Object Detection in Dynamic Environmental Conditions using Evolutionary Multimodal Approach

Thesis by

Anup Vasant Kale

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy in Software Engineering

Submitted to the Graduate School of

University of Technology Sydney

University of Technology Sydney

New South Wales, Australia

31st January 2018

Dedication

I dedicate this thesis to my PhD supervisor Dr Zenon Chaczko for his support, guidance and a genuine reflection on my work. Without his encouragement and appropriate directions, this work would not have shaped up.

I also dedicate this thesis to my wife Manasi and my son Arnav for all along support during this long journey. Their sacrifices throughout these years allowed me to focus on this work.

Declaration of Authorship and Originality

I, Anup Vasant Kale, state that this thesis was authored by me, and that no material has been used in part or whole from other origins without full acknowledgement. All designs, results and theories of others that have been used in my thesis are referenced, and that assistance sources are recognised appropriately.

Signature of Doctoral Candidate:

Production Note:
Signature removed prior to publication.

Acknowledgements

I would like to sincerely thank my supervisor Dr Zenon Chaczko for his continuous support for my work in Multimodal Object Detection domains. Also, I give my regards to co-supervisors Prof. Robin Braun and Dr Perez Moses for their scholarly guidance. Also, I would like to appreciate all the help in organising this report by Dr Chris Chiu. I would like to thank all friends from the University of Technology Sydney including Pranav Kamod, Dr Pakawat Pupatwibul and all the researchers who were part of this journey with me. Finally, I would like to express my respect and love towards my family members including my mother and my in-laws who were always a high motivation for me.

Abstract

Environmental dynamism and uncertainty can play a critical role in many problems involving camera-based detection of real-life objects. Uncertainty is witnessed due to the presence of climatic irregularities including illumination changes, smoke, heat-waves, dust and rain. In such scenarios, the visibility of an object can severely be influenced by both signal-noise and occlusion. With the recent developments in sensing technology and computing domains, it is still possible to overcome the shortcomings of uncertainty. Multimodal image processing techniques provide very encouraging results by reducing noise and improving visibility. However, the multimodality needs further improvements to enhance accuracy, performance and robustness. Here, an evolutionary multimodal method is proposed to succeed over the discussed limitations. An evolutionary biological inspiration is applied to create a set of computing models. The proposed set of innovative evolutionary algorithms allows to reduce redundancies in datasets and improve the detection process. Experimental validation is performed for testing proposed algorithms. A formal simulation method for data modelling process was incorporated in the testing scheme to emulate environmental variations. Rigorous experimenting and analysis show the merits of the proposed methodology. Notably, both the accuracy and performance can be improved significantly using the proposed evolutionary apparatus.

Contents

List of Algorithms

List of Figures

List of Tables

Nomenclature

A

AI - Artificial intelligence is intelligence displayed by machines, in contrast with the natural intelligence displayed by animals.

ANN - Artificial Neural Network a concept inspired from biological neurons and used as a classification tool.

B

Bayesian Inference - A statistical inference which uses Bayes' Theorem to update the current probability for a hypothesis.

C

Classifier - A statistical classification algorithm in a software or computing engineering context used for analytical purposes.

D

DCA - Discriminant Classification Analysis is a type of classifier where groups of populations are known a priori and observations are classified into known populations.

Decision Trees - Decision Trees have a tree-like structure to model decisions and their possible consequences.

Deep Learning - A neural network variation which uses multiple layers of neurons to perform a heuristic analysis.

Dempster-Schafer - Dempster-Schafer approach is an evidence-based mathematical theory which is a generalization of Bayesian inference theory.

DWT - Discrete Wavelet Transform is a technique uses a set of small waves or wavelets to represent a signal.

E

EA - Evolutionary Algorithm is a subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm.

EC - Evolutionary Computing or Computation is a family of algorithms for global optimization inspired by biological evolution.

ED - Environmental Dynamism is defined as any noticeable change in an environment due to natural or artificial causes.

G

GA - Genetic Algorithm is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms.

H

Heuristic - Heuristic is an approach to problem-solving that employs a practical method not guaranteed to be optimal, but sufficient for the immediate goals.

I

IEEE - The Institute of Electrical and Electronics Engineers is a non-profit professional association, fostering technical research and knowledge amongst academia and the engineering industry.

Inference - A conclusion calculated and thus determined based on the evidences collected from various cues.

K

KNN - K Nearest Neighbours is an algorithm very commonly used as a classification tool, a non-parametric method used for classification and regression.

L

LDA - Linear Discriminant Analysis is a statistical classification algorithm categorized under supervised machine learning.

Learning Classifier - A rule-based machine learning where the classifier learns from training datasets.

M

Multimodality - Multimodality provides multiple sources of data-channels for the same target for improved sensory perception.

N

Neural Network - A network of Biological neurons in the brain where every neuron represents an individual processing unit.

P

PCA - Principal Component Analysis is a computing procedure using orthogonal transformation to convert an observation set into a value set of linearly uncorrelated variables called principal components.

S

SVM - Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

Part I

Theory, Concepts and State of the Art Methods

1 Introduction

In many real-life object detection problems, accuracy is restricted due to uncertain and dynamic operating environments [38, 39, 47, 76]. In general, Environmental Dynamism (ED) can be defined as any noticeable change in an environment due to natural or artificial causes. Principle factors causing ED include natural variations, weather changes, pollution, demographic factors and so forth. Circumstances causing ED may be predictable or unpredictable. The after effects of ED contain uncertain environmental conditions. Handling dynamism and uncertainty in optimisation without losing essential details is one of the challenging research areas. Dynamic Environment is also unpredictable and devoid of patterns and regularities. Thus, the detection process of moving peak benchmarks in the environments involving dynamism is a computationally intensive and complicated task.

1.1 Problem Context

Few approaches for handling such stochastic situations are multimodal-sensor-fusion, heuristic algorithms and parallel-processing [43, 105, 107, 139, 130]. In the case of computer vision applications, environmental variations may consist of dynamism in illumination, weather changes (rain, fog, heat waves), dust particle density, smoke and so on. Few of the serious implications due to dynamic environmental conditions include noisy images, blurred images and occlusion of objects. In all the above cases, image quality is deteriorated with time and needs an adaptive selection in the image analysis software system for an object detection process. Apart from the algorithmic work, multimodality of the data acquisition system is an essential criterion to achieve better accuracy, stability and performance.

1.1.1 Effects of Uncertain and Dynamic Environments on Image Data Quality

Environmental Dynamism can affect the image quality in one or many ways. Image sharpness is a measure of the amount of detail conveyed to the observer. Image sharpness can be changed due to variability and incompleteness in environmental illumination. Variable climatic conditions can alter the visibility of an object due to foggy conditions, rainfall, snowfall, mirage, smoke and dust particles above specific density. Any clarity affecting an environmental parameter can also occlude the object entirely or partially. Other effects of climatic variations include a change in contrast level, impact on colour accuracy and distortion of the image.

1.1.2 Evolutionary Multimodal Detection

An evolutionary multimodal computational approach is proposed in this thesis to improve accuracy and reliability of object detection in dynamic environmental conditions. The proposed method can be applied either at modality level or feature level to reduce redundancy and noise in multimodal datasets. In this process, evolutionary biological concepts like genotypes and phenotypes are used to model these datasets and adapt them to changed environmental conditions by applying selection rules are used. The selection process typically involves the use of a feedback mechanism based on benchmark datasets and datasets in processing. The environmental dynamism is handled at a spatial level within datasets, and temporal scales within a sequence of datasets acquired one after another. Thus, this approach tries to create a blueprint of modal or feature level datasets to adjust as per the changed operating conditions.

1.2 Rationale

Currently, a multimodal-imaging is applied in many classes of applications including satellite-imagery, healthcare, gaming, robotics, etc. Developing an adaptive algorithmic solution to work with moving optima type problems would allow improving these applications further to achieve specific goals. An enhanced machine learning model would enable extending its applicability in various application domains since at the core of the computational model, the evolutionary processing will be done on metadata level. Evolutionary Algorithms are especially useful for search and optimisation type problems where there can be multiple solutions, and search space is usually very vast.

1.3 Aims and Objectives

This research study aims to investigate the application of evolutionary algorithms to overcome moving optima and environmental variations in the case of multimodal image detection applications. Since multimodality offers alternate sources of information for a targeted objective, its usage selectively can improve classification and performance of the detection process. In the case of computer vision applications, current methods offer all these advantages in a limited manner. During this research study, an attempt is made to enhance those benefits further by reducing redundant and noisy data-elements.

