recently joint training methods based on deep neural networks.

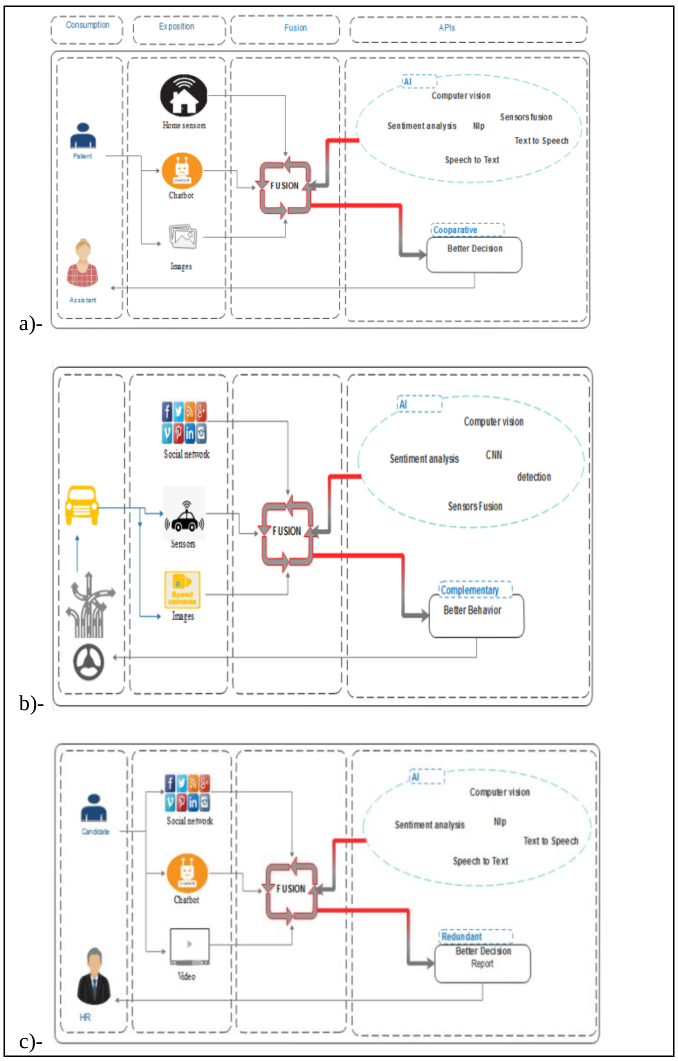


Figure 4: Smart Domains

C. *Multimodal Deep Learning for Heterogeneous big Data Fusion*

Multimodal learning is learning that involves multiple modalities, this can manifest itself in different ways: Input is one modality output is another, multiple modalities are learned jointly or one modality assists in the learning of another. What is difficult in multimodal learning is the different representations and missing data. We can solve these problems by combining separate models for single modalities at a higher level and Pre-train models on single-modality data. How do we combine these models? By Modelling two modalities jointly generate one modality from another or use one modality as labels for the other. The emergence of deep learning has improved the feature of multi-source heterogeneous data and made great progress in feature fusion effect, which makes the study of multi-source heterogeneous data fusion a big step forward. Deep learning is capable of the analysis and learning of massive amounts of unsupervised data. What's more, deep learning could blend the fusion, the feature of multi-source heterogeneous data, which is of considerable effect.

[27] Demonstrated cross modality feature learning, where better features for one modality, can be learned if multiple modalities (e.g., audio and video) are present at feature learning time. Show how to learn a shared representation between modalities and evaluate it on a unique task, where the classifier is trained with audio-only data but tested with video-only data and vice-versa.[27]Multimodal learning involves relating information from multiple sources. They considered three learning settings – multimodal fusion, cross modality learning, and shared representation learning. A prime example in the general space of multimodal deep learning is audio-visual speech recognition, where much work has been done using neural networks. A number of neural networks have been proposed to perform multimodal deep learning, including CNN, RBM and RNN. The choice of neural network often depends on the type of recognition involved, as there is currently no consensus on which network would best work. For instance, in tasks where sequential data is involved (e.g. image sentence description).

