ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 3, pp. 379-427
DOI:10.33168/JSMS.2022.0319

ASEMMS: The Adaptive Smart Environment Multi-Modal System

Doaa Mohey Eldin, Aboul Ella Hassanein, Ehab E. Hassanien

Information System Department, Faculty of Computers and Artificial Intelligence, Cairo University, Egypt

d.mohey@alumni.fci-cu.edu.eg, aboitcairo@gmail.com, E.Ezat@fci-cu.edu.eg

Abstract. The primary objective of this research is to make use of diverse information such as textual, audio, and visual to learn a model. Effective interaction between these approaches often leads to a better-performing system. In addition, a new fusion taxonomy based on a data perspective introduces a fusion-based classification model that can neglect the context domain and focus on several features of the input data. The proposed fusion taxonomy is constructed based on the data perspective not on context perspective. This interpretation relies on data modality types, the importance of extracted features, the data noise, and the streaming time. It improves an interpretation of the hybrid fusion technique. The fundamental challenge of the smart environment is illustrated by infusing big sensory data extracted from IoT sensors and devices to support the main objective. Also, an adaptive smart environment multi-modal system is proposed as a solution for the modality fusion challenge to improve classification and prediction in various contexts, whether one or more data modality types. We have proposed a new adaptive smart environment multi-modal system (ASEMMS) for improving the classifications accuracy results. It relies on the common characteristics of smart applications such as smart health for monitoring patients remotely and smart military for improving the hyper spectral in night mode. The adaptive smart environment multi-modal system (ASEMMS) is designed based on constructing five layers, a software-defined fusion layer, pre-processing layer, dynamic classification layer, hybrid fusion layer, and evaluation layer. A software-defined fusion layer is considered a controller for managing data types, model types, and noisy data. A hybrid fusion layer is designed based on a tailored neural network for making a combination between Dempster-shafer and Concatenation fusion techniques for getting bigger number of features. It measures the accuracy and optimization for the classification results. For validation, we make two comparisons between the proposed adaptive system and the baseline of Dempsyer-shafer fusion technique and the baseline of concatenation fusion technique.

Keywords: Multi-modal, context-aware, neural networks, data fusion, smart environment

1. Introduction

Smart environment (SE) is defined formally as an intelligent factor that realizes the state of real environments with the physical sensors and smart devices to automate management systems with powerful performance and effective optimization (Alberti, A.M., et al, 2019). Another definition is introduced in a smart environment, which means a connection between smart devices and machines improves interpretation and task achievement (Raun, N.F. et al., 2017). Any smart environment system relies on various objectives, characteristics, and conditions for making decisions. The importance of constructing smart environments is illustrated in improving the making decisions and remotely management. It can monitor and track objects that can save time and cost.

According to the analytics website of statistics, the current usage of 2020, the expected connected sensors via the Internet for many smart contexts achieve 30 billion sensors (Statista, 2016). According to the Statista website, this statistic will increase to 75 sensors in 2025 (Vailshery, L.S., 2022). So, the smart environment becomes significant for discussion research to solve the problems of the real intelligent systems as presented in equation 1.

$$\#\, No.\ of\ used\ IoT\ Devices\ > \#\, No.\ of\ Poplution, \tag{1}$$

The statistics results are introduced from smart devices and sensors that are better than the number of Population world, as shown in Figure.1. The effective data-driven is very effective for following data or goals and prior status for any object Data-driven. These data are valuable in an industry that can utilize in data analytics, marketing, or sales. It also enhances the decision-making by following the real status of each goal that is very effective in the market.

Constructing any smart environment relies on a combination of the artificial intelligence of data analytics and Internet-of-things (IoT) to improve management control with high accuracy results. The main significance in a smart environment is the data and how to get it and interpret it. Any smart system requires fusing multiple data sources with the same data type or fusing multiple data types from the same data source.

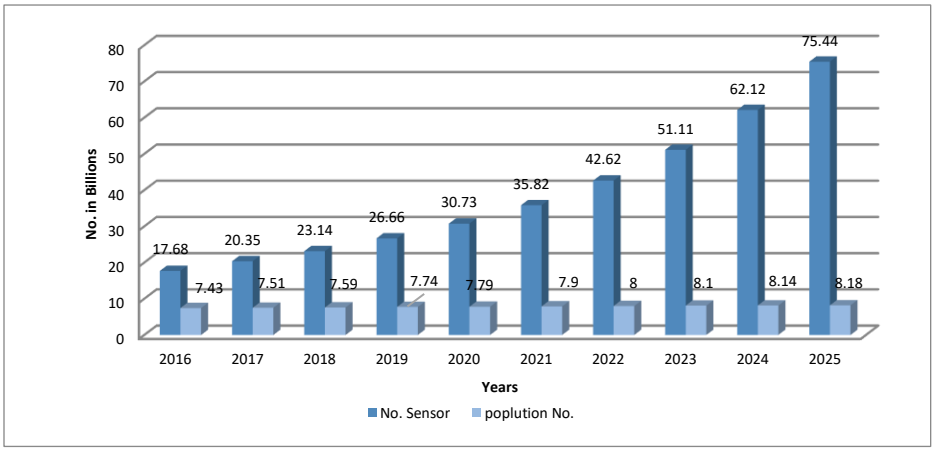


Fig.1: The statistics between population analysis & the double amount of used IoT sensors around the world (Vailshery, L.S., 2022)

Multi data fusion is the process of inference data from various data sources with various characteristics or different features. Each data source is different in target, organized architecture, input, context, and output. The input formers type can classify the outcomes as direct and indirect fusion. The extracted data requires making many processing to serve the main objective of smart information systems,

- Interpreting multi-Modality challenge that can be interpreted as data with complex relationships from multiple data types or multiple data features.
- Fusion multiple data sources: is an inference operation for integrating multi data sources that uses for being more consistent, accurate, and powerful data.
- Understand the Context-awareness for the capability of system components and parameters to fuse data about given intelligent environment system that depends on making adapt behaviors with respect to given time.

This paper introduces a proposed solution for the modality and context-aware challenges in various smart environment systems. It explores the design of adaptive smart environment multi-modal system. There is a requirement to be aware of the integrated perspective, with the objective of acquiring a solution where aspects related to both adaptive and multimodal systems are considered. The proposed adaptive multi-modal system explores the adaptive capabilities impact directly over the process of multimodal classification and fusion operations. It is designed for adapting to a differentiated context, including diverse user's requirements, execution platform, and specific intelligent environment. It is considered an evolution from past systems by consolidating aspects specific to multimodal interfaces directly in the development of an adaptive platform.

Previously, researchers use fusion techniques for improving classification accuracy and prediction results. However, there are still faced obstacles in accuracy

and validation results. Recently, there are some motivations to use machine learning (ML) for improving and solving previous challenges (Roy, CT., et al., 2017). Although many motivations try to solve these problems, no system can achieve a suitable solution for multiple data types or multiple data sources. So there is a need to design a solution for smart information systems to improve the prediction and accuracy results with a full vision of multiple data types or multiple data sources.

This thesis presents an information system for solving the modality of the context-aware challenge as mentioned in smart system. It constructs an adaptive smart environment multi-modal system that can solve information smart system challenges in combining characteristics, interfaces, and modalities. It aims to high effective of the modality in smart environments. A proposed adaptive smart environment multi-modal system improves the accuracy of classification and prediction results in various domains/contexts. The data fusion extracted from sensors with various physical characteristics that improves the interpretation analysis and produces the planning, decision-making, and the management of autonomous and intelligent machines.

Smart systems face many obstacles in "Modality" and "context-aware interpretation" problems to reach the full vision of extracted big data from the intelligent devices. The generated data hold big, heterogonous, and complex information data with different objectives to reach a full vision for data. The context-aware challenge refers to construct each system for each domain based on understanding parameters, conditions, and types. The modality challenge is interpreted into fusing multiple data sources with the same data type (image only, text only, or audio only) and fusing multiple data types from the same source (as images, video, audio, or text) in various smart environments.

The hardness of fusion technique can apply in many contexts, such as smart military or smart health.

1. It faces a challenge for weapon classifications in various spectrums especially in night mode, in the smart military (Yang,Z. et al., 2019).
2. It faces a challenge of fusion of health smart sensors for monitoring COVID-19 patients based on multiple data types, video, audio or text, in smart health (Kuang, S., and Davison, B.D, 2017)(Khoie, M.R., 2017).

These systems have faced the difficulty of making decisions for smart environments due to complex data, performance, data ambiguity, and unification target detection. They manage the hardness of dealing with many users in various smart environments that are entitled social Internet-of-Things (SIoT) and visualizing them by individual interfaces. It measures the classification accuracy and optimization of classification results. In addition, presents the validation based on two comparable systems.

The rest of paper is managed as the following: Section 2, related works, section 3, the proposed adaptive system, section 4, presents the experiments and evaluation

and validation results, section 5, Discussion, finally section 6, aims to the conclusion outlines.

2. Literature Review

Smart environment represents the Smart information systems in various contexts. These systems are different in features, parameters, and conditions. Smart environment systems or smart information systems are designed based on the communication between multiple sensors and smart devices. Smart environment systems extract big sensory data with heterogeneous data, diverse parameters, different data types, and various conditions. The main architecture of smart environment systems is designed as shown in Figure.2.

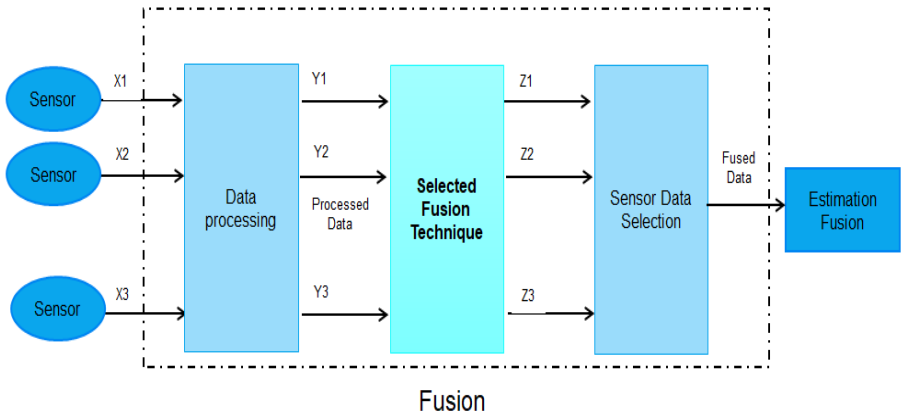


Fig. 2: The architecture of smart environment

2.1. Multi-modal fusion (modality)

Multimodal fusion is one of the recent research trends of artificial intelligence. Multimodal fusion achieves a benefit of the complementarity of heterogeneous data and provides trustworthy classification (Nasr, M. et al., 2021)(Atzori, L.,2012). It converts data from multiple single-mode representations to a compact multimodal representation. Multimodal fusion is an extremely important research direction and core technology in multimodal field research (as shown in Table.1). The main obstacle of data fusion is that no unification technique can be suitable to more than context, whether the same data type or variant data type. The fusion process of modalities from various data formats is a significant process to reach the highest performance and harvest relevant data.

Table 1: Modality data types in various smart environment systems

Modality data type	Features
Text	Detect language, syntax, semantic, and grammar.
Audio	Speed, length, and type of speech.
Image	Dimension type, resolution of the image, and image type.
Video	Number of frames, real-stream or offline

There are four classes of multimodal research that examine behavioral, computational, interaction, and suitable machine or deep learning techniques based on applying multimodal. Multimodal alignment faces many obstacles,

1. There are few datasets with explicitly annotated alignments;
2. It is hard to draw similarity metrics between modalities;
3. There occur many possible alignments, and not all elements in one modality have correspondences in another.

2.2. Multimodal data fusion

Multimodal fusion is one of the recent research trends of artificial intelligence. Multimodal fusion achieves a benefit of the complementarity of heterogeneous data and provides reliable classification for the model (Gaw, N., 2021).

Researchers have paid more attention to Multi-Modal Learning (MML) in (Sahu,S., and Vechtomova, O., 2021). Multimodal data fusion (MMDF) is the operation of integration diverse data streams (of disparate dimensionality, resolution, type, etc.) to produce information in a form that is more obvious or usable. Multimodal data fusion uses to improve information processing capabilities and warning performance of the monitoring system. That is different of multimodal data which is interpreted as Data is available across a wide range of modalities, from visual data in the form of images and video, language data in the form of text and audio data, video data for example music. The importance of multimodal data fusion is a key use of multimodal data is for educators to improve their teaching practices. Multimodal fusion has a very broad range of applications, including audio-visual speech recognition, multimodal emotion recognition.