1.3.1 Targeted Aims

- **Accuracy Improvement:**

  Usage of evolutionary algorithms for optimising multimodal input data at

modal, feature or inference level can assist to improve the accuracy of detection objectives. In the case of a discussed set of applications, running optimisation at metadata level to train an algorithm to select either specific modalities or features of the dataset allows reducing redundancies. Due to this factor, all data supplied to a classification algorithm can have inputs with minimum common cross-section enabling a reduction in confusion and improvement in accuracy.

- **Robustness:**

  Since the proposed algorithms can keep its optimum evolving as per the environmental variations, they can improve the robustness of the detection process. Any sudden or gradual changes in the environment causing its effects on datasets will need to be accommodated using an adaptive algorithm. Especially abrupt variations can be managed by the proposed set of evolutionary algorithms.

- **Reliability:**

  In case of the evolutionary algorithms, one of the primary objectives is to optimize supplied datasets. During this process, the quality of datasets can be enhanced, and the detection process can be made more reliable.

- **Performance Enhancement:**

  Feeding data selectively to fusion, parallel-processing or classification algorithms means a reduction in input scale which may allow improving performance. Reduced size of either data itself or extracted features helps to carry out remaining processing in a more quick and efficient way. Such a reduction can especially help in a real-time application where performance requirements are essential or in specific applications where accuracy requirement is critical.

- **Scalability Expansion:**

  Selective input processing can allow the increasing scale of the input data since it creates an allowance for additional information by cutting on unrequired redundant elements. Thus, a same bandwidth of inputs can allow feeding higher number of information channels of images. Therefore, the proposed method can help to improve the scalability of the detection inputs.

1.3.2 Objectives

The purpose of this investigation is to improve, develop and apply evolutionary computational models to multimodal image processing/computer vision applications to improve accuracy, performance and investigate the effect on the scalability of contemporary detection algorithms. The evolutionary computing methods will be developed either using currently available algorithms like Genetic Algorithms or by developing entirely new concepts inspired from evolutionary biology. Various aspects of imaging applications like spatial and temporal components will be improved by forming adaptive solutions towards a dynamic environmental scenario.

1.4 Hypothesis

"An evolutionary multimodal algorithmic approach can be used to improve accuracy, performance and robustness of the object detection process in dynamic and uncertain environments."

The dynamic optimisation problems are known for their stochastic behaviour. For variable environmental conditions, the pattern verification process needs to ap-

proach as same as moving optimum type of concerns like the dynamic optimisation problems. Tracking such a moving optimum can be beyond the capabilities of traditional numerical methods due to their steady-state or fixed optimum characteristics. The mathematical techniques are relatively computationally expensive and tend to consume enormous computational resources. The computational expense increased is due to the high variability required for large size data handling and processing.

Thus, stochastic techniques can be an alternative for solving such problems. One of the classes of such stochastic techniques is the self-adaptive evolutionary algorithm which provides a state of the art solution for overcoming the moving optima. Furthermore, a meta-modelling approach can improve the robustness of evolutionary algorithms used in dynamic environment detection using multiple sensory resources.

1.5 Innovation and Contribution to Various Bodies of Knowledge

Research work performed so far and to be completed adds value to various engineering application areas. Contributions to specific knowledge bodies are listed below:

- **Pattern Recognition:**

  A new approach specifically for variable environmental conditions is developed where an optimum continuously keeps moving. This method adds value to the current pattern recognition methods by decreasing redundancies in variable inputs.

- **Evolutionary Algorithms:**

  An attempt is given to enhance adaptivity of current multimodal detection and classification processes by developing new variations of evolutionary algorithms. Beyond the widely used concept, 'crossover and mutation', few other

principles like 'phenotypic plasticity', 'bet-hedging' and 'gene-silencing' are implemented to add to this knowledge stream.

- **Data Fusion:**

  A novel heuristic approach is invented for improving data fusion applications which mainly require high adaptivity to the dynamic environments. Addition of the evolutionary mathematical operations allows reducing noisy and redundant data elements. Therefore, the fusion-based methods being critical in multimodal computer vision applications are benefited with the increased accuracy and robustness.

- **Multimodal Computer Vision:**

  A new computationally efficient and intelligent methodology for object detection in various environmental irregularities is discovered. This method not only improves the overall accuracy but also allows to increase the efficiency of input datasets by reducing redundant elements.

Apart from knowledge bodies various application domains as well can be benefited from this research exercise. In the current research topic, the main focus is on developing an intelligent heuristic method for a new generation of low-cost sensory applications. An inexpensive and smart vision system will be the solution to many engineering applications which run in uncertain and dynamic environmental conditions. A number of the applicable areas are listed below:

- **Robotics:**

  A new generation of mobile robots is mostly equipped with such type of vi-

sion system. Use of the evolutionary computing methods can be beneficial to explore objects in constrained areas and variable environmental conditions.

- **Machine Inspection:**

  A new generation of mobile robots is mostly equipped with such type of vision system. Use of the evolutionary computing methods can be beneficial to explore objects in constrained areas and variable environmental conditions. Having a presence of obstacles in variable lighting conditions can be sensed using an adaptive vision system.

- **Surveillance & Security:**

  Many of the surveillance applications are supposed to monitor in a relatively vast verity of working conditions. Any suspicious movement during sudden environmental changes can be challenging to track by using a multimodal sensory approach. Adding an evolutionary computational intelligence to filter situation specific events and anomalies adaptively can make these applications better equipped and accurate.

- **Healthcare Applications:**

  Many conventional multimodal healthcare and medical applications need accurate predictive methods to detect human body parameters and disease symptoms. Using the proposed method can enhance their prediction accuracy. Furthermore, the growth in IoT and cheaper sensing options connected to a cloud server provides newer ways for remote healthcare applications. Right from data optimisation to event level processing, many scenarios can be benefited by using proposed adaptive computational methods.

1.6 Validation of Proposed Methodology

For validating the proposed methodology, a computational architecture is developed which can run on any of the modern computer platforms. This design can be implemented using any procedural or an object-oriented programming language. In the implementation, the object detection process is followed in a serial manner, starting with data-acquisition then data-optimisation and finally completed with the object classification. In this case, few of the most prominently practiced algorithms for every phase of the detection process are pre-programmed in functional blocks and stored. Use of such pre-programmed blocks allows swift and straightforward implementation of experimental scenarios. This application is an essential requirement since the proposed evolutionary algorithms need to work at various levels and phases of the detection process. Having readily available code blocks give an option to use them in a plug and play manner with the proposed evolutionary methods.

Another important aspect of the proposed validation process is to simulate gradual and sudden environmental variations. Here, different noise levels are added from smaller to a more considerable extent in a systematic and stochastic manner to provide the datasets with an additional variability. These added variations allow simulating environmental fluctuations to a certain degree. Added noise creates further variety to the datasets which are used for the experimental validation. Thus, a base dataset is expanded to verify the effects environmental fluctuations on the functioning of the proposed algorithms.

Furthermore, the proposed algorithms are adopted mainly for optimisation purposes. These algorithms are utilized at different levels from choosing a modality to a feature subset selection process. It intends these algorithms need to operate in different scenarios in cooperation with traditional methods used for the multimodal detection processes. In the experimental validation, proposed techniques

are employed to validate them along with the conventional fusion and classification methods. Apart from a combinational technique like data-fusion the validation is also conducted on an entire set of modalities without fusing them. Hence, the experimental validation attempts to verify the proposed hypothesis from contextual, application and computational perspectives.

1.7 Limitations and Constraints

The proposed evolutionary algorithms mainly focus on their applicability and results for multimodal imaging applications. Here, an approach is explicitly studied for computer vision applications using visual, infra-red and ultra-violet modalities. One of the primary focus is to improve object detection process which uses multiple camera modalities. During the experimental validation, there is no attempt made to apply the proposed set of computational models beyond the computer vision applications.

Another constraint is the type of data used and processed by the proposed algorithms. This investigation mostly focuses on a use of spatial domain datasets with their feature level representation. These features are derived further by using sets representing them into real life physical properties like size, shape, colour, etc. of the detected region of interests or ROIs in the multimodal dataspaces. The emphasis is given to improve detection results by reducing redundancy with the application of evolutionary algorithms and then validating the optimized features with classification algorithms.

Finally, the proposed method does not intend to replace the conventional classification algorithms or currently used multimodal fusion algorithms. These algorithms operate as an additional sequence in the detection process to increase accuracy and

reliability by optimising input datasets at different phases.

1.8 Structure of Thesis

This thesis is divided into two main sections with seven chapters:

- This chapter gives an overview of the research topic and the report itself;
- The second chapter explains various applications and theories involved in Multimodal Imaging/Vision area;
- The third chapter provided in-depth knowledge on the proposed computational models for realizing the proposed hypothesis and achieving research goals;
- The forth chapter discusses validation model from software and hardware perspectives with experimental results;
- The fifth chapter presents cases studies associated with the proposed evolutionary computational approach;
- The sixth chapter lists all related research contributions in various international conferences and journals; and
- The seventh chapter finishes the thesis with concluding remarks on the research findings with future directions.