There is a lot of work about Multimodal Deep Learning Model; [28]discovered fusion architectures that exhibit state-of-the-art performance for problems with different domain and dataset. [27](image +audio) they focus on learning representations for speech audio which are coupled with videos of the lips. four components are included in [29](Image + text): visual feature learning, textual feature learning, multi-class classification, and label quantity prediction , [24] assume having one neural net per modality, capable of inferring a decision from each modality taken in isolation, and want to combine them. AVEF model [30] , Include four parts, i.e., audio-network, visual-network segment fusion model and global fusion model. In the context of user profiling in a social network, [25] helps to predict users' gender and age by modeling both users' profile pictures and their posts, [31](text +

image + audio) for video sentiment analysis solution, [32] for Multimodal data fusion , [33] for Behavior smart driving and other papers [34][9][35][36] .

D. *Hybrid 3 ∑ W.I.S.D.O.M platform for smart cities*

In this work, we propose hybrid architecture designed for analyzing big data in a Domain X using a big data fusion model that welcomes real-time and offline data. To do so, various machines are used for behavior observatory systems (images, videos and sensors), social network or web data and the data from databases.

In this article, we proposed a multimodal 3∑ architecture that is based on the fusion model technique. The proposed architecture efficiently processed and analyzed real-time and offline data for better decision and behavior. It is important to develop efficient big data cleansing approaches to improve data quality. Data virtualization and data lakes are powerful approaches that help data integration and the Fusion of multiple sources of heterogeneous data for a comprehensive analysis the use of multi-source heterogeneous data contained in the complementary information.

The proposed architecture for analyzing big data is generic (Figure 5), which can be used for any big data analysis in any Domain, remote sensing application, social networking application, and networking application. Furthermore, the capabilities of collecting data, fusion model, and Hadoop parallel processing of only useful information is performed by discarding all other extra data. The contribution of the proposed architecture is composed of various components and is summarized as following (Figure 5):