Researchers in (Ahmad, J., et al., 2016) present a gender recognition strategy which employs Dempster-Shafer theory based reasoning process after the classification phase. The SVM probabilistic output makes the data analysis, validation measurement, used as evidence in the cause operation. Since the validation measurement phase, the fusion evidence is not satisfied enough and ignored. It makes many editions of the belief values for the two gender classes. The accuracy results reach 93.6%. Researchers in (Liu, K. et al.,2018) present new ways to integrate multimodal data that accounts for heterogeneity of modality signal strength across modalities for two levels, generally or partiality. The multimodal classification obstacle is considered with a focus on addressing the weak modality obstacle. A deep learning integration function is chosen automatically with suitable modalities per

sample and neglects weak modalities. The function applies to work on different neural network structures and is jointly trained in an end to-end fashion. A novel function to automatically choose mixtures of modalities is presented and evaluated. This function raises model capacity to catch possible correlations and complementariness across modalities. Researchers in (Ortega,J. D. S., et al., 2019) present a multi modal framework that is designed based on two layers. The lower layer refers to an extensible set of Deep Partitioned Autoencoders (DPAs). It delivers the input data from a single modality on uni-platform. It is useful for high dimensionality for enhancing motion-based tracking. But it is still complex joint statistics. Researchers in (Suk,H-II. et al., 2016) present a novel deep neural network (DNN) for multimodal fusion of audio, video and text modalities for emotion recognition. The proposed DNN has reached CCCs of 0.606, 0.534, and 0.170 on the development partition of the dataset for forecasting arousal, valence and liking, respectively. The proposed DNN has achieved CCCs of 0.606, 0.534, and 0.170 on the development partition of the dataset for predicting arousal, valence and liking, respectively. Researchers in (Akbari, H., 2021) a self-supervised multimodal representation learning framework based on Transformer architecture. With pure attention-based model on multimodal video inputs, our study investigation suggests that large-scale self-supervised pre-training is a propose direction to lighten the data burden for Transformer architectures and allows the Transformers to triumph Convolutional Neural Network (CNN) on various downstream functions. The learned representations by self-supervised learning across different modalities. The accuracy reaches 75%.

Our observation explores the new neural network is more powerful for improving the multimodal fusion as shown in previous four researches. The fusion statistical methods as Dempster-shafer theory are important for improving classification results. There is a need to developing an effective learning mechanism to dynamically determine weights used in the weighted regulatory architecture to enable online adaptation as shown in (Liu, Y-T., 2018). It creates an appropriate framework to allocate confidence for different sensory sources and establishing a comprehensive approach to handle different types of uncertainty. As shown in (Che, C. et al., 2020), there is a need for analyzing multiple fault diagnosis with various times and frequency domain features, and explores fault diagnosis early, middle and the full life cycle lately. There is a necessary for finding a method for improving robustness and generalization ability of the proposed method under continuous changing working conditions dynamically. Findings in (Shvetsova, N., 2021) a hot direction of research for domain adaptation or generalization in context multi-modal zero-shot recognition based multi-modal video processing in general.

Our observation presents the traditional multimodal learning methods contains early fusion, late fusion, and hybrid fusion. These fusion methods rely on as multiple kernel learning, and deep neural networks are highly active that discussed multimodal

fusion. Pervious motivations have been used to fuse information for audio-visual emotion classification, gesture recognition, affect analysis, and video description generation. While the modalities used, architectures, and optimization techniques might make a diverse, the generalize concept of inference data in joint hidden layer of a neural network stays the same. Recent motivations try to make a combination of the similarity probabilities across different feature components. The probabilities of dissimilarity between pairs of objects go through multiplication to generate the ultimate probability of be dissimilar, thus picking out the most confident component. Another used neural technique shows in ensembles multiple layers of a convolution network with a down-weighting objective function which is a specialized instance. There is a limitation of more generalization and flexible system for interpreting the modality combination and their strategies to address multimodal classification challenges. Single modality depends on the down-weights every class loss by one minus its own probability. Researchers use attention techniques for improving the combination multiple modalities based on multiplicative methods of feature level instead of decision level. Other multimodal tasks include LSTM for fusing sentence in a joint representation. Joint multimodal representation learning is utilized for media question answering (QA), visual integrity assessment, and personalized recommendation. Existing multi-modal methods do not adopt a joint technique to catching synergies among various modalities while simultaneously filtering noise and fixing conflicts of a per sample basis.

The significant limitation connects to the intrusiveness for intelligent sensing devices which make this modality impracticable for most context scenarios. Moreover the processing still cannot handle the existence of more than one sound of the audio stream. There is a shortage of investigations of fusing the visual and auditory streams. Prior motivations rely on the affect recognition. Moreover, there is not simple edit of the data based on the computational cost particularly in real time.

2.3. Adaptive multimodal fusion

The basic of adaptive fusion previous motivations are adaptive multimodal fusions that present two data types only with various smart contexts (Snidaro, L., 2015). The essential idea of constructing an adaptive fusion (also called quality fusion) is to give various weights associated with a modality. It improves the quality of modality with various characteristics and different features. The adaptive fusion presents dynamic fusion between multiple statistical or machine learning techniques. It is very important in multiple data types in various contexts. Context-awareness plays a vital role in adapting to context changes, conditions, and parameters that can reach the goal of computing services (Durkan, C. et al., 2018).

Prior research in 2005 in (Fierrez, J. et al., 2005) made many motivations for building adaptive multi modal framework based on using Bayesian classifiers and improving performance based on optimization functions. In 2009 (Dumas, B., et al.,

2009), researchers create a multi modal framework based on using the Bayesian classifiers and combined to function of optimization to improve classification results but that is not enough to improve the accuracy results. In 2011, researchers in (Kim, J.I. et al., 2011) present a design of human-centric adaptive multimodal interfaces that is designed based on a heuristic algorithm and existing ontology for improving the classification results. Although the motivation of multi modal for interpreting multiple modalities, the accuracy is not enough to understand multiple contexts. It also uses random algorithms with heuristic that makes the results not dependent.

In 2016, (Mezai, L., and Hachouf, F., 2016), present the fusion scheme to modalities of arbitrary nature and applied it on video and audio datasets. Each of the visual input modalities captures spatial data at a specific spatial scale. It depends on using multimodal deep learning in more detail and pay special attention. The accuracy reaches 96% with respect to domain dependent. In 2017,(Abidin et al., 2017) build a multimodal applications including in augmented reality environment. This research presents a conceptual framework to illustrate the adaptive multimodal interface in mobile augmented reality. It is considered a guide for developers to build a mobile AR applications with an adaptive multimodal interfaces. That uses for decrease the expert human inference in workloads. That applies on videos, Gps application on mobile. That has a limitation to applicable in many domains such as tourism.

In 2021, Researchers in (Zhou, B., et al. 2021) introduce an Adaptive Cross-modal weighting (ACmW) scheme to present complementarity properties from RGB-D data in this study. The scheme presents a relations learning among multiple modalities by integrating the features of different data streams. ACmW can automatically analyze the relationship between the complementary features from different streams, and can infer data between them based on the spatial and temporal dimensions. In 2021, Rameswar (Panda, R. et al., 2021) presents an adaptive multimodal learning framework, entitled AdaMML, that chooses on-the-fly the optimal modalities for each segment conditioned on the input for efficient video recognition. AdaMML uses neural network for interpret video segments for audio and frames. The experimental accuracy results improve 35%-55%. It improves the prediction decision results with respect to the target of reaching both competitive accuracy and efficiency.

Our observation finds from 2005 to 2010 researchers provide Multi-modal frameworks aspects that are of relevance in one domain context with interpreting the domain conditions or features. If it uses context to provide relevant information and/or services to users that relevancy depends on the user's task. From 2011 to 2017, researchers explores the modalities and uses many heuristic and probabilistic techniques for improves classification accuracy results for one context and multiple modalities. Since 2018, research goes forward to build adaptive multi modal for context-aware. The adaptively is shown in the flexibility of using in various cases for the specific input modalities and the same known context that is still faced many

obstacles to interpret multiple context with various conditions or features. In fact, both naive strategy and attention based approaches fuse the multi-modal features in a single way, which is not enough to extract complementary information and then limits the performance. In 2020, researchers present the need of adaptive multi modal context-aware system due to improve the entities consistency based on various situations. Context-awareness refers to as the operation of determining elements based on multiple functions. There is a challenge to interpreting cross-domain features and parameters or conditions. Any context-awareness application is designed data taxonomy due to interpret meanings. The classical context-aware extensive learning environment sustains for many physical limitations. Recently, the open research challenge of context-aware goes forward to interpret implicit and explicit meanings.

In 2017, researchers in (Chiu, C-K., et al.2017) presents a blended context-aware ubiquitous learning (b-learning) that is designed based on a navigation algorithm for the novel learning framework to be a guide students to learn efficiently in the b-learning environment. This framework relies on navigation algorithm, a mixed context-aware extensive learning system to support mechanism based on B-MONS was developed.

In 2019, Researchers in (Hasanov et al., 2019) present a new definition and open research area of Adaptive context-aware learning environments. The support for context-awareness and adaptation is essential in these systems so that they can learn features with relevant context. In 2021, Researchers in (Surve, A.R., and Ghorpade, V.R., 2017), presents a context-aware data fusion for smart health. Data management proceedings for healthcare systems are also represented, which contains preprocessing, context-aware data fusion and data processing and storage. It presents an approach for interpreting the context-aware for intelligent health system. It contains the context acquisition and filtering, situation building, and reasoning and intelligent inference. This explains how context-aware data fusion is executed by next doubled subtasks. In 2021, (Zhao, S., et al. 2021) presents propose an adaptive strategy for fusing the symmetric gated due to produce data fusion for the multi-modal contextual parallel. It uses for recovering a dense depth map from sparse LiDAR information input and dense RGB data an adaptive symmetric gated fusion strategy to make inference of data from multi-modal contextual representations parallel that impalements ACMNet for interpreting high quality depth maps.

Our observation explores pervious adaptive multimodal systems which are powerful but there are not usable enough due to the limitation of not applicable with multiple contexts. In addition, the hardness is shown in interpreting multiple the contexts modalities. So there is a need to construct a new adaptive multimodal system to fill the gap in the previous motivations and save time and efforts of expert people. The hardness of multi-modal context-aware system is presented in interpreting multiple features and their relationships. Another obstacle presents in interpreting in

the flexibility of usability in multiple contexts. So, there is a need to construct a new multi data fusion context-aware system to fill the gap in the previous motivations.

2.4. Data fusion levels

Data fusion depends on fusing various multi sources with diverse context types that hold some operations of data mining. Data fusion includes preprocessing data, identifying patterns, and visualizing these data. Data fusion levels examine three fusion levels, data fusion level, feature fusion level, and decision fusion level (Almasri, M. and Elleithy K., 2015). Data fusion applications are constructed based on fusion techniques, machine learning, and hybrid fusion techniques, as shown in Figure 2.8.

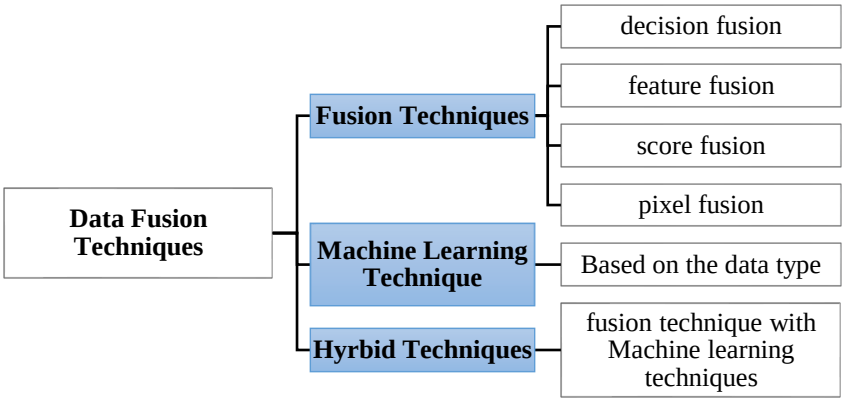


Fig..3: Data fusion techniques

2.5. Fusion taxonomy

Smart environment extracts big sensory data of pervious taxonomies of fusion. The design of smart environment requires interpretation of sensory data and parameters in context-awareness. Although there are four current fusion taxonomies, there is a need of context-aware interpretation for various smart environments. Researchers in (Emmanouilidis, C., 2013) make a motivation to construct a fusion taxonomy based on the abstraction and architecture levels. It relies on eight categories for managing the fusion control. It faces a problem in the lack of context storage. It also faces a problem of the criteria to secure data and secure the network. Authors (Almasri, M., and Elleithy, K. 2015) constructed a fusion taxonomy based on five categories for improving mobile guides. The hardness of this taxonomy is appeared of a lack of technology evolution, a lack of hard devices, and less performance, as shown in that focus on the complexity of the context system. Researchers construct a fusion taxonomy in (Gumawardama, A., and Shani, G. 2009) that improves saving power and benefit from data redundancy. This taxonomy meets a problem of no guaranteed spatial-temporal infusion process concurrently. The last fusion taxonomy is constructed in (Mujtaba , E.Y., and Elmustafa, S.A.A., 2019) to improve the fusion's

usability and prevent ambiguities. The comparison of previous fusion taxonomies is shown in Table.2. Pervious motivations present the smart environment architecture, advantages, and the current challenges. These taxonomies aim to support research to construct systems in a specific domain. They can't work on multiple contexts easily due to the hardness of interpreting data.

3. The Proposed Fusion Taxonomy for Context-aware

The new fusion taxonomy is designed based on four dimensions that present solving the modality to improve the context-aware problem. This taxonomy supports the full vision of any smart environment system. It is designed to interpret meanings and parameters without expert knowledge in a smart context (as shown in Figure.4). On the other hand, the proposed taxonomy can support researchers to create smart systems with neglect context domain. It can solve various characteristics, opportunities, challenges, and techniques. Detection-based fusion algorithms can detect the presence of a mine-like object but not categorize them. They provide many features concerning data and objects. Classification-based machine learning algorithms can classify mine-like objects, and they can also provide an output that is a union of possible objects. Contextual information includes many properties, features, and conditions for each context domain

This taxonomy was constructed based on four criteria as shown in Figure 3.4 and described as follows,

1. Modality Data type: Interpreting the data type of context (such as images, videos, audio, or text and numerical).
2. Data reduction: The feature reduction process type.
3. Concerning Noisy or Outliers data: The data noise (such as outliers, events, or errors), an
4. Data Time streaming: The time of streaming this data

The proposed Fusion taxonomy introduces a new classification modality for interpreting the context-awareness. The taxonomy is designed based on four dimensions, the type of context, the reduction operation type, the data noise, and the time of streaming this data.