Part I: Theory, Concepts and State of the Art Methods			
Chapter Number	**Introduction**	**Background Context**	**State of the Art Methods**
Chapter 1	*Introduction:* Thesis topic with Aims, Objectives and Rationale		
Chapter 2		*Literature Review:* Multimodal Detection Techniques State of the Art	
Chapter 3			*Methodology*: Evolutionary Multimodal Computational Models

Part II: Research Contributions and Validation				
Chapter Number	**Experimental Famework**	**Validation**	**Action Research**	**Conclusion**
Chapter 4	*Experimental Testbed:* modelling and Design of Framework			
Chapter 5		*Experimental Work:* Validation and Evaluation of Proposed Evolutionary Methods		
Chapter 6			*Action Research:* Environmental, Healthcare and Industrial Applications	
Chapter 7				*Conclusion:* Final remarks with future directions

1.8.1 Structure of Thesis in Detail

The thesis is organised into three parts which are explained in detail below:

- **PART I: *Theory, Concepts and State of the Art Methods***

- **CHAPTER 1: Introduction**

  Pages 23-41

  This chapter provides a high-level perspective of the thesis with the hypothesis, aims, objectives and rationale. It also includes sequence and structure of the remaining dissertation. A sequence and structure of the remaining dissertation are presented in here. A short overview of every thesis chapter is provided. Finally, all published research to date in proceedings and journals is listed.

- **CHAPTER 2: Multimodal Detection Techniques Applications & Methods**

  Pages 41-80

  This chapter provides an overview of recent State of the Art methods applied for the multimodal imaging applications. The main classes were identified which are fusion, parallel processing and classification methods for processing multimodal images or their reduced feature forms. This chapter also provides an overview of various applications associated with Multimodal detection. The multimodality is applicable in many areas of real life which includes healthcare, bio-metrics, robotics, satellite imaging, etc. Very briefly the hardware and software level concepts applied by other researchers are discussed to bring in attention towards applicability of multimodality.

- **Chapter 3: Evolutionary Multimodal Algorithms**

  Pages 80-113

  This chapter discusses methodology in detail with all provided solutions to the problem to be addressed. Here, various multimodal evolutionary methods are presented with their computational architectures.

- **Part II: *Research Contributions and Validation***

- **Chapter 4: Experimental Design**

  Pages 113-135

  Chapter four outlines experimental framework and considerations. In this chapter, system specifications and architecture required for performing proposed experimental work is explained.

- **Chapter 5: Experimental Results**

  Pages 135-154

  This chapter provides experimental results of all the validation exercises for the proposed methodology. Qualitative and quantitative aspects of the proposed method is assessed through these results.

- **Chapter 6: Action Research**

  Pages: 154-166

  In this chapter, case studies providing the proposed research used in real-life scenarios are discussed. In here, application for environmental monitoring and health domain are presented with their architectures and prototypes.

- **Chapter 7: Conclusion**

  Pages: 166-182

  The seventh chapter discusses conclusion and future work to summarize the completed work and possible extension of the current work in future.

- **Bibliography: Appendices and References**

  Page: 182

  This part provides all references and appendices to this thesis.

1.9 Research Contributions

The following list details the most relevant research publications pertaining investigation of this research work:

1. Kale, A. and Chaczko, Z., "Deep Learning and Learning Classifiers for Multimodal Image Processing," In Proceedings of the 5th Asia Pacific Conference on Computer Aided System Engineering(APCASE'17), Guillin, China, November, 2017.

2. Kale, A. and Chaczko, Z., "IMMuNE: An IoT Methodology for Multimodal Networks using Evolutionary Algorithms in Healthcare Context," In Proceedings of the 5th Asia Pacific Conference on Computer Aided System Engineering(APCASE'17) , Guillin, China, November, 2017.

3. Kale, A. and Chaczko, Z., "iMuDS: An Internet of Multimodal Data Acquisition and Analysis Systems for Monitoring Urban Waterways," In Proceedings of the 25th International Conference on Systems Engineering (ICSEng) (pp. 431-437). August, 2017.

4. Kale, A., Chaczko, Z. and Slehat, S., 2015, July. HyMuDS: A Hybrid Multimodal Data Acquisition System. In Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on (pp. 107-112).

5. Kale, A. and Chaczko, Z., 2015. Evolutionary Feature optimisation and Classification for Monitoring Floating Objects. In Proceedings of Computational Intelligence and Efficiency in Engineering Systems (pp. 3-16). Springer International Publishing

6. Kale, A., Chaczko, Z. and Rudas, I., 2013, February. Managing Dynamism of Multimodal Detection in Machine Vision Using Selection of Phenotypes. In International Conference on Computer Aided Systems Theory (pp. 483-490). Springer, Berlin, Heidelberg

7. Kale, A., Chaczko, Z., Wazirali, R., 2012, November, Evolutionary Switching for Modality Selection in Water Contamination Monitoring, Proceedings of International Conference on Broadband and Biomedical Communications, IB2COM 2012

8. A Kale, C Chaczko, "Genetically Inspired modelling of the Multimodal Feature Space," 1st Australian Conference on the Applications of Systems Engineering ACASE'12, pp.26- 27, Feb 2012

9. A Kale, C Chaczko and I Rudas, "Parallel robot vision using genetic algorithm and object centroid," Proceedings of EUROCAST'11 Proceedings of the 13th international conference on Computer Aided Systems Theory - Volume Part II, Springer-Verlag Berlin, Heidelberg, pp. 170-178, 2012

10. Chaczko, Z. and Kale, A., 2011. Automated Tablet Quality Assurance and Identification for Hospital Pharmacies. International Journal of Electronics

and Telecommunications, 57(2), pp.153-158

11. Chaczko, Z., Kale, A. and Chiu, C., 2010, November, December. Intelligent healthcare—a motion analysis system for health practitioners. In Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010 Sixth International Conference on (pp. 303-308). IEEE

12. Kale, A., Chaczko, Z., Chiu, C., 2010, October, Software Framework for Intelligent Micro-Factory, Proceedings of 2nd MICEEI International Conference, Makassar, Indonesia

13. Shedge, S., Kale A., 2010, October, Remote Patient Monitoring Network Cluster, 2010 Proceedings of 2nd MICEEI International Conference, Makassar, Indonesia

2 State of the Art Techniques for Multimodal Detection

Multimodal object detection can be performed using various computational techniques. Based on the rigorous literature review these methods can be divided into three major classes: data fusion, parallel processing and heuristic or machine learning algorithms. The fusion techniques use either spatial or spectral form of registered images and combine them. This aggregation can allow complementing of modalities to each other to create an enhanced result. The success of these methods massively relies on registration of the input datasets. The fusion process can be implemented at data, feature and inference level. The second significant multimodal processing category is parallel processing methods. Implementation of parallel algorithms is mostly driven by computer hardware architectures. The heuristic or machine learning algorithms are often applied to multimodal detection for the data in feature form. Multimodal feature optimisation and classification are the most common applications of the machine learning algorithms.

This chapter has the first subsection to discuss these multimodal methods. The second subsection has case studies and examples of the applications which use multimodal detection algorithms. The final subsection summarizes the chapter with concluding remarks.

2.1 Computational Apparatus for Multimodal Detection Methods

2.1.1 Image Fusion Methods

As discussed before, many of image fusion processes are available to integrate data at various abstraction levels.Two of the most commonly applied fusion techniques are as below:

- **Spatial Fusion:**

  Mamdani, Wavelets and Principal Component Analysis

- **Inference Fusion:**

  Bayesian and Dempster Shafer methods

Spatial Fusion

The Mamdani technique[96] replaces one to one pixels at the same location with a maximum, minimum or average value of a pixel to create an output image. These methods tend to achieve moderate success with the low-quality input images. Any further analysis will highly rely on the accuracy of the image registration as well. Misaligned photos will add to errors of the fusion process and this will not allow correct detection of objects present in the acquired imagery. Thus, these algorithms offer a straightforward solution to many problems and can be used in cases where registration accuracy is guaranteed.

Wang et al. propose a General Image Fusion (GIF) method [124] which is a comprehensive framework for evaluating existing image fusion methods. In this research study, authors review many contemporary fusion methods which include HIS transform, Brovay transform, High-pass filtering, Wavelet transforms, High-pass Modulation, Principal Component Analysis, and Multi-resolution Analysis-Based Intensity Modulation.