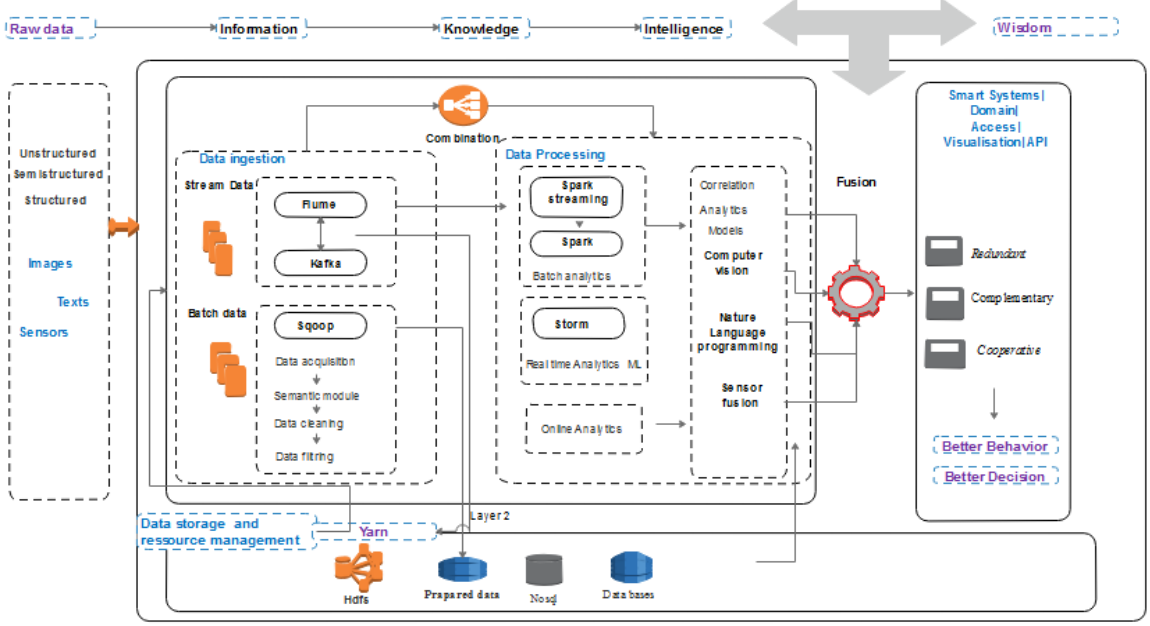


Figure 5: Hybrid 3 ∑ W.I.S.D.O.M architecture

V. DISCUSSION: UNIMODALITY AND EFFECTIVENESS LEAD TO INTELLIGENCE, MULTIMODALITY AND EFFICIENCY LEAD TO WISDOM

The reason becomes supple enough when the emotional being is widened with empathy or when intuition reveals knowledge. All these paths lead to wisdom. Wisdom is acquired in a similar way as it is also the accumulation of myriad experiences in life with one distinction of applied knowledge. By applying the lessons we learn through experience, wisdom is akin to intuition, insomuch as they both allow one to discern and judge with insight.

Intelligence is the ability to increase efficiency and the adeptness in problem solving and gaining information. Wisdom is the ability to increase effectiveness;
effective-adequate to accomplish a purpose, producing the intended or expected result, efficient-performing or functioning in the best possible manner with the least waste of time and effort. The difference between effectiveness and efficiency can be summed up shortly, sweetly and succinctly. While being efficient is about doing things right, effective is about doing the right things. "Intelligence tells you it's going to rain; wisdom tells you to head for shelter." One is a focus on "what is," the other is a focus on "what ought to be”. They are mutually reinforcing each other. Knowing to head inside when it is about to rain does not do you any good if, you have no way of knowing that it is about to rain. Intelligence is a largely static font of mental power from which one may draw to answer particular questions about the world. Wisdom is a dynamic system that guides behavior based on values, goals, and motivations. Intelligence is pursuit of knowledge, it tires the seeker. Wisdom is pursuit of truth, it inspires the seeker. Intelligence leads you. Wisdom guides you; an intelligent man understands what is being said, a wise man understands what is left unsaid.

When big data meet deep learning they find the solution for the bad side of big data and AI. Intelligence is from unimodality and wisdom is from multimodality. However, there are still many problems to be solved in Multi-source heterogeneous data fusion, such as how to solve heterogeneity problems among data universally, how to achieve better integration effect, how to stand at the height of artificial intelligence to realize multi-source heterogeneous data fusion at semantics level and so on. We still have a lot to do. To solve the problems existing in the feature extraction and multi-modal fusion better, the deep learning technology applied to various fields can play an important role.

Results	**Goals**	**Comment / explantation**
Result 1	From brain- body-environment to big data – Machine learning – Domain (Figure 1) => toward smart systems	The idea is to classify all heterogeneous sources of data and compare it with five senses; for example, humans have eyes for seeing and in our new platform, we have images as a sight sense, the same thing for the other senses. [4]
Result 2	From five to sixth senses ; the key to fusion	The intuition and perception , or the sixth sense are the result of the combination. Q4 How can we represent this intuition? The solution is the FUSION
Result 3	Result 1+ Result 2+ Smart domain(Figure 4)	Better decision and behavior.
Result 4	Result 1 +Result 2 + Result 3+multimodality(Figure 1)=>multimodality and efficiency lead to(3∑) W.S.D.O.M approach	The proposed approach welcomes 3∑ layers: • The first ∑ for all vs. of big data so it represents the big data layer. • The second ∑ for Fusion models • The third ∑ for Wisdom which look like W (horizontal)
Result 5	Result 1 +Result 2 + Result 3+ result 4 + Big data technologies and AI tools(Figure 5)	Five senses + intuition + better decision and behavior + deep multimodal learning => Hybrid 3 ∑ W.I.S.D.O.M architecture .
Result 6	Our definition of wisdom ; -**W**: In French is called "double V", or ∑Vswhich refers to all Vs. of big data. Volume, Velocity, Variety… -**I**: Intuition for sixth senses and Fusion.- **S** for Smart systems self-learning.-**D** for Domain and deep learning.-**O** for CC Cognitive Computing. Cognitive computing involves self-learning systems; automate decision-making and problem solving and embodying computational intelligence. **M** for Multimodal deep learning	AI is a way to navigate and gather insights in the world of Big Data, and like every tools they are used for good and bad. Our approach is related to the idea that, in reality we don't need just intelligence we need **W.I.S.D.O.M**