The Context-aware Fusion taxonomy supports multiple contexts based on fusing various parameters, conditions, and features automatically. It improves the accuracy results of the predictive analysis and classification analysis. The Context-aware Fusion taxonomy helps researchers for creating dynamic/adaptive multimodal systems based on the four mentioned dimensions. It benefits in reducing ambiguities and errors. Other benefits of the taxonomy deal with outliers or noisy data minimizes time, the dynamic approach based on data types, and neglecting the context domain. The essential architecture still faces obstacles in interpreting the reduction level, the

noisy of missed or outliers, and time streaming supervision that affects the results analytics and users' requirements.

This taxonomy was constructed based on four criteria as shown in **Figure.3** and described as follows, **Modality Data type:** Interpreting the data type of context (such as images, videos, audio, or text and numerical). **Data reduction**: The feature reduction process type. **Concerning Noisy or Outliers data**: The data noise (such as outliers, events, or errors), and **Data Time streaming**: The time of streaming this data.

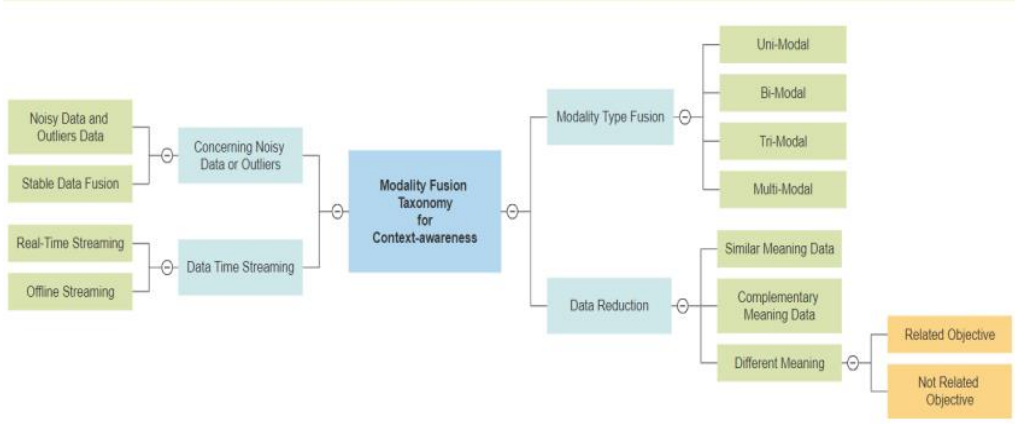


Fig. 4: The proposed context-aware fusion taxonomy architecture

Table.2: A comparative study for pervious fusion taxonomies for context-aware

Reference #	Taxonomy target	Categories #	Benefits	Limitations
(Almasri, M. and Elleithy K., 2015)	Makes supervision for the context-aware fusion taxonomy for implemented systems.	Eight	Presents an enhancement for controlling the fusion context of implemented systems.	The shortage of context storage and the scarcity of suitable criteria to protect data requirement.

(Emmanouilidis, C., 2013)	Produces mobile guides data fusion taxonomy	Five	Makes a classification of context to enhance mobile guides	The shortage of technology applications and hardness of used smart devices that causes of less performance.
(Almasri, M., and Elleithy, K. 2015)	Present a fusion classification taxonomy	Six	Saves power for getting a benefit of redundancy data.	No undertaking promise for spatial and temporal in fusion operation simultaneously.
(Gumawardama, A., and Shani, G. 2009)	The classification of fusion taxonomy depends on applications strategies	Five	Presents fusion taxonomy for classifying fusion strategies due to block ambiguously.	The high degree of complexity for constructing the fusion applications.

The fusion taxonomy architecture gives a full vision of the fusion characteristics for intelligent environment systems. They have been an integral part of the importance of the details of the fusion operation. The methodology aims to combine several topology structures into one type. The proposed architecture keens on all impacted properties (inner and outer) on the data fusion operation to achieve the highest fusion accuracy. The good fusion reaches good decision-making that refers to a direct correlation as the equation (2).

$$Good\ Fusion \propto Good\ Decision, \tag{2}$$

But other properties have a big impact in multi data fusion for a smart environment that is usually used to real-time stream with giving the data noise. However, the level of reduction relies on the user's requirements and user's number which is entitled Social Internet-of-Things as equations (3), (4), and (5).

$$RL \propto User\ Requirement, \tag{3}$$

$$\#\ Users \propto Reduction.level \tag{4}$$

$$RL = \sum_{Dt=1}^{n} \text{Relationship weight} + Modality\ type\ priorty \tag{5}$$

The interpretation of reduction level (RL) is interpreted by the weight of various relationships between parameters (Rweight) and the priority values between input data types (Mpriorty) as discussed in chapter 4. The relationships discuss parameters and their relationships. Various modality types interpret into vectors that can measure

in similar or different vectors. Complementary data discusses in various data types or characteristics. The reduction level is based on the similar vectors or different vectors.

The new taxonomy can support multiple contexts with various parameters, conditions, and features to improve predictive and classification analysis. It also can save time for selecting a suitable fusion technique. The proposed taxonomy helps researchers create dynamic multimodal systems based on the four mentioned dimensions. It benefits in reducing ambiguities and errors. It also presents many benefits in focusing on the main essential features, dealing with outliers or noisy data, Minimizing time, the Dynamic approach based on data types, and Neglecting the context domain.

4. The Adaptive Smart Environment Multi-Modal System (ASEMMS)

The adaptive smart environment multi-modal system (ASEMMS) aims to prepare the data for processing with data cleaning. It also provides handling the input data from noisy and outliers. An adaptive smart environment multi-modal system is constructed based on a new fusion taxonomy that interprets the input data perspective based on four parameters without focusing on the current context. It examines the main parameters (modality type, noisy data, reduction data, and time streaming). It provides the interpretation of the characteristics, features, properties, and conditions. It neglects the understanding of the domain expert system. An adaptive smart environment multi-modal system is a solution for multi- modalities, multiple data types in multiple unknown contexts. It also aims to

- Interpret modality type and number dynamically.
- Explicate the modality weights based on noisy and dataset size.
- Excavate the relationship between modalities from different perspectives.
- Improve fusion various information from multiple same or different data types.
- Reach high accuracy of classification.

An adaptive smart environment multi-modal system is designed based on four layers that are software-defined fusion layer, dynamic classification layer, hybrid fusion layer, and an evaluation layer as shown in Figure.5 and the detailed is shown in Figure.6. The concept of adaptive is developed in adapting to the context characteristics and preferences and broadening to the user spectrum. This adaptive smart environment multimodal system is used the semantic fusion approaches have been applied on multimodal text, speech, image, and video, for which the input modes number are coupled based on the context environment system. Adapting the input modal type interprets the quantity and method of data presented to both the user and display device. Although, the input modalities prepares different but complementary data that typically is combined at the utterance level. Semantic integration systems use individual recognizers that can be trained using Unimodal, bimodal, tri modal, or

multiple modals. With respect to semantic fusion, it can be scaled up user friendly when using input modes or vocabulary sizes. Adaptation of the ASEMMS system is effective in finding the relationships between features, excluding expert people, and keeping the performance of system based on the acceptable level. ASEMMS enables users to alter the user model or choose many assumptions for improving accuracy results. However, raises the complexity and convert attention from the major function. Divide the GUI interface in partition due to be adaptive and improve the predictable objects. The adaptive part presents make suitable suggestion classification and apply hybrid fusion on the GUI interface.

Fig. 5: The architecture of the adaptive smart environment multi-modal system (ASEMMS)

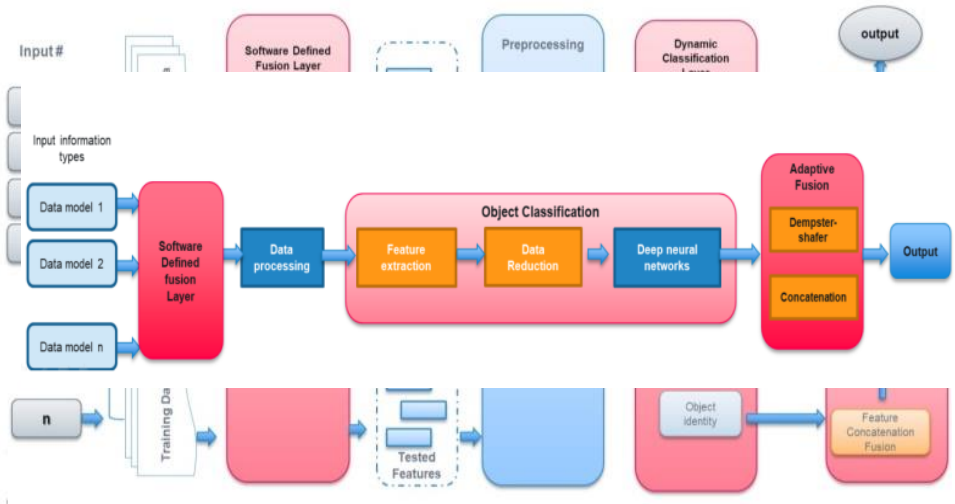


Fig.6: The detailed structure of the adaptive smart environment multi-modal system (ASEMMS)

The architecture of proposed system is designed based on four layers as the following,

4.1. Software-defined fusion laye

Designed as a controller for multiple modals types as the input whether Unimodal, bimodal, tri modal, or more multiple modals in various environmental systems as shown in Figure.6. A software-defined fusion layer plays a vital role in the managing and interpretation of parameters, features, and conditions. It increases the flexibility,

usability, and greater speed in constructing smart systems. It is based on five dimensions, modalities and data type's modality number, the weight of features relationships interpretation, and relationship weights. It aims to high accuracy and performance results for multiple data sources with multiple parameters. It is interpreting the input modalities type and the number of them. It also checks the relationship between multiple input data modalities and themselves. It extracts important conditions for the main target of the proposed modal. It deals with the priority data and deals with uncertain domain (unknown context).

a) Interpretation multiple modality data type inputs (Modalities): Interpret the inference of four modality data (image, text, audio, and video).

b) Dealing with multiple inputs numbers (Modalities #):

$$I(n) = \sum_{i=1}^{N} \sum_{j=1}^{x} Dt_{xN} \tag{6}$$

Proof

Let DT ➔ data type, let n= the total number of modality inputs,
Let Dt_x ➔ a type of degree of each data , Dt_N ➔ total number of inputs of each type

Input= DT#, DTx
$I(n) = \sum_{i=1}^{N} \sum_{j=1}^{x} Dt_{xN}$
$I(n) = \sum_{x}^{N} DT_{x1} + \sum_{x}^{N} DT_{x2} + \sum_{x}^{N} DT_{x3} + \sum_{x}^{N} DT_{x4}$
Let DT_{x1} ➔ Image data type, DT_{x2} ➔ Audio Data type, DT_{x3} ➔ Text data type, DT_{x4} ➔ Video
$\sum_{i=1}^{N} \sum_{j=1}^{x} Dt_{xN} = \sum_{x}^{N} DT_{x1} + \sum_{x}^{N} DT_{x2} + \sum_{x}^{N} DT_{x3} + \sum_{x}^{N} DT_{x4}$

c) The modality relationships:

It interprets the inference modalities in relationships. The weight factor of each dataset is computed based on the relationships between each dataset and neighbor dataset.

$$I(w) = \frac{\sum_{1}^{n} Dt_{x1N1}}{\sum_{x}^{N} Ds_{xN}} \tag{7}$$

Proof

Let DT ➔ data type, let n= the total number of modality inputs,
Let Ds_x ➔ total size of neighbor biggest datasets , Dt_{Nb} ➔ neighbor biggest dataset
Let Ds_{xc} ➔ Number of dataset size , Dt_{Nc} ➔ current dataset
Relationships (1-1, 1-m, m-m)
Input= Dt_{xcNc}, Ds_{xN}
If (Dt_{xcNc}, < Ds_{xN})
{
I(w)= $\frac{\sum_{1}^{n} Dt_{xcNc}}{\sum_{x}^{N} Ds_{xN}}$
c(w)= $\left|\frac{1}{I(w)}\right|$, ***let c(w)= computed weight of relationships (3)***
}
Else if (Dt_{xcNc}, >= Ds_{xN})
{
C(w)= I(w)
}

End if

d) Dealing with the priority data.