Another review by Bo et al. [136, 127] in various image fusion techniques discusses pixel level fusion methods which present a generalization of computational architecture. The overall sequence of this architecture includes multimodal image registration process followed by pre-processing, fusion, post-processing and finally visualization. Authors consider most commonly used imaging modalities including

visual, infrared, Synthetic Aperture Radar (SAR), Panchromatic (Pan), etc. for the fusion operations.

Weighted Averaging is one of the simplest pixel level fusion methods which uses the average value of every pixel from all source images.

$$C(x, y) = \alpha AA(x, y) + \alpha BB(x, y) \tag{1}$$

Where A and B are source images whereas αA and αB are scalar weights. The outcome of the fusion process is an image C.

The Principal Component Analysis or simply PCA [87] is a method which extracts major components of the data to decrease it to lower dimensions. It converts set of correlated variables into a smaller set of uncorrelated variables called principal components. In a simpler terminology PCA can create multiple component images at lower resolutions which are further used to fuse with transformed images from another modality. The biggest drawback for these methods is suppression of salient features since averaging operation treats all input modalities with equal importance and blindly uses every pixel value.

Multiscale-based fusion schemes use three essential steps to create a fused output. These three operations include:

- Decomposition of input images A and B into sub-images using either wavelet or pyramidal transform methods,
- The second step is to fuse all decomposed images employing a fusion rule $\emptyset$, and
- The final step is to reconstruct fused image C by applying an inverse transform.

Pyramidal decomposition allows simulation of multiscale information into multiple scales of human visual perception. A primary example of this method includes a Gaussian pyramid based on Gaussian kernel ω and a 2D downsampler. For an ith level of the Gaussian pyramid for image g(0) is:

$$g(i)(x,y) = [\omega(x,y) * g(i-1)(x,y)] \downarrow 2 \tag{2}$$

Where * is a convolution operator and $[\cdot]\downarrow 2$ is a down-sampling operator.

In case of Laplacian pyramid similar decomposition is achieved using following formula:

$$l(i-1)(x,y) = g(i-1)(x,y) - 4\omega(x,y) * [g(i)(x,y)] \uparrow 2 \tag{3}$$

The advantage of the wavelets is that it can provide representation in both time and frequency domains. In the case of wavelet decomposition, an input image is decomposed into low and high frequency components. In the case of fusion applications, multiple inputs are split into low and high-frequency components and then those are fused together. A typical one stage two-dimensional wavelet decomposition is explained here.

Fusion rules in are used to fuse decomposed images into a singular output. Most commonly seen three types of rules are as below:

- **Point-based Rules:**

  In this rule every pixel is considered as an independent entity and then fused together.

$$SC(x,y) = SA(x,y), |SA(x,y)| > |SB(x,y)| SB(x,y) \tag{4}$$

- **Area-based Rules:**

  In this type of rules, a correlation between the local area of the main image with other sub-images is used as a fusion rule.

- **Region-based Rules:**

  Use of edges or segmented regions is one of the conventional methods under this type of rule. Evaluation of fusion techniques can be done using four common techniques:

  - **Root Mean Square Error (RMSE):**

    Error computed between an ideal and fused image using following formula:

  - **Mutual Information (MI):**

    This method is more useful in case of heterogeneous fusion applications and this parameter is computed using following formula:

$$MI = H(A, B) + H(C) - H(A, B, C) \tag{5}$$

- Where:

  $H(C)$ is defines the entropy of the post-processed image C;

  $H(A, B)$ represents the collective entropy of the inputs A, B;

  $H(A, B, C)$ is the joint entropy of input and post-processed images.

  - **Edge Formation (EF):**

    A metric used in the research work [134] by Xydeas et al. In here, QAC

is edge preservation amount within inputs A and C; QBC is the value in B and C; wA and wB are weight functions described based on the edge intensity of inputs A and B, respectively.

– **Spatial Frequency (SF):**

This evaluation metric uses rows and columns as a basis for calculation: Here, f_{row} and f_{col} are spatial frequencies.

colour Space Fusion

colour Space Fusion is a linear combination of input image colour spaces into a fused image to create an imaginary with a new colour spectrum. The best example is combining RGB with IR spectrum creating another colour spectrum. Here, Markov Random Fields and Simulated Annealing methods use optimisation approach to fusion. An appropriate cost function is used to define a fusion goal and then an optimisation algorithm is used to find an optimum solution by treating the data as specified.

Kumar and Dass [75] suggest an algorithmic approach based on total variation for pixel-level image fusion. The authors propose the fusion process as an inverse problem and locally affine conjunction with PCA to estimate the image fusion output. The proposed algorithm estimates the sensor gain by splitting the input image into smaller blocks. The fusion process is implemented using a total variation approach over every pixel in a set of images.

Hamza et al. propose a pixel level fusion method [46] by treating it as an optimisation problem where a biorthogonal wavelet transform of every input image is calculated and then divergence based fusion algorithm constructs composite wavelet coefficients. Purpose of this research study is to integrate complementary informa-

tion from multiple images and enhance a base image to improve visual perception for human eyes. Authors claim superiority over other wavelet-based contemporary methods.

Yang and Li propose an image fusion method [135] using the signal sparse representation approach. In this work, the authors create a set of sparse representation on overlapping patches instead of an entire image. They also use simultaneous orthogonal matching pursuit to temporally align sparsely decomposed images into the same subset of a repository.

Petrović and Xydeas propose a technique[92] which uses multiscale image fusion to provide an efficient fusion process and to minimise aliasing effects in conventional multi-resolution fusion methods. Here, the fusion process involves combining of features at two different scales by using adaptive size averaging templates. Simple arithmetic fuses the larger sized images and then an efficient feature selection is applied to combine the more delicate details.

Thus, the spatial fusion techniques are very commonly applied with several variations as per the requirements of a targetted application.

Inference Methods

Inference methods [129] use a rule-based approach from the theory of probability. Bayes' formula is a prime for all inference methods due to its ability to define the relation between priories and hypothesis. This subsection discusses inference methods including the Bayesian model, the Dempster-Schaffer approach, and hybrid probabilistic models.

The Bayesian Inference model is a probabilistic model commonly employed for sensor fusion applications [30]. It provides a characterisation of probabilities to the truth and falsehood of events that are not random. It involves collecting an evidence that is either consistent or inconsistent with the desired or targeted hypothesis.

With the sufficiency of the evidence, an assumption of confidence in the hypothesis is obtained.

The simplest form of Bayesian Inference model has following representation:

$$P(H/E) = (P(E/H) * P(H))/P(E) \tag{6}$$

Where:

- H = A specific hypothesis which may or may not be a null hypothesis
- E = An observed evidence
- P(H) = A prior probability of H
- P(E/H) = A conditional likelihood of E to observe if H appears valid
- P(E) = A marginal likelihood of E is a priori likelihood of witnessing the new observation E under every potential hypothesis
- P(H/E) = A posterior probability of H provided E and a new measure of likelihood is an estimate of the likelihood that hypothesis H is valid, taking evidence E into a record

Thus, the Bayesian approach can be applied to applications involving mixed or uncertain information. This reflection is derived because it supports propagation of a belief and it can be implemented for modelling affiliations between diverse ranges of tentative information sets.

An investigation by Pavlin et al. [90] provides a case study on a Bayesian fusion of multisensor fusion. The authors assume scenarios where states can be represented in collections of arbitrary variables and then distributed into hidden

and observed events. Finally, the hidden events in this representation are evaluated for a classification process by the cooperation with Bayesian Networks. Here, the mapping between hidden states and hypothesis is defined by the Bayesian Network. Therefore, this entire process transitions with a state estimation and probabilistic inferring tasks in a sequential manner.

Singhal et al. propose [112] a multilevel Bayesian modelling approach for automatic subject detection in an image fusion application. The goal of this method is to overcome shortcomings in the previous techniques while addressing uncertainties and dependencies. In this process, statistical evidence is generated by applying region segmentation and feature extraction processes to the image. This evidence acts as an a priori information for the Bayesian network used for detecting the main subject. As per the authors, a tree-like structure of this system allows to analyse the statistical cues in greater detail and provides superior results as compared with the contemporary methods. Multilevel Bayesian modelling to construct a belief map by identifying the main subjects located in the image significantly improves classification accuracy. Thus, this study provides a detailed view of the applicability of a multilevel Bayesian inference theory towards a real-life problem.

Wei et al. propose a method [128] for fusion of hyperspectral and multispectral images using a hierarchical Bayesian model. In his work, a set of multiband images for the same scene is acquired from multiple sensors and applied with an observation model as below:

$$Zp = Fp(X) + Ep \tag{7}$$

- Where:

- Zp is an observed data,
- Fp(X) is a linear transformation model over a scene, and
- X and Ep are noise and modelling coefficient.