Table 4.Summary of results of this paper – Steps of results toward wisdom

VI. Future work and conclusion

This paper has presented our vision about wisdom and about several multimodal deep research works and has completed the above idea [4] about our new platform from five to sixth sense to pave the way for wisdom. When big data meet AI with the fusion, they pave the way for wisdom. The proposed architecture combines data from different data sources and fuses the result and performs complex analysis for better decision and behavior making in real time. In the future, even if the difficult side in multimodal learning in the different representations and Noisy and missing data, we will anticipate to design, implement, train, and test our fusion model in each domain suggested; smart recruitment, smart transportation and healthcare to improve accuracy and efficiency. A smart city is a city that is characterized as an "instrumented, interconnected, and intelligent, with our platform and vision of wisdom can we, in the future, talk about wise cities

VII. References ;

[1] L. Zhang, Y. Xie, L. Xidao, et X. Zhang, « Multi-source heterogeneous data fusion », in *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, Chengdu, mai 2018, p. 47-51, doi: 10.1109/ICAIBD.2018.8396165.

[2] « Artificial Intelligence and Big Data: A Perfect Match - DZone AI », *dzone.com*. https://dzone.com/articles/artificial-intelligence-and-big-data-a-perfect-mat (Acessed déc. 03, 2018).

[3] « The Difference Between (Artificial) Intelligence and Wisdom | LinkedIn ». https://www.linkedin.com/pulse/difference-between-artificial-intelligence-wisdom-stanley-bergman/ (Accessed déc. 03, 2018).

[4] F. Fathi, N. Abghour, et M. Ouzzif, « From Big Data Platforms to Smarter Solution, with Intelligent Learning: [PAV] 4 - Pave the Way for Intelligence », in *Proceedings of the 2017 International Conference on Cloud and Big Data Computing*, New York, NY, USA, 2017, p. 11–16, doi: 10.1145/3141128.3141143.

[5] L. Wang, « Heterogeneous Data and Big Data Analytics », *Automatic Control and Information Sciences*, vol. 3, nº 1, p. 8-15, août 2017, doi: 10.12691/acis-3-1-3.

[6] H. Hu, Y. Wen, T.-S. Chua, et X. Li, « Toward Scalable Systems for Big Data Analytics: A Technology Tutorial », *IEEE Access*, vol. 2, p. 652-687, 2014, doi: 10.1109/ACCESS.2014.2332453.

[7] A. Gandomi et M. Haider, « Beyond the hype: Big data concepts, methods, and analytics », *International Journal of Information Management*, vol. 35, nº 2, p. 137-144, avr. 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

[8] Q. Zhang, L. T. Yang, Z. Chen, et P. Li, « A survey on deep learning for big data », *Information Fusion*, vol. 42, p. 146-157, juill. 2018, doi: 10.1016/j.inffus.2017.10.006.

[9] P. K. Atrey, M. A. Hossain, A. El Saddik, et M. S. Kankanhalli, « Multimodal Fusion for Multimedia Analysis: A Survey », *Multimedia Syst.*, vol. 16, n° 6, p. 345–379, nov. 2010, doi: 10.1007/s00530-010-0182-0.