It is based on relationship between each data set with the lowest dataset. The adaptive smart environment multi modal system suggests the inference interpretion of the priority of each input modalities. This priority relies on the subtraction of each dataset size and the lowest dataset size that divides on the summation of the total size number of all input modalities.

$$\mathrm{P(f)} = \frac{\text{The difference between the Each dataset size with the lower dataset size}}{\text{Sum of the input datasize}}$$

$$\boldsymbol{p(f)} = \frac{\sum_1^n \boldsymbol{Dt_{xcNc-}} \sum_1^n \boldsymbol{Dt_{xlNL}}}{\sum_x^N \boldsymbol{Ds_{SN}}} \quad (8)$$

Proof

Let DT → data type, let N= the total number of modality inputs,
S[]= {s1,s2,……..,sn} , S → the sizes of all input modality datasets
Let DT_c → total size of the current input data size , Dt_L → lowest size of dataset size
Let DT_x → Number of dataset size , Dt_N → total number of input modalities dataset sizes

Input= Dt_{xcNc}, Ds_{xN}
p(f)= $\frac{\sum_1^n Dt_{xcNc-} \sum_1^n Dt_{xlNL}}{\sum_x^N Ds_{SN}}$

For example, smart health dataset#2 (which is mention in chapter 3) has 70.000 patient's metadata of Excel sheets with 1000 cough audio due to

$$p(f) = \frac{70000-1000}{71000} = 0.97$$

e) Dealing with uncertain domain (unknown context).

It is based on offline streaming for supervised learning that creates a model for improving decision making. It is based on various data types.

4.2. Preprocessing layer

A preprocessing layer refers to the normalization data from nulls or outliers and many augmented types on the input data. A preprocessing layer is constructed based on various data preprocessing in various modality types. Each modality type has a pre-processing data based on Normalization, Cleaning, and Augmentation. The pre-processing improves the data tuning and cleaning data or resizing data with various scales. The pre-processing handles the noisy data, or redundant data. It is applied to two methods checks with manual check on nulls or redundant data. Augmentation data enlarges the data size as mentioned in rotation, share, reflection, share, and scaling functions. An image data augmenter configures a group of preprocessing options for image augmentation, such as resizing, rotation, and reflection.

4.3. A dynamic classification Layer:

This dynamic classification layer is an automated layer for improving multi-object classifications and improving object detection that based on selecting a suitable neural network with respect to the input data types. A dynamic classification layer deals with various modalities input as vectors. Feature extraction is considered the operation of transforming raw data into numerical parameters that can be processed while preserving the data in the original data set. It yields better results than applying machine learning directly to the raw data. It relies on the automated feature extraction that uses specialized algorithms or deep networks to extract features automatically from signals or images without the need for human intervention. With the ascent of deep learning, feature extraction has been largely replaced by the first layers of deep networks – but mostly for image data. For signal and time-series systems, feature extraction remains the first obstacle that requires significant expertise before one can build effective predictive models.

- Image Feature Extraction: this dynamic classification layer applies transfer learning techniques of Alexnet, and Googlenet for images, video's frames, and audio's spectrogram. Image feature extraction influences the results and the resolution of each image that enhances the accuracy classification results. Colorized images, resolutions, and pixels are factors that should define from the first of the inference operation. Image feature extraction is made a zoom based on dividing into the image pixels twice times into half number of pixels. Audio feature extraction converts the audio into spectrogram images.
- Text Feature Extraction: although the text usually takes a care of the sentence's syntax, the context and intertextuality relies on the semantic meaning of the text. The text feature extraction interperats the core of the text that makes a fusion between multiple text documents which have relationship between theirselves such as excel sheets, or word documents. This dynamic classification layer relies on long-term short memory (LSTM). Feature extraction refers to the most discriminating properties in signals. Feature extraction applies to machine learning or various deep learning algorithms for

being more easily consume. Training machine learning or deep learning directly with raw signals often yields poor results due to the high data rate and information redundancy.

A dynamic classification layer makes compatibility between the appropriate types of neural networks based on the input data type (image, video, audio, and Text) as shown in Figure 5.10. It is a dynamic classification that is constructed based on:

- The number of input modalities and the type of input modalities.
- Making automated feature extraction for various modalities based on the learning models whether numerical features, and images features.
- It presents a solution for Multi-object classification. **Multi-object classification** is harder than one object classification due to need to interpret many features and characteristics of each input for improving the classification results. The dynamic classification layer include the classification utilizes predefined groups in which objects are distinguished (such images type dogs, cats, ext.), while clustering identifies similarities between objects (used for unsupervised) which it groups regarding to those properties in mutual and which differentiate them from other.

$$\{x_i, y\}^m \quad y \in \{1,2,3, \ldots\ldots N\} \tag{9}$$

A Dynamic classification layer uses two transfer learning techniques (Alexnet and Googlenet) that are very powerful for improving the accuracy results for images, Video' frames, and audio spectrograms' frames. The selection of Alexnet is the shortest direct pre-trained neural network which includes 25 layers using Matlab2022. The proposed Alexnet is development based on the modification in last three layers and freeze pervious weights and layers to benefit from the pre-trained model in Figure.7.

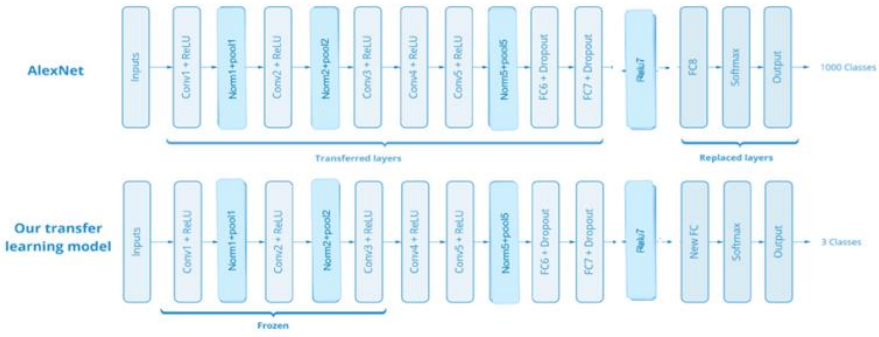


Fig. 7: A comparison between the current alexnet and proposed alexnet.

The selection of Googlenet is the longest neural network which is constructed based on 144 layers using Matlab2022. The proposed Googlenet is development based on the modification in last three layers and freeze pervious weights and layers to benefit from the pre-trained model in Figure.8.

Fig. 8: A proposed GoogleNet architecture

The Tailored Long-short term memory (LSTM) is constructed specifically for sequenced data based on time especially text classification. The proposed LSTM uses relu function and constructs based on 50 epochs and 50 hidden layers in Figure.9.

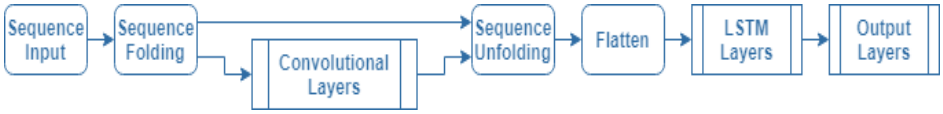


Fig. 9: The tailored LSTM for text modality

The classification layer has many properties for feature extraction and modality classification. It is respect to four factors, learn rate drop factor, learn rate drop period, MaxEpochs, and minibatchsize.

- Learn rate drop factor refers to a multiplicative factor to apply to the learning rate every time a certain number of epoch's passes, and the default value is 0.2 due to.
- Learn rate drop period means , and the default value is 5 due to.
- MaxEpochs refers to too many epochs can reach overfitting of the training dataset, and the default value is 20 due to.
- Minibatchsize aims to a subset of the training set that is utilized to measure the gradient of the loss function and update the weights, and the default value is 64 due to.

A dynamic classification layer uses many solvers for improving results SGDM, ADAM, and RMSprop.

- SGDM measurement default optimizer is based on working image, video, and audio SGDM/Gradient descent is an optimization algorithm that follows the negative gradient of an objective function in order to locate the minimum of the function. SGD with Momentum is a stochastic optimization function that increments a momentum term to a regular stochastic gradient descent.
- ADAM optimization refers to an extension to stochastic gradient descent which has newly shown broader adoption for deep learning applications in computer vision (CV) and natural language processing (NLP). Adam measures the text optimizer. It evaluates the adaptive learning rates for each characteristic. Adam is a replacement optimization technique for stochastic gradient descent for applying to deep learning models.
- RMSprop refers to Root Mean Square Propogation has an interesting history. RMSProp takes the scales of learning rate so the algorithms go forward through saddle point faster than most. It uses for making the neural network faster.

Adam (short for Adaptive Moment Estimation) gets the best of both worlds of Momentum and RMSProp. This layer requires installation many Matlab toolboxes for image processing, statiscial and machine learning, deep learning for Alexnet, deeplearning, deeplearning for Googlenet, TextAnalytics for text.

4.4. A hybrid fusion layer

This hybrid fusion layer consists of a hybrid between Dempter-Shafer statistical technique and concatenation technique with reducing the redundant vectors that refers to not important features. Various modality types interpret into vectors that can measure in similar or different vectors. Complementary data discusses in various data types or characteristics.

$$Reduction\ level = \sum_{Dt=1}^{n} \boldsymbol{R}weight \ + \boldsymbol{M}priorty \qquad (10)$$

Proof

Let f ➔ fusion,
LetRL ➔ reduction level , Dt_N ➔ count data type input
Let Rweight ➔ Number of dataset size , $M_{Pariorty}$ ➔ current dataset

$$f(RL) = \sum_{Dt=1}^{n} \boldsymbol{Rweight} \ + \boldsymbol{Mpriorty} \qquad (10)$$

The interpretation of reduction level is interpreted by the weight of various relationships between parameters **(Pweight)** and the priority values between input modality types (**Mpriorty**). The relationships discuss parameters and their relationships. The reduction level is based on the similar vectors or different vectors. Data Reduction relies on the parameters and their relationships or conditions between themselves and the priority between modality data inputs.

It is designed based on tailor neural network for constructing Dempster-Shafer technique and depends on suggestion of the belief and evidence. This hybrid fusion layer benefits from enlarging the number of features and parameters with a short time consuming. It is used for improving the classification accuracy results and prediction results. Figure.10 shows the hybrid decision fusion.

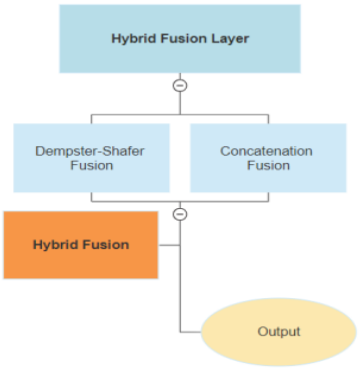


Fig.10: The proposed hybrid fusion approach for multiple data modals

This hybrid fusion consists of the integration between the Dempster-Shafer and concatenation fusion techniques with respect to feature reduction. The tailored neural network is analysis the network behavior in various conditions, Bayesian theory is

only concerned about single evidences and Bayesian probability cannot describe ignorance. It can accept a huge number of features or parameters use a hybrid with concatenation technique. The uncertainty in this model is given by:

1. Consider all possible outcomes.
2. Belief will lead to believe in some possibility by bringing out some evidence.
3. Plausibility will produce the evidence compatible with possible outcomes.

The concatenated neural network: cluster characteristics estimation for multi-modals was collected by multiple sources/sensors in offline mode. Dempster-Shafer neural network relies on the suggested belief and evidence. The subtraction of the redundant vector of feature classes detected in the first vector by the classes revealed on the second vector, which is equivalent to the initialization of our data fusion algorithm. Having computed the mass, plausibility and belief values for each simple and compound hypothesis of the multisource model, we design a criterion, which is entitled "decision rule," to determine which hypothesis is the additional "realistic." Recently, the option of this criterion stays system-dependent. The three most familiar decision rules are: 1) maximum of plausibility, 2) maximum of belief, and 3) maximum of belief without overlapping of belief intervals. Rule 1) is judged as the best by many researchers; maximum belief over the simple hypotheses is the most used; rule 3), entitled absolute decision rule, is very strict. It relies on weighted fusion neural network, belief, evidence, and plausibility as shown in Figure.11.

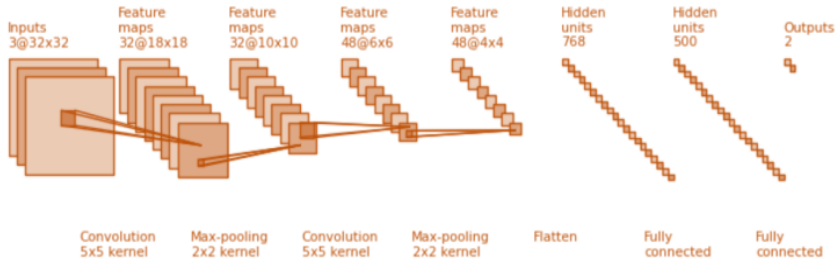


Fig. 11: The proposed hybrid fusion neural network

The importance of Dempster-Shafer theory DST is an evidence theory; it combines all possible outcomes of the problem. Hence it is used to solve problems where there may be a chance that different evidence will lead to some different result. That differs to Bayesian theory is only concerned about single evidence. Bayesian probability cannot describe ignorance.

So, the proposed solution makes a hybrid technique between concatenation and Demspter-Shafer to make the fusion in two levels with respect to the high level of classification decisions and the low level with interpreting some features with concatenation feature fusion technique. Concatenation is one of the popular fusion methods of building a joint representation of feature vectors from multiple modalities. Therefore, concatenation of the vectors occurs after the learners encode their

respective inputs. This is not an effective way to fuse multimodal inputs as modeling non-linear intermodal interaction becomes difficult in such scenarios. The concatenation process relies on the fusion type, is implemented the learning process for the desired entities as shown in Figure 4.20 and Figure 4.21. For example, if using early fusion, output vectors of individual learners are concatenated. For example, the fusion between two models that are classified with extracted many features in vectors. It can calculate the vector length and unify the length group. Then it makes the normalization of the feature vectors. The Output includes the fused feature vector from the two inputs. The feature fusion includes the Input,

$$D[\ ] = \{x1, x2, ..., xn\}; \quad (11)$$

$$W[\ ] = \{y1, y2, ..., yn\}; \quad (12)$$

The hybrid fusion approach is designed based on making two fusions in two different levels, high and low. It makes parallel fusions of the Demspster-shafer and concatenation then extracting the not important features and reducing these features as shown in Figure 4.9. The hybrid fusion layer refers to draw the full vision of the modals classification. That provides the unification target of multiple sensory data classifications in various smart environment systems.