In this approach, prior information on a spectral distribution of images is used as a probability distribution for the Bayesian inference model. The Bayesian estimators are computed from a scene of interest by using an appropriate sampling strategy towards the images captured by sensors. Thus, the Bayesian modelling allows an efficient decision level fusion between the pictures with different spectral dimensions.

Guerriero et al. propose a strategy for detecting a target using a distributed method [42] utilising a network of sensors. In this approach, multiple sensors are scattered in different geographical locations to track the possible locations of a target under consideration. Here, every sensor is responsible for sending its observation efficiently by using a quantised form of data. A view collection is done separately at a system called a fusion centre. A Bayesian model is embedded in the fusion centre for calculating a possible location of the targeted subject. Thus, the Bayesian model along with a decentralised architecture helps to improve the system efficiency and reduces the data transmission bandwidth.

Wu & Wong use a signal strength of wireless sensor network to improve the target tracking scheme [118, 131, 136]. A collaborative effort by the sensors along with cross-modality and cross-sensor fusion allows accurate tracking of objects. The Bayesian network here formulates the sensory signal strengths to calculate the probable location of the object to be detected. Thus, the Bayesian inference modelling effectively establishes cross-modal and cross-sensor decision level fusion within a group of sensors.

Prodanov and Drygajlo propose a speech error handling method [97] for an interaction between a robot and human in a high acoustic noise. Here, the acoustic modality alone cannot be used to the high level of interference in the environment, and it is supported by a laser scanning modality. A Bayesian modelling of acoustic and scanner data reduces the speech error and improves the accuracy of the system. Again, a Bayesian approach allows a decision level fusion using a heterogeneous combination of the modalities.

Thus, the Bayesian methods for sensor fusion tasks are essential techniques which allow a decision level fusion by isolating the actual data level constraints. Other advantages include improved energy efficiency, reduced bandwidth and increased geographical scope of the application. The most notable limitation of this approach is its heavy reliance on the evidence collected from different sources. In a probabilistic approach, the evidence can only serve to improve the probability of the hypothesis, or in case of inadequate evidence, the likelihood can drastically decrease.

Franko and Boyer propose an image fusion method [36]for silhouettes using a Bayesian inference approach. In this work, they recommend the use of the inference technique to a forward sensor model with cause and relationship model. A joint probability calculation using Bayes rule used to calculate the outcome of the inference level fusion process.

Gros et al. propose a data fusion model [31, 32, 41, 43] for Non-Destructive Testing (NDT) at the pixel level using a Bayesian inference approach. The process here includes the input of eddy current and infrared thermography, pre-processing and finally fusion using the Bayesian inference technique. Authors use a conditional probability of occurrence for both parameters (Eddy current and Infrared thermographic images) to calculate the combined inference for NDT results. Thus, two distinct modalities are fused at inference level by the authors using a Bayesian

model. Here, authors do consider the possibility of either additive or subtractive effect of the fusion process.

Murtagh et al. propose a Bayesian inference method for multiband image segmentation [86] with model-based image segmentation. Here, authors use a Bayesian selection model for choosing multiple clusters. Authors consider a number of candidates $G = Gmin, \ldots..Gmax$. Every cluster G in this set represents a separate statistical model for the dataset MG. Here, the Bayesian formula is used to combine inferences received from all clusters for a multiband fusion process. Authors claim superior results due to the Bayesian method they applied.

Another commonly applied method classified under a Bayesian model is a Hidden Markov model. The Markov method models the hidden observation states for further processing. A case study [110] by Shivappa et al. performs a multimodal tracking of objects using hidden states or filtered observations. The hidden states are decoded from observations coming through a principal modality. A posteriori likelihood is determined and applied as extrinsic knowledge to the next sensor to decode hidden states of subsequent modality. The method proceeds furthermore until the final modality is assessed using external knowledge acquired from its predecessor. Following step of an algorithm is to threshold the aggregate probability of the state sequences.

The Dempster-Schafer approach is an evidence-based mathematical theory which it is a generalisation of Bayesian inference theory [48]. Evidence collected from different sources with a degree of beliefs of all available evidence is applied for an effective sensor fusion process. Due to the limited success of Bayesian implementations to handle the problem of 'uncertainty', a separate attempt is needed to overcome this issue. The Dempster-Shafer approach, though, a belief function can help with an additional attribute for measurement of states. A belief in a hypothesis is described

using beliefs and plausibility. A belief is measured by summing quantities of each subset of the likelihood. Plausibility is a difference between the maximum probability and all quantities not intercrossing with the hypothesis. Thus, the belief represents the support to hypothesis and plausibility.

Wu et al. discuss an approach applied to a sensor fusion application using Dempster-Shafer theory [130]. The authors assume that the context knowledge can be expressed in metaphorical and numerical form. The mapping of data flow is performed from sensory observations to represent context level description. Based on these premises, an intermediate level or higher-level modalities are evaluated for application-specific inferences. An observed sensor contribution is applied to establish a belief in the context of realities or events. Based on the set of belief masses, the plausibility of every hypothesis is determined. A further value is added by specifying weight to Dempster-Shafer evidence combination rules. The weighted Dempster-Shafer combination of evidence is produced by estimating historical data from sensor performance. Finally, the historical data of a sensor for a particular situation is used for calculating the level of trust in the current case.

In another study, Pong and Challa [95]explore various uncertainty measures under an evidential reasoning framework and Dempster-Shafer Fusion Frameworks. Here, the sensor selection process is based on the measure of uncertainty level. Total Uncertainty measurement with a proposed search needs further work for making sensor fusion adaptive. Fuzzy Logic combined with probabilistic approaches.

A method by Sun et al. uses a combined approach with a K means clustering algorithm and Dempster-Shafer evidential rules [117]. A transformer error detection application was developed by applying a multisensory fusion method. The sensory data is fuzzified for detecting fault patterns and fed to Dempster-Shafer inference engine. Thus, fuzzy logic is used to a fuzzification of the sensory states helps to

handle the dynamic behaviour of events. A mixed approach with a fuzzy and probabilistic method is applied for estimating the authenticity of the fusion system. A reliability evaluation algorithm uses a failure frequency as an ambiguous parameter.

An extensive review and trials on multimodal detection[146] by Yi Zheng and Ping Zheng shows significant results. The authors use a fuzzy inference method to combine visual and infrared spectrums. The fuzzy inference is particularly used to improve the image fusion quality. The case study suggested by the authors required detecting objects in variable environmental conditions. The proposed method shows enhanced results.

An investigation by Kumar et al. explicitly concentrates on addressing target discovery in changing environmental conditions by applying multimodal sensory approach[75]. Multimodal sensory data are utilised to create beliefs about a target in the atmosphere. Authors use an intermediate or a high-level belief fusion approach. The process begins by evaluating of Region of Interest (ROI) or blob using each modality. Every modality detects the right target candidates with specific dimensions and locations. During the next step of the process, a belief mass is calculated for every modality employing appearance proportion and similarity calculation parameters. The further algorithm follows a context-based belief fusion and Kalman filter based on target position integration for additional precision. The proposed algorithm is particularly useful for the system operating in environmental variations because of lighting, perceptibility and climatic conditions.

A robotic application[6] proposed by Amantiadis et al. uses fuzzy fusion process for navigation purposes. The intended application uses a hybrid combination of modalities including vision and RFID. Operational conditions are examined and classified into measurable characteristics by a multimodal sensor fusion process. A relationship between knowledge attributes in surroundings in a feature form is as-

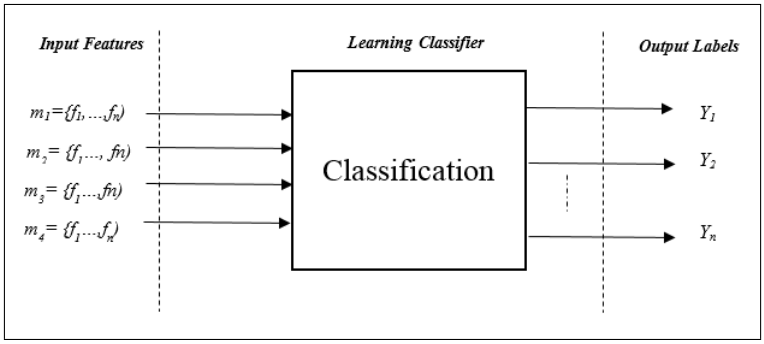
sessed. Conclusively, a qualitative property is needed to estimate the RFID efficiency as a degree of probability for each location estimate.