[10] O. Cameron, « We're Building an Open Source Self-Driving Car », *Udacity Inc*, sept. 29, 2016. https://medium.com/udacity/were-building-an-open-source-self-driving-car-ac3e973cd163#.fb7vtlrfn (Accessed june 20, 2018).

[11] « Teaching a car to drive using Deep Learning | LinkedIn ». https://www.linkedin.com/pulse/teaching-car-how-drive-using-deep-learning-muhieddine-el-kaissi/ (Accessed june 20, 2018).

[12] S. Raval, *How_to_simulate_a_self_driving_car: This is the code for « How to Simulate a Self-Driving Car » by Siraj Raval on Youtube*. 2018.

[13] SavitribaiPhule Pune University, S. G Panchal, et R. S. Apare, « Real Time Traffic Detection using Twitter Tweets Analysis », *International Journal of Engineering Trends and Technology*, vol. 47, n° 8, p. 458-461, may 2017.

[14] « How Sensor Fusion Works for Self-Driving Cars | LinkedIn ». https://www.linkedin.com/pulse/how-sensor-fusion-works-self-driving-cars-david-silver/ (Accessed june 21, 2018).

[15] « Smart Interviews: AI-Powered Recruitment - DZone AI ». https://dzone.com/articles/smart-interview-a-new-way-for-recruiting-candidate (Accessed jan. 10, 2019).

[16] F. Fathi, N. Abghour, et M. Ouzzif, « From Big Data to Better Behavior in Self-Driving Cars », in *Proceedings of the 2018 2Nd International Conference on Cloud and Big Data Computing*, New York, NY, USA, 2018, p. 42–46, doi: 10.1145/3264560.3264572.

[17] S. Poria, E. Cambria, R. Bajpai, et A. Hussain, « A review of affective computing: From unimodal analysis to multimodal fusion », *Information Fusion*, vol. 37, p. 98-125, sept. 2017, doi: 10.1016/j.inffus.2017.02.003.

[18] D. Lahat, T. Adali, et C. Jutten, « Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects », *Proceedings of the IEEE*, vol. 103, n° 9, p. 1449-1477, sept. 2015.

[19] F. Wang, L. Hu, J. Zhou, J. Hu, et K. Zhao, « A Semantics-based Approach to Multi-source Heterogeneous Information Fusion in the Internet of Things », *Soft Comput.*, vol. 21, n° 8, p. 2005–2013, avr. 2017, doi: 10.1007/s00500-015-1899-7.

[20] L. Zhang *et al.*, « Multimodal Fusion for Cognitive Load Measurement in an Adaptive Virtual Reality Driving Task for Autism Intervention », in *Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being*, Cham, 2015, p. 709-720, doi: 10.1007/978-3-319-20684-4_68.

[21] S. K. D'mello et J. Kory, « A Review and Meta-Analysis of Multimodal Affect Detection Systems », *ACM Comput. Surv.*, vol. 47, n° 3, p. 43:1–43:36, févr. 2015, doi: 10.1145/2682899.

[22] Y. Zheng, « Methodologies for Cross-Domain Data Fusion: An Overview », *IEEE Transactions on Big Data*, vol. 1, n° 1, p. 16-34, mars 2015, doi: 10.1109/TBDATA.2015.2465959.

[23] C. T. Duong, R. Lebret, et K. Aberer, « Multimodal Classification for Analysing Social Media », *arXiv:1708.02099 [cs]*, août 2017, Consulté le: déc. 15, 2018. [En ligne]. Disponible sur: http://arxiv.org/abs/1708.02099.

[24] V. Vielzeuf, A. Lechervy, S. Pateux, et F. Jurie, « CentralNet: A Multilayer Approach for Multimodal Fusion », in *Computer Vision – ECCV 2018 Workshops*, vol. 11134, L. Leal-Taixé et S. Roth, Éd. Cham: Springer International Publishing, 2019, p. 575-589.