The fusion techniques two types are:

- **Voting-based:** In the voting-based decision fusion techniques, majority voting is the most popular and is widely utilized. Some of the other techniques include weighted voting in which a weight to each classifier is attached and then decision fusion is proceed. There is a borda count that refers to another technique in which the aggregates of reverse degrees are computed to perform decision fusion. Other voting techniques are probability-based, such as fuzzy rules, Naïve-Bayes, Dempster-Shafer theory, and so forth.
- **Divide and conquer:** In this decision fusion technique, the training dataset is split to subsets of equal sizes, and then the classification is executed keep track of a decision fusion on the results of those lower dataset classifications. These divide and conquer functions contain the ideas of bagging and boosting.

The presentation of current research implements the majority voting of the different CNN-based pre-trained models with a parallel structure of the decision fusion technique.

4.5. Evaluation layer

This evaluation layer measures the accuracy and optimization results in multiple smart context systems. This layer classifies data into two types, training data and testing data. It measures the accuracy, precision, recall, F1-measures. It also presents a comparison between default optimization Bayesian and swarm particle optimization.

The measurements of proposed adaptive smart environment multi-modal system rely on two dimensions, accuracy and optimization measurements.

5. Experiments and Results

The adaptive smart environment multi-modal system makes the interpreting various features, conditions, and characteristics of each input modal. ASEMMS examines two smart environment contexts, smart military and smart health for four experiments with same of different modalities types. This section includes four experiments for evaluation and makes comparison between the baselines of pervious fusion techniques Demspter-shafer and Concatenation for validation the adaptive hybrid fusion technique.

5.1. Measurements

a) Accuracy results

The accuracy evaluation measurement computes the classification accuracy of various classification modals [36]. The classification accuracy refers to a number of relevant documents retrieved based on the total number of the existing relevant documents. Precision-Recall is a beneficial measurement of success of forecasting when the classes have a big variance.

Table 3: the tracing of infected people with no symptoms or mild symptoms

Predicted condition	**Conditions**	**True conditions**	
		PCR	Symptoms
	Predicted condition positive	TP	FP
	Predicted condition negative	FN	TN

The formula for measuring the accuracy (as shown in Table.3) is,

Precision (P) is known as the number of true positives (Tp) over the number of true positives plus the number of false positives (Fp) as illustrated in equation (13).

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

Recall (R) is known as the number of true positives (Tp) over the number of true positives plus the number of false negatives (Fn), as shown in equation (14).

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

These amounts are also related to the (F1) score, which is defined as the harmonic mean of precision and recall as mention in equation (15). These quantities are also regarding to the (F1) score, which is known as the harmonic mean of precision and recall.

$$F - measure = \frac{2\,Recall \cdot Precsision}{Recall + Precision} \quad (15)$$

Where: TP = True positive; FP = False positive; TN = True negative; FN = False negative.

b) Optimization

The optimization measures default optimization SGDM and ADAM optimizers. The objective of the optimization is optimizing the classification results. It applies three optimization functions,

- **SGDM** measurement default optimizer is based on working image, video, and audio SGDM/Gradient descent is an optimization algorithm that follows the negative gradient of an objective function in order to locate the minimum of the function. SGD with Momentum is a stochastic optimization function that inserts a momentum term to a regular stochastic gradient descent.
- **ADAM** optimization refers to an extension to stochastic gradient descent that has lately shown broader adoption for deep learning systems in computer vision (CV) and natural language processing (NLP). Adam measures the text optimizer. It evaluates the adaptive learning rates for each characteristic. Adam is a replacement optimization algorithm for stochastic gradient descent that makes the training deep learning models.
- **Bayesian Optimization:** improves classification accuracy results that are discussed in chapter 5. It is a computational solver function to optimize a classification problem by repetitive attempt to enhance a candidate solution with regard to a given quality estimation.
- **Swarm particle optimization (SPO)** enhances the quality based on mathematical formula. Particle swarm optimization (PSO) is a computational method to optimize a problem by iteratively to improve a candidate solution with regard to a given estimate of the quality. Each particle's movement is affected by its local best-known position, but is also made a brief toward the best-known positions in the search-space, which are upgraded as better positions are established by other particles.

5.2. Datasets

A) Dataset 1: Smart Military (Multi Images Modalities)

Objectives: This dataset aims to solve the problem of military classification objects in night mode based on multiple spectrums (as shown in Figure.12). Although researchers in [4] present a classification for military objects, they faced a lack of military datasets and a problem of classification objects in many spectrums.

Description: The thesis creates a new purified dataset collected and tuned from five benchmark datasets (TNO Image Fusion, Gun objects dataset, Multi-Spectral Images, Flir-Starter Thermal, and Terravic Weapon IR) in night mode. This dataset

includes six spectrums, Thermal, Long Wave Infrared (LWIR), Near Infrared (NIR), RGB, and DHV (VIS, NIR, 0) as shown in Table.4.

Table.4: The purified dataset is collected from the five mentioned datasets

Ref	Dataset	Description	Dataset Size
(Toet, A., 2014).	TNO Image Fusion	Visual (0.4–0.7μm), near-infrared (NIR, 0.7–1.0μm), and long-wave infrared (LWIR, 8–14μm).	579 images
(Sasank, S., 2019)	Gun objects dataset	Images dataset in real object photos	333 images
(BIIC, 2019)	multi-spectral images	Dataset includes 7 objects that are classification into 3 spectrums as the following, Visible (VIS), Near-Infrared (NIR), and Thermal spectrums.	420 images
(Teledyne Flir, 2015)	Flir-starter thermal	Images Thermal Datasets based on LiDAR sensor.	119,491 images
(Miezianko, R. 2018)	Terravic Weapon IR	Weapon detection and weapon discharge detection with thermal imagery	

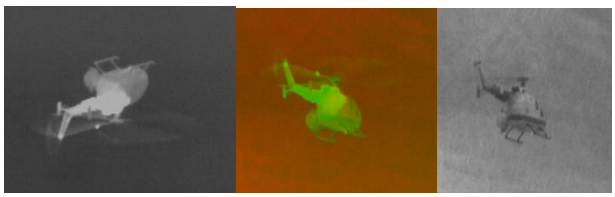


Fig.12: Examples of a helicopter in multi-spectral object detection

B) Dataset 2: Smart Military (Images and Videos Modalities)

Image is 11,131 images and Video is 30 files and applies data augmentation to achieve over 1 million images and video frames files (as shown in Figure.13). Video modal is interpreted based on required many libraries of image, color and it converting the video modal into images and count frames number and computes the image's sequence.

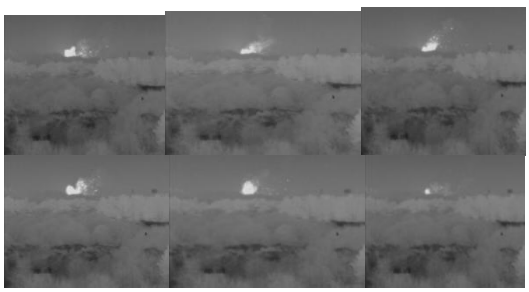


Fig. 13: The terrorist Fire sequenced video object recognition test in Thermal spectrum

C) Dataset3: Smart Health Dataset for cardio disease classifications

Objectives: This dataset aims to classify cardio diseases based on the sound of cough and metadata of text about patients.

Description: This dataset consists of a text sheet fused with cardio sound for improving the classification of COVID-19 disease from sound and patient metadata from the Kaggle website (Ulianova, S., 2019) and (Lisphilar, 2019).

Samples: It presents a solution for classifying Cardio disease and COVID-19 from cardio sound and meta-data about patients as mentioned in Table.5, whether normal or abnormal cases as shown in Figure.14.

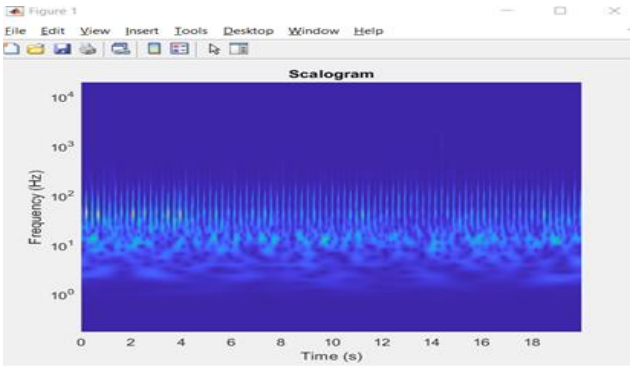


Fig. 14: A sample cough cardio sound converts to spectrogram image

Table 5: Smart health dataset description for cardio patients

Dataset	Ref	Description	Dataset size
Cardiovascular Disease dataset	(Ulianova, S., 2019)	Including 11 properties About Patients (it includes patient's profile age, height, weight, gender) with a medical profile)	70 000 records of patient's data
Respiratory Sound Database	(Lisphilar, 2019)	The annotation text files have four columns: Beginning of respiratory cycles, End of respiratory cycles, Presence of crackles, Absence of wheezes.	920Audio for patients

D) Dataset 4:

Objectives: This dataset aims to classify cardio diseases based on the sound of cough and metadata of text about patients.

Description: This dataset consists of a text sheet fused for improving the classification of COVID-19 disease from X-ray images and patient text metadata from the Kaggle website (Mooney, P., 2018) and (Maontoya, F.J. and Ledesma, M.M.,2020) as shown in Figure 15(a,b).

Data sizeA sample of predictive analysis for future data and time about 70.000 records regarding historical data. Image is 357 records and augmented by reflection,

rotation, scaling, share, cropping and add many noisy on the images that reach 8000 images.

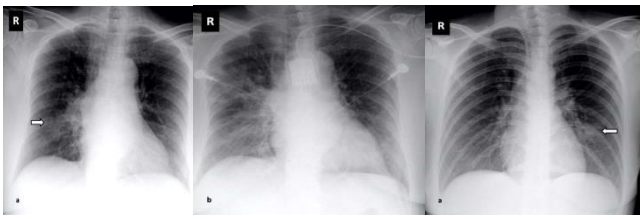


Fig.15 (a): A sample of Chest X-Ray Images for lung patients

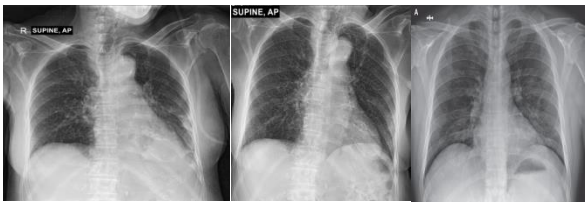


Fig.15 (b): A sample of Chest X-Ray Images for lung patients

5.3. Experiments

The experiments include different or same modalities types. They are applied on three comparisons between as the following,

a) The classification and fusion results

b) The adaptive hybrid fusion mechanism with the baseline of Demspter-shafer theory fusion technique and the baseline of concatenation fusion technique.

c) The optimization functions Bayesian optimizer and Swarm Particle optimizer for improving fusion accuracy results [47].

The experiments are four that are applied on datasets as mention in the same modalities,

- The experimental analysis for images only on military dataset
- And, in the different modalities,
- The experiment on fusing images and videos modalities on military dataset.
- The experiment on fusing Text & Audio modalities on smart health,
- The experiment on fusing Text & Images modalities in smart health.

The discussion of the experiments is shown as the following,

(A) Experiment 1:

That is designed based on dataset1 for multi images modalities as shown in Figure.16. the adaptive system uses for improving the classification results for multi spectrums.