2.1.2 Multimodal Detection using Machine Learning Methods

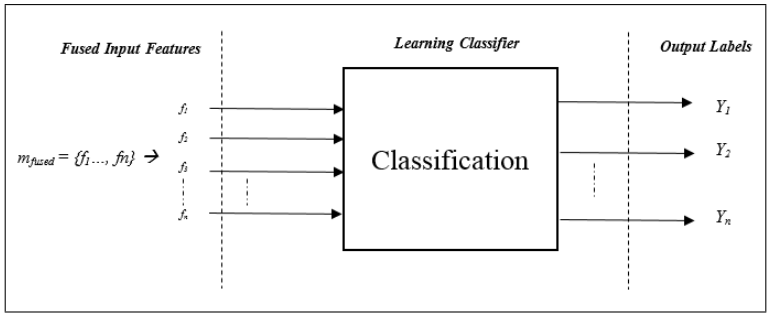
Machine Learning methods are very commonly applied for multimodal detection purposes [43]. In most cases, the unprocessed data is learned and classified at a feature level to identify objects and labels within. The process has two phases, learning and validation. The algorithms are supplied with a portion of datasets for learning, and the remaining portion is validated. A common observation in most of the application is the learning requires two-thirds of the dataset. As understood from the literature review, these algorithms require a massive effort in the learning process.

Following are the most commonly used multimodal learning algorithms for multimodal detection applications [71, 143, 144]:

- Decision Tress (DT),
- Artificial Neural Networks (ANN),
- Support Vector Machines (SVM),
- Random Forests (RF),
- K Nearest Neighbors (KNN),
- Naive Bayes (NB), and
- Discriminant Classifiers (DC)



(a) Classification without fusion



(b) Classification with fusion

Figure 1: Feature Classification Process in Multimodal Detection Context

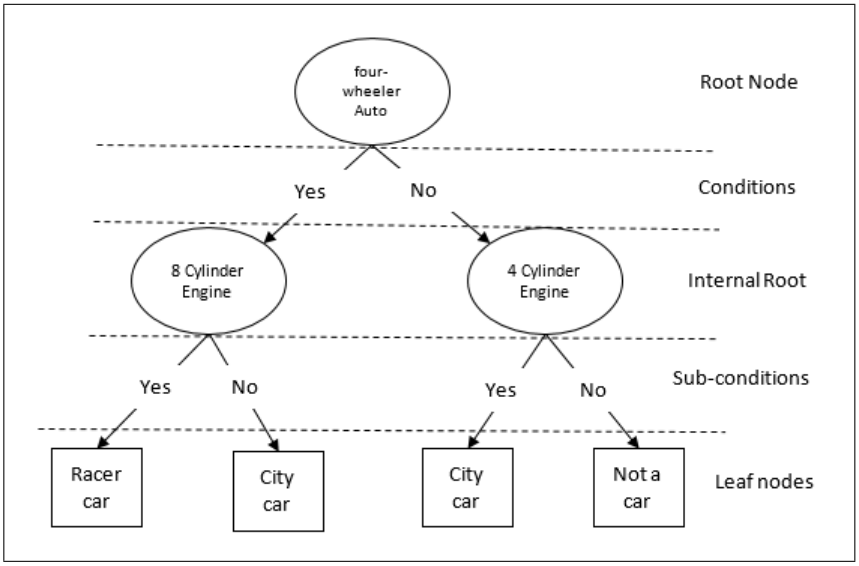


Figure 2: Decision Tree Classifier Example

This subsection discusses the theory and applications of the multimodal classification algorithms.

Decision Trees and Random Forests

Decision Trees [104] have a tree-like structure as depicted diagrammatically in Figure 2 to model decisions and their possible consequences. The decision tree is one of the simplest models which has been very commonly implemented by the research community. Random Forest algorithms [80] use the Decision Trees in a large volume to stochastically select branches within to form a classification model.

Su et al. use AdaBoost along with Decision Stumps for classifying music emotions in multimodal feature space [115]. The primary task here is to categorise the set of songs into emotion categories. Authors use a decision stump method where single-level decision trees are used to express confidence in decisions. The Multimodal feature set includes audio and lyrics features which are expected to correlate with emotions. Input features are extracted from audio files and lyrics text. The

AdaBoost algorithm is used to maintain weight distribution over all input features. Multiple training cycles are performed over Decision Stumps or so-called 'Weak Classifiers'. A linear combination of all iterations is used for the final classification. Each piece of music with lyrics has an emotion label attached to it. Here, classification success performance depends on a massive training process conducted by AdaBoost. Another observation seen here is that it does not adequately utilise Decision Trees and uses the only single layer as a weak classifier.

Kerr et al. use decision-trees for classification of multimodal data to diagnose epilepsy [66]. Combination of data belonging to different Epilepsy patients was used to create a multimodal feature pool. Main three modalities used here include CI, MRI, and PET from the real-world data of patients. Since this data belongs to known clinical cases, all datasets are prelabeled. Decision tree classification of this multimodal feature-space provides better accuracy as compared to single modality feature classification. Combined modalities can stretch the accuracy more than 80% as compared to 58.7% maximum for an individual modality classification. Still being a clinical dataset, the proposed application demands more correct results and relies heavily on human intervention. Additionally, highly conditional and decision cues are required to improve this model.

Chen et al. propose a Multimodal Data Mining framework [21] for soccer goal detection using a multimedia data. This study combines a multimodal analysis and the decision tree algorithm to analyse multimedia datasets. The multimodal feature set is formed out of audio-visual features of soccer game videos. The feature set creation process requires audio-visual data extracted for each shot at different granularities. The feature-set is filtered to omit noise and then used for the classification process. The iterative partitioning approach is applied to training datasets to divide the data into smaller branches using context-specific attributes. The decision tree

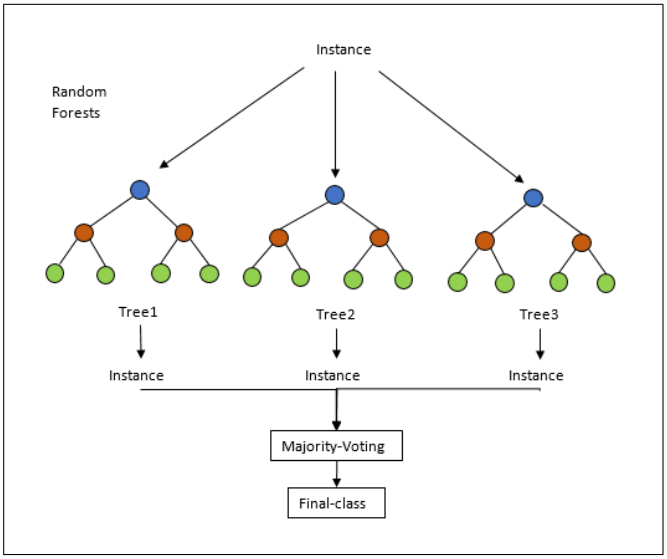


Figure 3: Random Forests Classifier Architecture

creation process stops at leaf or class label creation. Thus, the decision tree for classifying a feature set into a goal is developed. Author claim very high accuracy in their experimental results. It may be possible due to binary classification with only one class label is required to find - that is, a goal scored from an audio-video sequence.

Artificial Neural Network

The 'Artificial Neural Networks' or ANN is a technique [29] inspired by biological neurons in a human brain where every single processing unit comprises a processing capability. The ANN has multiple layers with every layer having various nodes to receive and process the data supplied. A typical implementation of an individual neuron and a multi-layered ANN is described in Figure 4 and 5 respectively.

Hosseini and Moghadam propose an approach [52] based on Artificial Neural

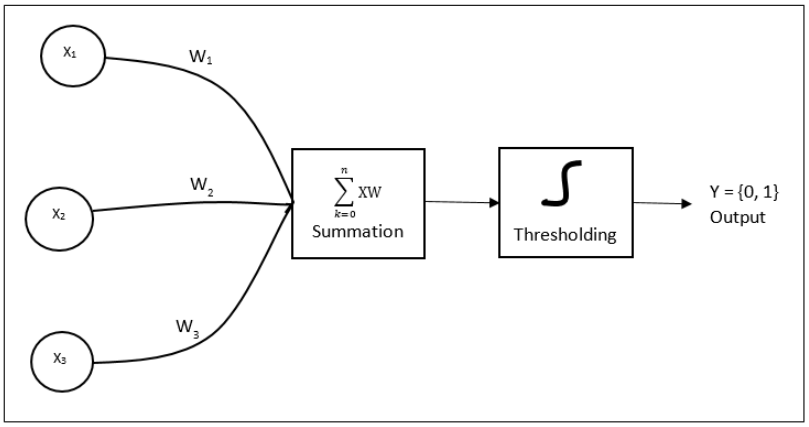


Figure 4: Single Neuron in ANN

Network to classify a multimodal featureset of a video. The authors propose a Neuro-Fuzzy approach to conduct an analysis of broadcast soccer videos. Set of audio-visual features is used as an input to the proposed classification algorithm and expected classification results as events occurred. Fuzzy logic is used to overcome unpredictability in video data. Authors use the set of fuzzy rules for various semantic events in the audio-video dataset. The proposed computational model allows authors to analyze various soccer matches to find events like a goal, foul, corner, offside, etc. The proposed method is learning intensive and needs plenty of data with prelabeled classes.