[25] K. Liu, Y. Li, N. Xu, et P. Natarajan, « Learn to Combine Modalities in Multimodal Deep Learning », *arXiv:1805.11730 [cs, stat]*, mai 2018, Consulté le: sept. 10, 2019. [En ligne]. Disponible sur: http://arxiv.org/abs/1805.11730.

[26] E. F. Nakamura, A. A. F. Loureiro, et A. C. Frery, « Information fusion for wireless sensor networks: Methods, models, and classifications », *ACM Comput. Surv.*, vol. 39, n° 3, p. 9–es, sept. 2007, doi: 10.1145/1267070.1267073.

[27] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, et A. Y. Ng, « Multimodal deep learning », in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, Bellevue, Washington, USA, juin 2011, p. 689–696, Consulté le: mars 09, 2020. [En ligne].

[28] J.-M. Pérez-Rúa, V. Vielzeuf, S. Pateux, M. Baccouche, et F. Jurie, « MFAS: Multimodal Fusion Architecture Search », *arXiv:1903.06496 [cs]*, mars 2019, Consulté le: sept. 17, 2019. [En ligne]. Disponible sur: http://arxiv.org/abs/1903.06496.

[29] Y. Niu, Z. Lu, J.-R. Wen, T. Xiang, et S.-F. Chang, « Multi-Modal Multi-Scale Deep Learning for Large-Scale Image Annotation », *arXiv:1709.01220 [cs]*, sept. 2017, Consulté le: sept. 10, 2019. [En ligne]. Disponible sur: http://arxiv.org/abs/1709.01220.

[30] A. Zadeh, M. Chen, S. Poria, E. Cambria, et L.-P. Morency, « Tensor Fusion Network for Multimodal Sentiment Analysis », *arXiv:1707.07250 [cs]*, juill. 2017, Consulté le: janv. 20, 2019. [En ligne]. Disponible sur: http://arxiv.org/abs/1707.07250.

[31] J. Williams, R. Comanescu, O. Radu, et L. Tian, « DNN Multimodal Fusion Techniques for Predicting Video Sentiment », in *Proceedings of Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML)*, Melbourne, Australia, 2018, p. 64-72, doi: 10.18653/v1/W18-3309.

[32] D. Lahat, T. Adali, et C. Jutten, « Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects », *Proc. IEEE*, vol. 103, n° 9, p. 1449-1477, sept. 2015, doi: 10.1109/JPROC.2015.2460697.

[33] K. P. Seng, L. Ang, A. W.-C. Liew, et J. Gao, « Multimodal Information Processing and Big Data Analytics in a Digital World », in *Multimodal Analytics for Next-Generation Big Data Technologies and Applications*, K. P. Seng, L. Ang, A. W.-C. Liew, et J. Gao, Éd. Cham: Springer International Publishing, 2019, p. 3-9.

[34] V. Radu *et al.*, « Multimodal Deep Learning for Activity and Context Recognition », *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, n° 4, p. 1-27, janv. 2018, doi: 10.1145/3161174.

[35] F. Huang, X. Zhang, Z. Li, T. Mei, Y. He, et Z. Zhao, « Learning Social Image Embedding with Deep Multimodal Attention Networks », *Proceedings of the on Thematic Workshops of ACM Multimedia 2017 - Thematic Workshops '17*, p. 460-468, 2017.

[36] H. V. Le, T. Murata, et M. Iguchi, « Deep Modular Multimodal Fusion on Multiple Sensors for Volcano Activity Recognition », in *Machine Learning and Knowledge Discovery in Databases*, vol. 11053, U. Brefeld, E. Curry, E. Daly, B. MacNamee, A. Marascu, F. Pinelli, M. Berlingerio, et N. Hurley, Éd. Cham: Springer International Publishing, 2019, p. 602-617.