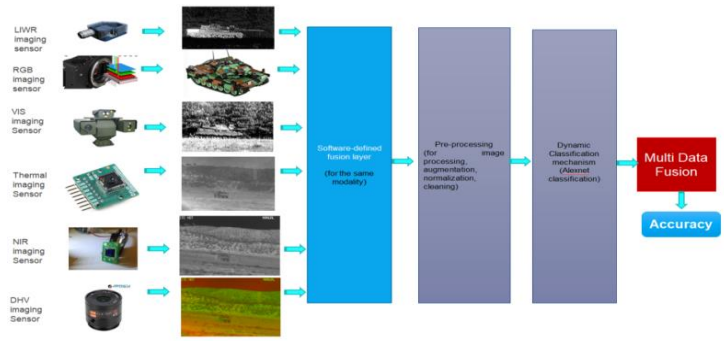
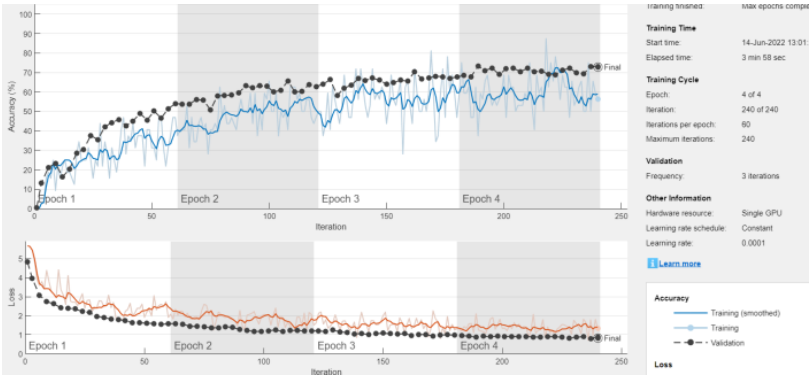
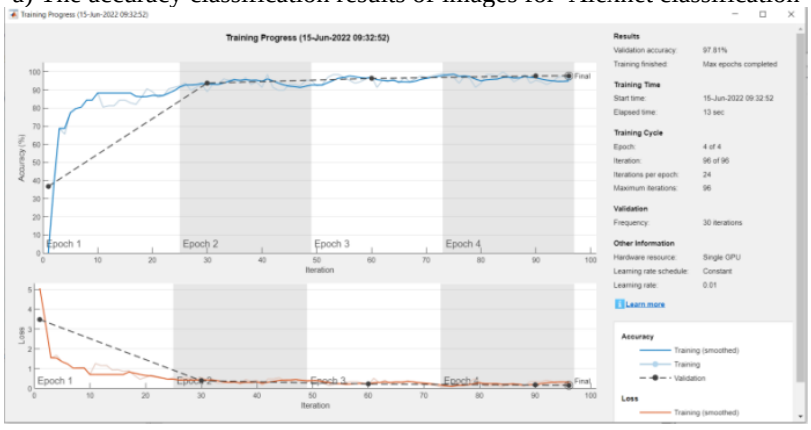


Fig.16: A smart military challenge architecture from six spectrums



a) The accuracy classification results of images for Alexnet classification



b) The accuracy classification results of images for hybrid fusion

Fig. 17(a,b): A comparison between pre-trained alexnet classification and the hybrid fusion for classification objects

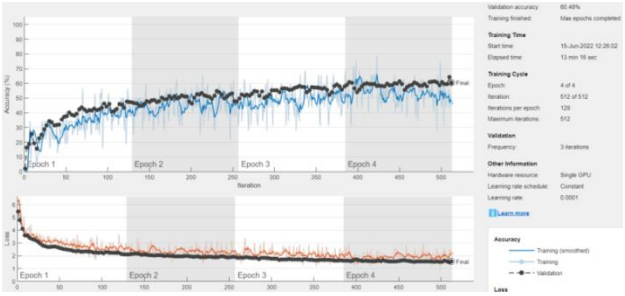
Table.6: A comparative study for the classification accuracy of the military spectrums in night mode.

Metrics /Model	NIR	VIS	Thermal	LIWR	DHV	RGB
Classification Accuracy	72.2	71.3	74.3	72.3	70.1	71.4
Fusion Accuracy	97.2	96.5	96.9	97.1	96.7	97.8

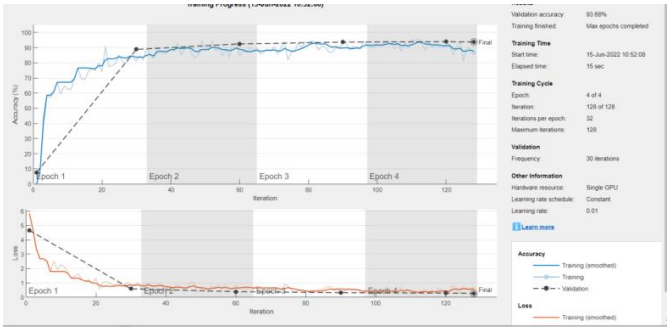
The classification accuracy results approximately 75% for two spectrums (NIR and VIS) and thermal spectrum reaches 89% approximately, the accuracy classification achieves to 91% for LWIR and DHV spectrums, RGB accuracy classification results achieves to ~ 92%. Table 5.2 presents the classified precisions, recall, and F1-score results for the accuracy computation in object detection and classification. The hybrid fusion is better than the normal statistical fusion approach that achieves to 98.6%. The adaptive smart-environment multi modal system improves the accuracy results that reach to 96%. The hybrid fusion improves the classification accuracy results 98.3% and the optimization reaches 97%.

(B) Experiment 2: smart military for images and videos modalities

The adaptive smart environment multi-modal system improves the classification accuracy by 97.4 % and enhances the optimization results using Bayesian optimizer by 98.7% and swarm particle optimization results by 99.1% for improving the classification accuracy results.



a) Classification of Videos classification objects



b) Accuracy Classification of hybrid fusion

Fig. 18(a,b): A comparison between pre-trained alexnet classification and the hybrid fusion for classification objects

Table.11: The accuarcy comparison results of hybrid fusion for object and videos classifications

Metrics /Model	Object Classification	Video Classification
Classification Accuracy	73.2%	63%
Fusion Accuracy	95.5%	96%

Table 12: The optimization comparison results of hybrid fusion for object and videos classifications

Metrics /Model	Optimization results
Bayesian optimizer	95.7%
Swarm optimizer	96.1%

(C) Experiment 3: smart health for audio and text modalities

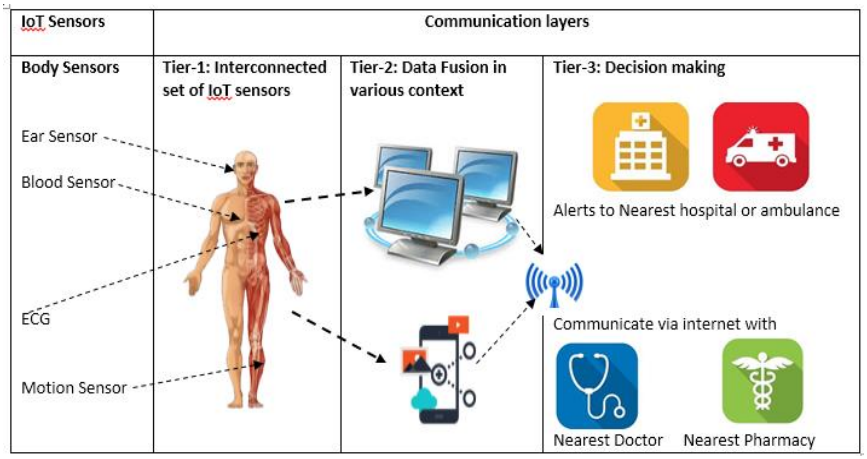


Fig. 19: Smart military for monitoring patients based on cough audio and metadata problem

The adaptive smart environment multi-modal system improves the classification accuracy by 96.7 % and enhances the optimization results using Baseyian optimizer by 97.8% and swarm particle optimization results by 98.8% for improving the classification accuracy results as shown in Table.13.

Table 13:The optimization results of fusion Graphs and precentages

Metrics /Model	Optimization results
Bayesian optimizer	97.8%
Swarm optimizer	98.4%

(D) Experiment 4: smart health for text and images modalities

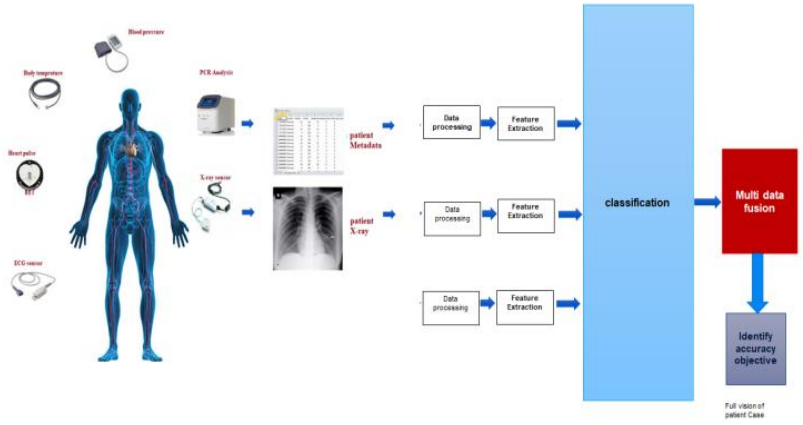


Fig. 20: Smart military for images and text problem

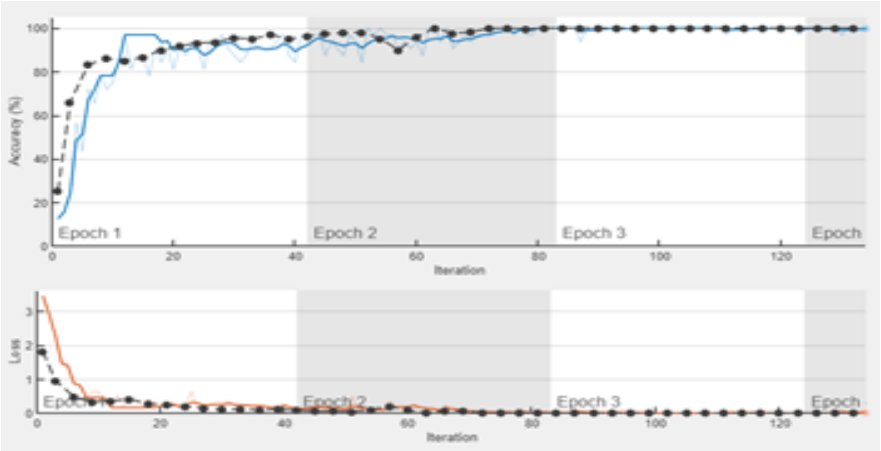


Fig. 21: A sample example of smart military classification

Table 14: The optimization comparison results of hybrid fusion for object and videos classifications

Metrics /Model	Optimization results
Bayesian optimizer	100%
Swarm optimizer	100%

The adaptive smart environment multi-modal system improves the classification accuracy by 99 % and enhances the optimization results using Bayesian optimizer by 100% and swarm particle optimization results by100% for improving the classification accuracy results as shown in Table.14.

6. Validation Experiments

This section develops a validation measurement with making a comparable analysis with two previous motivations. This comparative study depends on the adaptive hybrid fusion technique, the baseline of Demspter-shafer baseline (Chen, Q. et al. 2014), and the baseline of Concatenation (Gogate, M. et al. 2017). This comparison is applied on the accuracy results for classification for multi modalities.

6.1. Experiment 1: Smart military (multi-images modalities)

The accuracy classification comparison between hybrid fusion with 97% with the baseline of Dempster-shafer theory with 85% and the baseline of concatenation with 72%. In addition, the optimization of Bayesian optimizer for adaptive hybrid fusion mechanism reaches 98.2% and for the baseline of Dempster 86%, and for the baseline of concatenation reaches 72.8%. However, the swarm-particle optimizer improves the adaptive hybrid fusion with 2% to achieve to 99.4% and 86% for Dempster-shafer and 75 for Concatenation.

Table 15: The accuracy classification comparison between hybrid fusion with the baseline of Dempster-shafer theory and the baseline of concatenation

	Classification Accuracy results
Adaptive hybrid fusion	97.4%
Baseline Dempster-shafer	85%
Baseline of concatenation	72%

Table 16: A comparative study between Bayesian and swarm-particle optimizers on three techniques

	Bayesian optimizer	Swarm-particle optimizer
Adaptive hybrid fusion	98%	99.4%
Baseline dempster-shafer	85%%	86%%
Baseline of concatenation	72%	75%

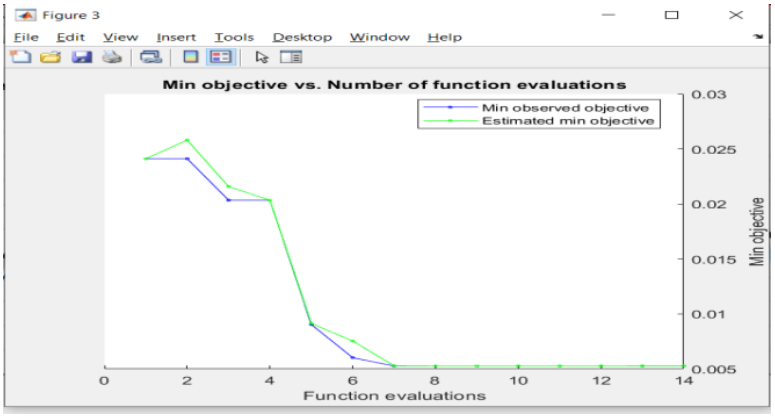


Fig. 21: Comaprison between the adaptive hybrid fusion and the baseline of Dempster-shafer and the baseline of Concatentation

6.2. Experiment 2: Smart military (images & videos modalities)

The accuracy classification comparison between hybrid fusion with 94.2% with the baseline of Dempster-shafer theory with 92% and the baseline of concatenation with 62%. In addition, the optimization of Baseyain optimizer for adaptive hybrid fusion mechanism reaches 95.2% and for the baseline of Dempster-shafer 86%, and for the baseline of concatenation reaches 72.8%. however, the swarm-particle optimizer improves the adaptive hybrid fusion with 2.2% to achieve to 94.2% to 97.4%, Dempster-shafer achieves to 87% and 82% for the Concatenation.