Liu et al. propose a method based on stacked contractive encoders [81, 82] or deep-networks for multimodal feature learning. In this approach, authors try to accommodate three modalities in a video signal which include an image, audio and text. The process starts with building a single network to address a feature classification of an individual modality and then scale up to classify remaining feature subsets. For images, feature-set includes visual aspects of shape, geometry, texture, colour, etc. Whereas, audio and texture include their modality-specific features. The entire classification model contains two main layers named as Autoencoder and

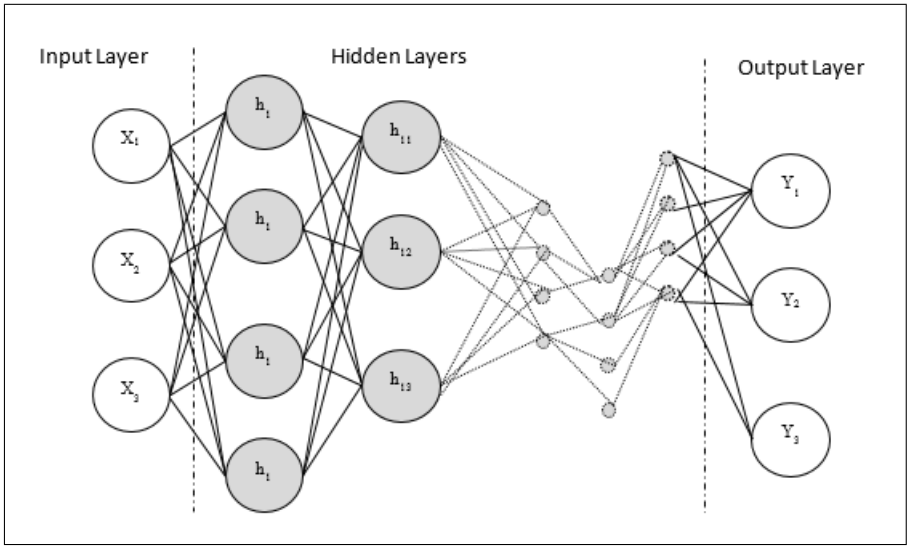


Figure 5: Multilayer ANN Architecture

Decoder. An autoencoder is a set of neurons x, which include input, hidden and reconstruction layers. Whereas, the decoder is a representation of outputs y or a processed form of inputs x in several layers of neurons. The authors use the sigmoid as a base function for every neuron. As per the authors, results of their work show better classification accuracy as compared with contemporary methods.

Gustafson and Paplinski propose a bimodal integration of phonemes and letters using Self Organised Maps [45] which is a variant of the ANN. In this case, a two-dimensional organised plane of neurons is used to receive inputs. The model mainly works on three principles of competition, co-operation and synaptic adaptation within neurons. Competitiveness is a method to find the best possible weighing vector which will give the Autoencoder least error level. The process starts with random weight allocation to neurons, then selecting a vector randomly from the weighted neuron pool and finally by assessing the Euclidean distance between inputs and weights to get a set with minimum distance. This method is beneficial when output data is not available or is limited in quantity.

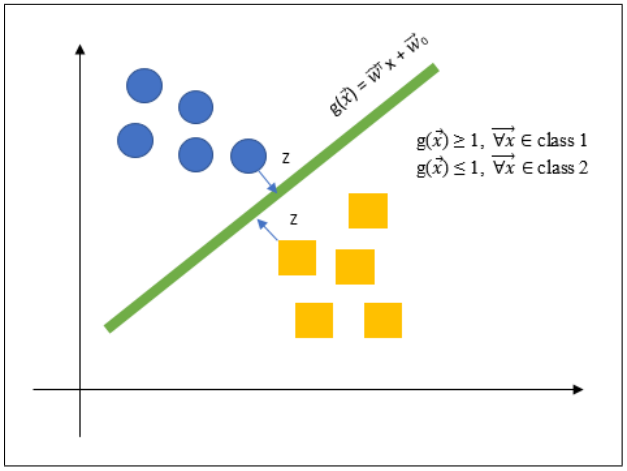


Figure 6: Binary SVM Classifier

Support Vector Machine

Support Vector Machine [109] or SVM is a learning classifier which can have one to an infinite number of planes to separate feature space into separate classes. Figure 6 shows a typical example of a two-class SVM. SVM is a widely used classification algorithm in many multimodal imaging applications.

Fleury et al. propose an SVM based multimodal classification [34] of activities in smart homes from a health perspective. This method combines sensory data coming from occupancy, sound, temperature and camera sensing modalities. This approach allows combining location, activities of a resident with environmental parameters like temperature to analyse events in their daily lives. The authors use SVM for classification purposes since it requires relatively smaller training data.

Chen et al. propose a multimodal classification approach [18] for documents using feature enhanced rank class (F-Rank Class). The authors combine textual with image data to develop a hybrid or a heterogeneous featureset. This method allows them to combine interesting features from both modalities to complement each other and classify documents. The method forms a heterogeneous information

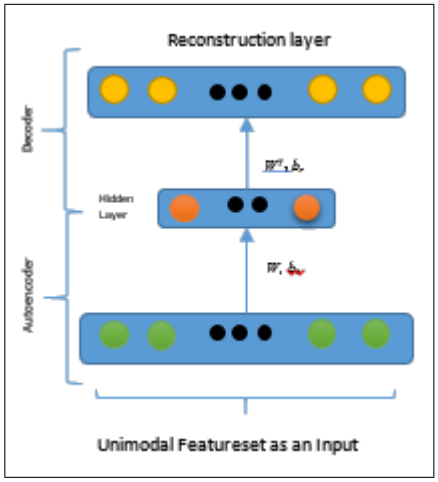


Figure 7: Autoencoder for Documents

network, then developing a ranking distribution and finally using the F-RankClass framework to classify the multimodal featureset. The comparative study with SVM was conducted to show the merits in the proposed method.

Xue et al. proposed a method [133] to classify image-types present in artefacts from biomedical journals. The objective of the authors is to sort images into two major classes: regular images which can be X-ray, photographs, MRI, etc. and illustrative images which may include mimics, charts, graphs, etc. Challenges in this application involve images surrounded by text, images compounded or combined with other images. Authors propose two algorithms based on SVM and Deep Learning approaches which try to classify a binary set of features. In the proposed method, features are used by authors called as 'semantic concept features' or 'Bag of concepts'. Authors claim better accuracy with the contemporary methods.

Verma and Tiwari propose a multimodal fusion framework [122] for emotion classification using physiological signals. Their work involves modelling signals using wavelet transform and classification using SVM and other contemporary classifiers. Binary classification capability of SVM is used with a Kernel called Poly-Kernal at

a certain value. Authors attempt to classify signal data either in separate channels or in a fused manner. Thus, physiological signal vs emotion relationship is achieved by using classification algorithms and sensor fusion.

Naive Bayes and Discriminant Classifiers

Naive Bayes classifier [85] is one of the probabilistic classifiers which is based on the Bayes' theorem by applying strong or naive confidence within the features. The usual Bayes' rule where the posterior probability is proportional to prior, likelihood and evidence is enhanced further to calculate conditional distribution over the class variable. Various examples of the use of Naive Bayes' classifier in multimodality include biomedical, biometric, machine vision and multimedia applications.

Chang et al. apply the Bayes approach [67] in annotating multimodal images for retrieving purposes. The authors employ a process which starts with the annotation of training images by labelling them with a single semantic label. Then the binary classifiers are trained to predict label memberships. Authors use SVM and Bayes methods for classification accuracy comparison. Finally, authors find Bayes method with a better annotation quality than the SVM-based ensemble used in multimodal image retrieval.

Gupta et al. provide a multimodal sentiment analysis [44] and context determination by using perplexed Bayesian classification. In this work, the authors focus mainly on internet multimedia content to extract emotions out of them and classify into sentiments or contexts. Application of perplexed version of Bayesian classifier over the traditional Bayesian classifier provides relatively better results. Linear Discriminant Analysis (Discriminant Classifier) or LDA or DC is another statistical classification algorithm [8] categorised under supervised machine learning. The LDA was invented by Ronald Fisher and mostly used as a dimensionality reduction technique. But in cases, it is also practised as a classifier. The LDA in most of

the cases requires steps to compute mean vectors of the classes involved and then computing scatter-matrices and eigenvectors. Later these eigenvectors are sorted, and finally, only high value-vectors are selected.