Table 17: A comparison between hybrid fusion with the baseline of Dempster-shafer theory and concatenation

	Classification Accuracy results
Adaptive hybrid fusion	94.2%
Baseline dempster-shafer	92%%
Baseline of concatenation	62%

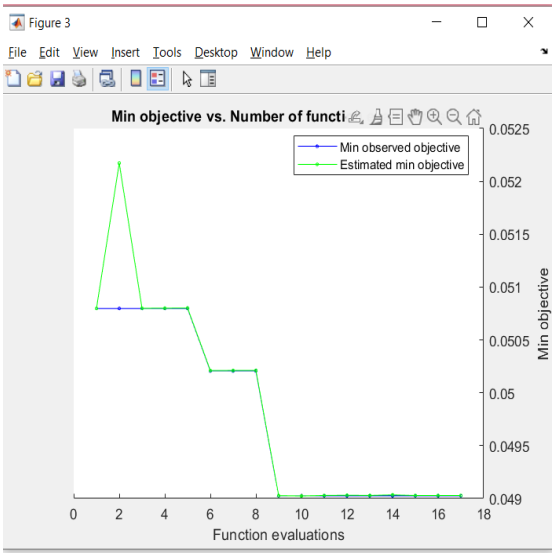


Fig. 22: Tracing for best fit optimizer

Table 18: A comparison between hybrid fusion with the baseline of Dempster-shafer theory and concatenation

	Bayesian optimizer	Swarm-particle optimizer
Adaptive hybrid fusion	95.2%	97.4%
Baseline dempster-shafer	85%%	85%%
Baseline of concatenation	80%	80%

6.3. Experiment 3: Smart health (text &audio modalities)

The accuracy classification comparison between hybrid fusion with 94.2% with the baseline of Dempster-shafer theory with 92% and the baseline of concatenation with 62%. In addition, the optimization of Baseyian optimizer for adaptive hybrid fusion mechanism reaches 97.2% and for the baseline of Dempster-shafer 84%, and for the baseline of concatenation reaches 65%. However, the swarm-particle optimizer improves the adaptive hybrid fusion achieves to 98.4% Dempster-shafer achieves to 85% and 66% for the Concatenation.

Table 19: A comparison between hybrid fusion with the baseline of Dempster-shafer theory `and concatenation.

	Bayesian optimizer	Swarm-particle optimizer
Adaptive hybrid fusion	97.8%	98.4%
Baseline dempster-shafer	84%%	85%%

Baseline of Concatenation	65%	66%

6.4. Experiment 4: Smart health (text & images modalities)

The accuracy classification comparison between hybrid fusion with 99% with the baseline of Dempster-shafer theory with 95% and the baseline of concatenation with 92%. In addition, the optimization of Baseyian optimizer for adaptive hybrid fusion mechanism reaches 100% and for the baseline of Dempster-shafer 96%, and for the baseline of concatenation reaches 93%. however, the swarm-particle optimizer improves the adaptive hybrid fusion achieves to 100%, Dempster-shafer achieves to 98% and 94% for the Concatenation.

Table 20: A comparison between hybrid fusion with the baseline of Dempster-shafer theory and concatenation

	Classification Accuracy results
Adaptive hybrid fusion	99%
Baseline Dempster-shafer	95%
Baseline of concatenation	92%

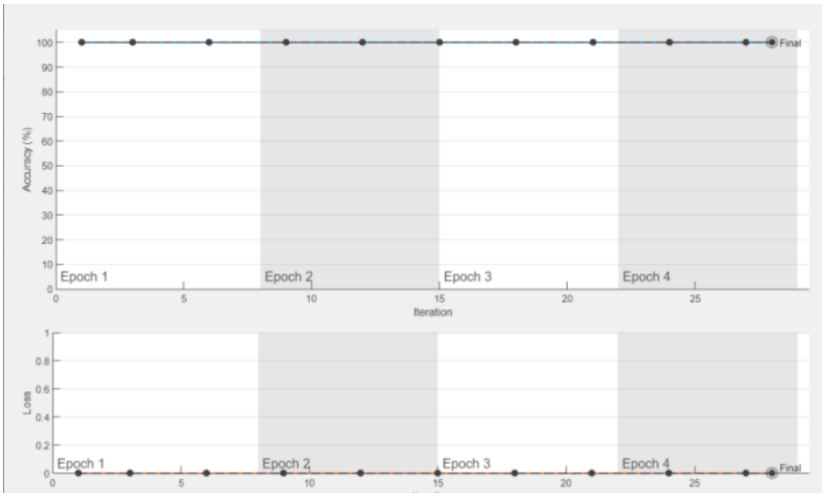


Fig. 23: Tracing for hybrid fusion accuracy results

Table 21: A comparative study between Bayesian and swarm-particle optimizers on three techniques

	Bayesian optimizer	Swarm-particle optimizer
Adaptive hybrid fusion	100%	100%

Baseline dempster-shafer	96%%	98%
Baseline of concatenation	93%	94%

7. Discussion

Main advantages of the adaptive smart environment multi-modal system are shown as the following,

- The benefit of software defined fusion layer is a controller layer and interpreting relationships between modalities input types. It also can deal with multiple input from 1 to 16 sources based on 2^n by 4*4 modalities inputs. In addition, new relationships between modalities input datasets for making consistency and interpreting better meaning.
- The advantage of pre-processing layer is bigger size of input datasets, normalized and tuning them.
- The powerful of a dynamic classification deal with input data as vectors. It improves feature extraction and improves Multi classification objects. This method is different from any specific classification system based on specific condition's number. Although a dynamic classification layer takes longer time. It reaches higher accuracy of classification results and finds new affected parameters and features in the same domain. For example, classifying patient disease depends on two conditions that are the age and smoked or not from the metadata patient sheet input. The classification is different in the result from the specific system when our proposed dynamic classification layer finds a new affected feature as location due to the proposed layer checks on all parameters relationships with themselves and finds the most affected on themselves.
- The hybrid fusion layer benefit is improving classification results with respect enlarging number of parameters/features and finding many relationships between them. A hybrid fusion layer benefits from tailored neural network for improving the plausibility of the event based on evidence and belief and getting big number of features for enhancing the accuracy classification results.
- Our system presents a balance of the relationship between complementation and redundancy of multimodal. It makes an optimization of the tradeoff between the exemplification capability and complexity of the fusion model. It reaches fine interactions between multiple modalities. In addition, our technique applies on the different fusion experiments in smart environments for improving classification results. It is designed based on Demspter-shafer belief function and Particle Swarm Optimization (PSO) for reaching the best accuracy for full vision of multiple sources for supervised learning. In

order to measure the performance of the proposed method, multiple indicators are identified as illustrated below, the ability to solve all problem types of various contexts. The ability selects projects with the lowest uncertainty. The solving large number of parameters challenge.

For validation the proposed system, we make a comparison between Proposed AHF and Baseline of Dempster-shafer to shown the superiority of the proposed function in making solution for the problems while uncertainty exists. It changes input data to reach higher-level of the abstraction. Hybrid fusion acquires knowledge of a joint exemplification of diverse modalities.

The validation measurement makes a comparable analysis with two previous motivations. The validation of the proposed adaptive smart environment multi-modal system relies on making two comparisons between the proposed system and two previous systems based on verified dataset.

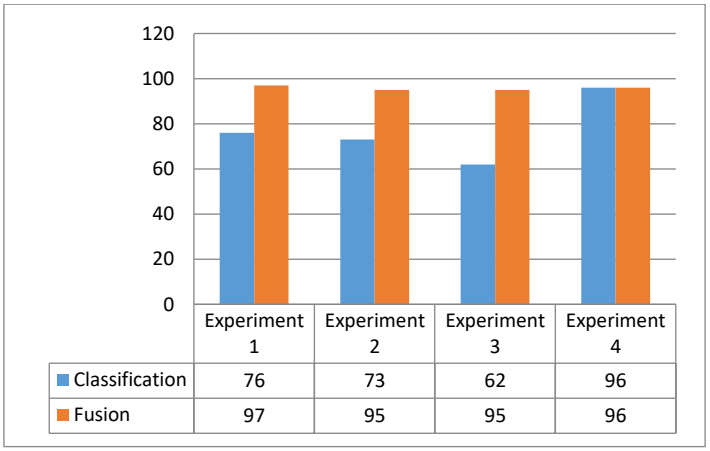
A. The Average of adaptive system results improves the accuracy classification results by 20%.

B. It compares to the baseline Dempster-shafer technique by 10 %.

C. And the baseline of concatenation technique by 30%.

For validation our adaptive system, we make a comparison between the baseline of Concatenation fusion technique, the baseline of Dempster-shafer fusion technique, and our Adaptive hybrid fusion technique. We apply this comparison on the previous two experiments and the results achieve to 97% to 99 %.

We make a comparison the hybrid fusion technique presented in this research to the baseline of concatenation and Demspter-shafer techniques for supervised data fusion (as shown in Figure 24, 25, and 26).



	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Classification	76	73	62	96
Fusion	97	95	95	96

Fig. 24: The behavior analysis of pre-trained classification results and classification fusion results in many experiments for various modalities inputs

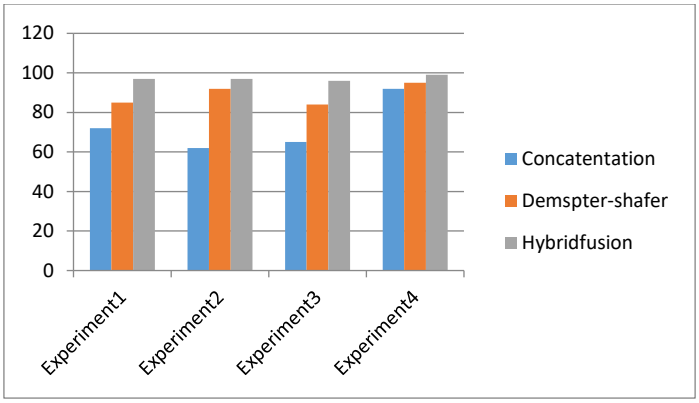


Fig. 25: The behavior analysis of fusion techniques in many experiments for various modalities inputs

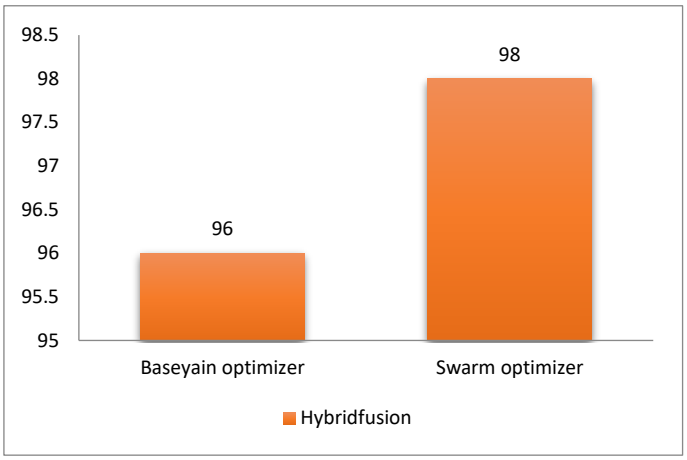


Fig. 26: The behavior analysis of hybrid fusion for various modalities inputs

8. Implementation

The implementation of the constructed decision technique relies on MATLAP 5. The implementation was testing on CPU and multi-graphical processing unit (GPU) specific. GPU is constrGPU results are faster, better performances and higher accuracies.

9. Conclusion and Future works

Multi-source information fusion is a complex estimation process that allows users to more accurately assess complex situations by effectively combining key evidence in large, diverse, and sometimes conflicting data from multiple sources. It implicates inference information from these sources to generate standardized, specific, and comprehensive measures about an entity, activity, or event.

This thesis presented an adaptive smart environment multi-modal system that is a solution for the modality and context-aware interpretation problems and to understand extracted big data from intelligent devices fully. The generated data hold big, heterogonous, and complex data. The context-aware refers to constructing each domain system based on understanding parameters, conditions, and data types. The modality is interpreted into fusing multiple data sources with the same data type and fusing multiple data types from the same source in various smart environments. This thesis presents an information system for solving smart systems construction challenges in "context-ware" and "modality" challenges. It introduces an adaptive smart environment multi-modal system for improving the accuracy and optimization results. This adaptive system relies on a combination of deep neural networks and hybrid fusion techniques for interpreting various data types or data characteristics. The interpretation of multi-modals input depends on the weight of each input. There are two types interpretation for the inputs, the same weights by default, and dynamic choice based on the user selected. The adaptive system consists of five layers, software-defined fusion layer, pre-processing layer, dynamic classification layer, hybrid fusion layer, and evaluation layer. A software-defined layer is considered a controller layer for data cleaning, pre-processing, and managing data characteristics. It also handles the outliers and noisy data. It is based on three dimensions, data type, the importance of extraction features, and detecting anomalies in the data. A Software-defined fusion layer plays a vital role in controlling fusion processes based on multiple data types and multiple context-aware. It aims to high accuracy and performance results for multiple data sources with multiple parameters. A dynamic classification layer is an automated classification layer for choosing the appropriate neural network based on the input data types. This layer uses the suitable classification based on the input data type. A hybrid fusion layer consists of a hybrid between two fusion techniques parallel that are Dempster-shafer and concatenation statistical techniques. The powerful of this hybrid is shown in improving the belief probabilities of classified results with decision fusion and the detailed features in fused vectors for the concatenation technique. The hybrid is also useful due to reducing not all features in the fused vectors. The important fused features in vector draw the full vision of classifications. It is used for improving the classification accuracy and prediction results. It is designed based on a hybrid fusion between neural networks and Dempster-shafer statistical theory based on the evidence and proof for fusing multiple spectrums in night vision. Dempster-shafer is also known as by the "evidence theory" or the "belief function theory" is a formal system for reasoning with partial, uncertain and imprecise information. The basic ideas of evidence theory are belief architecture, belief and plausibility methods. Concatenation is converting features which interpreting into vectors for one vector with unification target. This hybrid layer can interpret a big number of features without redundant between fusion techniques based on check similar between

redundant vectors and remove it for drawing the full features in two levels of fusion. An Evaluation layer relies on two parts, Part one is evaluating the accuracy and optimization results of multiple smart systems. Part two is a comparative analysis with two previous motivations. The fusion output graphs the fusion accuracy for classification results from 96% to 98% in various smart systems. The optimization of classification results are % and %. The experiments are applied in multiple contexts as smart military and smart health.

The experiments are applied to evaluate validation results based on two parts,

Part 1, Evaluation layer: includes four experiments with four datasets in smart military and smart health contexts. Experiment#1: Smart military for images modalities input that is applied for improving the full vision of accuracy results. It is suitable in other smart domains such as Smart military that can fuse multiple spectrums and classify objects and actions from images. Dataset size includes 11.320 that are increased by augmentation functions reaching 875,970 images. The adaptive smart environment multi-modal system improves the classification accuracy by 97.4 % and enhances the optimization results using Bayesian optimizer by 98.7% and swarm particle optimization results by 99.1% for improving the classification accuracy results. Experiment#2: smart military for images and videos modalities input that is applied for improving the classification accuracy for military objects. The adaptive smart environment multi-modal system improves the classification accuracy by 94.7 % and enhances the optimization results using Bayesian optimizer by 96.8% and swarm particle optimization results by 97.8% for improving the classification accuracy results. Experiment#3: in Smart health, the challenge of monitoring COVID-19 patients remotely based on classifying multivariate objects and regression analysis based on various data types inputs (text and cough audio). Dataset size includes Text is 70.000 records and normalized them. Audio is 920 files and applies data augmentation to achieve files. The adaptive smart environment multi-modal system improves the classification accuracy by 20% that achieves to 96.6%. It enhances the optimization results using ADAM by 96.6%. The proposed adaptive system improves the optimization results by 97.7 % for improving the classification accuracy results. Experiment#4: in smart health, the objective is health classification objects. Dataset size includes Text modal is interpreted based on a sample of predictive analysis for future information and time about 70.000 records concerning historical data. Image is 357 records and augmented by reflection, rotation, scaling, share, cropping and add many noisy on the images that reach 8000 images. Image modal is interpreted based on required many libraries of image, color and it converting the video modal into images and count frames number and computes the image's sequence. The adaptive smart environment multi-modal system improves the classification accuracy by 99 % and enhances the optimization results using Bayesian optimizer by 100% and swarm particle optimization results by100% for improving the classification accuracy results.

Part 2 introduces the validation which measures the quality of the proposed adaptive smart environment multi-modal system. It relies on making two comparisons between previous motivations researches. Validation is applied to two experiments and compared between them. We compare the hybrid fusion algorithm presented in this research with the baseline of concatenation and Demspter-shafer techniques for supervised data fusion.

The Average of adaptive system results improves the accuracy classification results by 20%. It compares to the baseline Dempster-shafer technique by 10 % and the baseline of concatenation technique by 30%. For validation our adaptive system, we make a comparison between the baseline of Concatenation fusion technique, the baseline of Dempster-shafer fusion technique, and our Adaptive hybrid fusion technique. We apply this comparison on the previous two experiments and the results achieve to 97% to 99 %.

For future works, Smart data aims to filter noise data and generate valuable information that can be effectively utilized to businesses and governments to plan, operate, monitor and control and make smart decisions. Although an unprecedented amount of data can be made available as advanced data merging technologies advance, the key is to explore how big data can become smart data and deliver intelligent information. Developed big data modeling and analytics are necessary to discover embedded data infrastructures and getting more intelligent data. There is an open research in using multimodal in Visual Question Answering (VQA) has become a hot topic in computer vision. A crucial solution to VQA exists in how to fuse multi-modal features extracted from images and questions. In this paper, we present that integrates visual relationships and attention reaches more fine-grained feature fusion. Particularly, we construct an effective and efficient module to reason the complex relationship between visual objects.

References

Abidin, R. Z., Shukri, S. A., & Arshad, H. (2017)Adaptive multimodal interaction in obile augmented reality: A conceptual framework. *Journal of Telecommunication, Electronic and Computer Engineering*, 9 (2-11), 97-103.

Ahmad, J., Muhammad, K., Kwon, S-i., Baik, S.W., & Rho, S. (2016). Dempster-Shafer fusion based gender recognition for speech analysis applications. *2016 International Conference on Platform Technology and Service (PlatCon)*, Jeju, South Korea, 1-4.

Almasri, M. & Elleithy K. (2015). Data fusion in WSNs: Architecture, taxonomy, evaluation of techniques, and challenges. *International Journal of Scientific & Engineering Research*, 6(4).

Akbari, H., Yuan, L., Qian, R., Chuang, W. -H., Chang, S. -F., Cui, Y., & Gong, B. (2021). Vatt: transformers for multimodal self-supervised learning from raw video. audio, and text, arXiv:2104.11178v1 [cs.CV].

Alberti, A. M., Santos, M. A. S., Da Silva, H. D. L., & Souza, R. (2019). Platforms for smart environments and future internet design: A survey. *IEEE Access*, (4), 1-33.

Almasri, M. & Elleithy, K. (2015). Data fusion in WSNs: Architecture, taxonomy, evaluation of techniques, and challenges. *International Journal of Scientific & Engineering Research*, 6 (4).

Atzori, L., Lera, A., Morabito, G., & Nitti, M., (2012). The social internet of things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization. *Computer Networks Journal*, 56 (16), 3594-3608.

Biometrics and Identification Innovation Center-BIIC. (2019). Multispectral Dataset. available online on : https://biic.wvu.edu/data-sets/multispectral-dataset

Che, C., Wang, H., Ni, X., & Lin, R. (2020). Hybrid multimodal fusion with deep learning for rolling bearing fault diagnosis, Measurement.

Chen, Q.,Whitbrook, A., Aickelin, U., & Roadknight, C. (2014). Data classification using the Dempster-Shafer method. *Journal of Experimental & Theoretical Artificial Intelligence*. 26 (4), 493-517.

Chiu, C. -K., Tseng, J. C. R., & Hsu, T. -Y. (2017). Blended contextaware ubiquitous learning in museums: Environment, navigation support and system development, *Personal and Ubiquitous Computing*, 21(2), 355–363.

Dumas, B., Lalanne, D., & Oviatt,S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. In: Lalanne D, Kohlas J, editors. Human Machine Interaction: Research Results of the MMI Program (chapter 1). Springer Berlin Heidelberg, 5440, 3-26.

Durkan, C., Storkey, A., & Edwards, H. (2018). The context-aware learner, *ICLR*, 1-18.

Emmanouilidis, C., Koutsiamanis, R. -A., & Tasidou, A. (2013). Mobile guides: Taxonomy of architectures, context awareness, technologies and applications. *Journal of Network and Computer Applications*, 36 (1), 103-125.

Fierrez, J., Ortega-Gracia, J., Gracia-Romero, D., &Gonzalez-Rodriguez, J. (2005). Bayesian adaptation for user-dependent multimodal biometric authentication, *Pattern Recognition*, 38(8), 1317-1319.

Gaw, N., Yousefi, S., & Gahrooei, M. R., (2021). Multimodal data fusion for systems improvement: A review. *IISE Transactions, Taylor & Francis*, 1-19.

Gogate, M., Adeel, A., & Hussain, A. (2017). A novel brain-inspired compression-based optimised multimodal fusion for emotion recognition. *IEEE Symposium Series on Computational Intelligence (SSCI)*. Honolulu, HI,USA,1-7.

Gumawardama, A. & Shani, G. (2009), A survey of accuracy evaluation metrics of recommendation tasks, *Journal of Machine Learning Research*, 10, 2935-2962.

Hasanov, A., Laine, T., & Chung, T. -S. (2019). A survey of adaptive context-aware learning environments. *Journal of Ambient Intelligence and Smart Environments*, 11, 403–428.

https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database

Lisphilar, (2019). COVID-19 dataset in Japan Dataset. available online on:

https://www.kaggle.com/datasets/lisphilar/covid19-dataset-in-japan

Khoie, M. R., Tabrizi, T. S, Khorasani, E. S., Rahimi, S., & Marhamati, N. (2019). A hospital recommendation system based on patient satisfaction survey. *Applied Sciences*, 7(10), 2076-3417.

Kim, J. I., Park, I. W., & Lee, H. H., (2011). An intelligent context-aware learning system based on mobile augmented reality. UCMA2011LUbiquitous Computing and Multimedia Applications, International Conference on Ubiquitous Computing and Multimedia Applications,Communications in Computer and Information Science book series (CCIS), 151, 255-264.

Kuang, S. & Davison, B. D (2017). Learning word embeddings with chi-square weights for healthcare tweet classification. *Applied Sciences*, 7(8), 846, 1-12.

Liu, K., Li, Y., Xu , N., & Natarajan, P. (2018). Learn to combine modalities in multimodal deep learning. arXiv:1805.11730v1 [stat.ML], 1-15.

Liu, Y. -T., Pal, N. R., Marathe, A. R., & Lin, C. -T, (2016). Weighted fuzzy dempster-shafer framework for multi-modal information integration, *IEEE*, 1063-6706.

Maontoya, F. J. & Ledesma, M. M. (2020), Performance evaluation of the particle swarm optimization algorithm to unambiguously estimate plasma parameters from incoherent scatter radar signals. *Martínez-Ledesma and Jaramillo Montoya, Earth, planets, and space*, 72:172.

Mezai, L. & Hachouf, F. (2016). Adaptive multimodal biometric fusion algorithm using particle swarm optimization and belief functions. *IEEE 2016 4th International Conference on Biometrics and Forensics (IWBF)* - Limassol, Cyprus, 1-6.

Miezianko, R. (2018),Terravic weapon IR Dataset- weapon presence/discharge detection. Available online on: http://vcipl-okstate.org/pbvs/bench/

Mujtaba, E. Y. & Elmustafa, S. A. A. (2019). Internet-of-things in smart environment: Concept, applications, challenges, and future directions. *World scientific news : An international scientific journal*, WSN 134(1), 1-51.

Mooney, P. (2018). Chest x-ray images (pneumonia) Dataset. available online on: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

Nasr, M., Islam, M., Shehata, S., Karray, F., & Quıntana Y. (2021). Smart healthcare in the age of AI: Recent advances, challenges, and future prospects arXiv:2107.03924 [cs.CY], 1-24.

Ortega, J. D. S., Senoussaoui, M., Granger, E., Pedersoli,M., Cardinal, P.,& Koerich, A. L. (2019). multimodal fusion with deep neural networks for audio-video emotion recognition,Xiv:1907.03196v1 [cs.CV].

Panda, R., Chen, C. -F., Fan, Q., Sun,X., Saenko, K., Oliva, A., & Feris, R. AdaMML: Adaptive multi-modal learning for efficient video recognition, ICCV,Virtual, Boston University, MIT-IBM Waston AI Lab, 7576-7585.

Raun, N. F. (2016). Smart environment using internet of things (IOTS) - A review. *IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*.

Roy, C. T., Lakshmi, D. S., Kumar, G. A., & Vishwas, H. N. (2017). Smart environment using IoT. In: *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE*, 1232–1237.

Sahu,S. & Vechtomova, O. (2021). Adaptive fusion techniques for multimodal data. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 3156–3166.

Sasank, S. (2019). Gun dataset. Available online on: https://www.kaggle.com/datasets/issaisasank/guns-object-detection

Shvetsova, N., Chen, B., & Rouditchenko, A. (2021). Everything at once– multi-modal fusion transformer for video retrieval, arXiv:211.04446v1 [cs.CV], 1-15.

Snidaro, L., Gracia, J., & Llinas, J. (2015). Context-based information fusion: A survey and discussion. *Information Fusion*, 25, 16-31.

Statista, (2016). Available online on: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

Suk, H. -II., Lee, S. -W., & Shen, D., (2016). Initiative, deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain Struct Funct*. 221(5), 2569–87.

Surve, A. R. & Ghorpade, V. R. (2017). Pervasive context-aware computing survey of context-aware ubiquitous middleware system. *International journal of engineering research and technology*, 10(1), 411-415.

Teledyne Flir (2015). FREE Teledyne FLIR Thermal Dataset for Algorithm Training. Available online on: https://www.flir.eu/oem/adas/adas-dataset-form/

Toet, A. (2014). TNO image fusion dataset. Available online on: https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029

Ulianova, S. (2019). Cardiovascular disease dataset. Available online on: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

Vailshery, L. S. (2022). Available online on: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

Vbookshelf, (2019). Respiratory Sound Database Dataset. Avaiable online

Yang, Z., Yu, W., Liang,P., Guo, H., Xia, L., Zhang, F., Ma, Y., & Ma, J. (2019). Deep transfer learning for military object recognition under small training set condition. *Neural Computing and Applications, Springer*, 31, 6469–6478.

Zhao, S., Gong, M., Fu, H.,& Tao, D. (2021). adaptive context-aware multi-modal network for depth completion. *IEEE Transactions On Image Processing*, 30, 5264-5275.

Zhou, B., Wan, J., Liang, Y., & Guo, G. (2021). Adaptive cross-fusion learning for multi-modal gesture recognition. *Virtual Reality &Intelligent Hardware*, 3(3), 235-247.