Kim and Kitler propose a multimodal face detection algorithm [67] using an LDA with locally linear transformations. Authors focus on nonlinear variations of the facial features due to changes in its pose. Due to the local linear transformations, the facial classification in such a situation is seen improved. Li et al. apply Fisher's discriminant analysis for dimensionality reduction [79] and class assignment of hyperspectral imagery. The primary objective of this use of Fisher's method is to preserve multimodality with a reduced dimensionality of the subject datasets. Employing local Fisher's LDA model shows a significant success in achieving goals set by the authors. Torre and Kanade propose an LDA implementation [77] to overcome the general limitations with high-dimensionality and multimodality. In this work, a Multimodal Discriminant Analysis is applied to overcome these constraints in a conventional LDA. A comparative analysis shows superiority over contemporary methods. Thus, different variants of Bayesian and LDA classifiers can be applied for multimodal imaging applications in various scenarios.

Adaptivity with Evolutionary Algorithms

An evolutionary algorithm [72] is a class of algorithms based on a biological emulation of a theory of evolution. In this type of algorithm, the Darwinian principle of "survival of the fittest" is applied to search the optimisation. The algorithmic process involves initialisation of population, fitness or an objective checking and reproduction of the next generation. It commences with a random generation of an encoded population of solution approximations. In the following event, every population member is evaluated for its fitness index against an objective function of the solution. If desired fitness level is found, then the process is terminated else contin-

ued with a reproduction of next-generation chromosomes. Reproduction operation consists of crossover and mutation of population members. During the crossover operation, a chromosome bit sequence varies by swapping one or more bits of its binary code with another population member. Mutation creates random bit swapping in the population to maintain a diversified population in the population pool. Thus, this cycle of a fitness check and reproduction continues until the termination criterion is matched. The termination criterion can be either a number of iterations performed or reaching a desired fitness level of the solution.

Maslov et al. propose an extensive investigation [83] on an application of evolutionary algorithms (EAs) for a multi-sensor fusion process. Principle sub-domains of evolutionary algorithms including Genetic Algorithms and Genetic Programming are discussed from sensor fusion process perspectives. It indicates that evolutionary techniques are gaining popularity due to their superiority and simplicity over traditional optimisation techniques. Conventional optimisation techniques are based on fixed states and are made for deterministic problem spaces. Evolutionary algorithmic methods provide an automated approach and handle stochastic situations very well.

A few of the significant applications of EA in sensor fusion applications are discussed below. An image registration process is a critical process in remote sensing, industrial and medical applications and EA is successfully applied in complex image registration processes. In the case of a complex form, which involves multiple computers, receiving imagery information from numerous sources, the registration process becomes computationally very expensive. In such a scenario, system performance becomes vulnerable if data burden is not reduced. Using EA, the problem specific data can be separated and fused together in the registration process. So far, many researchers have contributed in this direction and proved the usefulness

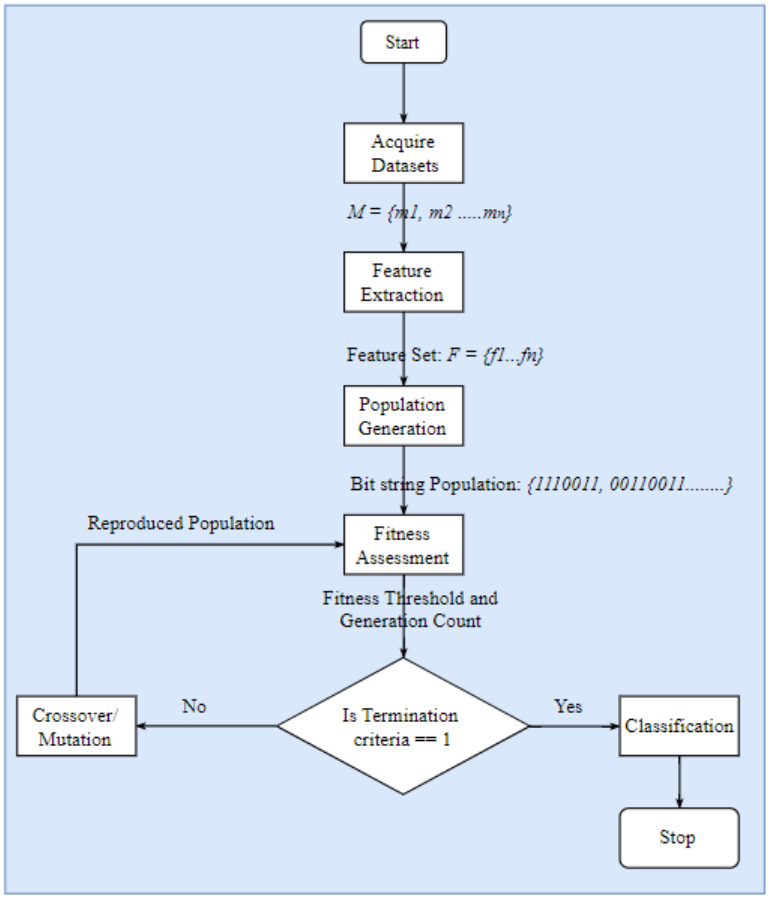


Figure 8: Example Genetic Algorithm Implementation for Feature optimisation

of this method [17, 23, 27, 29, 49, 55, 102, 106, 121]. A scene interpretation process can be performed efficiently and reliably applying the EA approach in case of noisy and obscured sensory data. Best examples for such applications include ocean current classification [7], along with classification and labelling of objects in dynamic environmental conditions [13].

Feature Selection and Classification is another important application in case of sensor fusion. Feature subsets are essential to define properties of the object to be identified. Feature recognition is performed by selecting the only feature specific information from the problem domain. One of the most popular applications uses feature extraction approach is Optical Character Recognition (OCR) [56, 108, 120]. Dynamic optimisation problems are defined as problems with moving optimum. In the case of Dynamic Environmental conditions, a pattern recognition process needs to address a fluctuating optima type of concerns. Tracking moving optimum may prove complex task for deterministic mathematical techniques due to their steady state measurement nature. Apart from the deterministic quality, another paramount matter is a high computational cost. Stochastic techniques are essential for solving pattern recognition problems in large data sizes as well. Self-Adaptive Evolutionary algorithms can provide state of the art solutions for moving optima type of problem domains. A meta-modelling approach can add robustness to Evolutionary Algorithms in case the dynamic environment is to be detected using composite sensory resources.

2.1.3 Concurrent Data Stream Processing

Parallel processing [10, 64] can be an attractive option for processing multiple data streams based on the availability of appropriate computer hardware. Usually, it is

due to the concurrent mapping of processing units improving the system performance. With the current multiprocessor architectures, it is possible to map modalities over individual processor cores. Use of Field Programmable Gate Array (FPGA) can as well provide a flexible and rapid development for concurrent image processing [12]. Figure 9 shows a typical example of multimodal parallel processing architecture. All workflows presented in the dotted boxes represent individual threads or processing cores. As shown in Figure 9, every modality is mapped onto a single core and lastly, an aggregation is performed to calculate the outcome. Few of the examples of this parallel processing approach include remote sensing, biometrics and security applications. Here, we discuss a few of the example research studies under this body of knowledge.

In case of remote sensing applications, a study by Plaza et al. [94] shows an example of parallel processing of spatial and spectral information. Here, both type of information is divided into separate blocks and those blocks are assigned to separate processing units. Whereas, other approaches use parallel processing methods like partitioning and pipelining the data into parallel streams and process it. Another example includes multimodal detection of pedestrians [33] using a computational model and application of a parallel processing hardware architecture to process the low-level pixel data. Thus, parallel processing is as well one of the fundamental approaches, which can be applied to treat massively acquired real-time and parallel sensory data. But many of the parallel processing architectures heavily rely on the hardware and complex scheduling of processes.

Hossain et al. propose a parallel implementation of edge detection [51] using the CUDA framework. In this research, they apply the CUDA framework using a parallel processing method for multiple images in different colour schemes to detect edges. Parallel implementation, in this case, allows improving performance since

computing tasks scheduled concurrently saves times. But, the limitation here is, it is a heavy hardware dependent solution and cannot be easily applied to single processor systems.

Xue et al. apply the parallel fusion method [132] to remote sensing images by splitting wavelet frequency components into some high and low-frequency spectrums. Two parallel operations fuse these two frequency components help to improve it. In this parallel activity sequence of operation include wavelet transforming, fusion and then inverse transforming of the same to achieve fused output. Limitation of this technique is though it presumes all spectral components are with limited or minimum noise and usable for fusion purposes.

Ino et al. propose a parallel image registration algorithm [54] applied by using a hardware cluster. This method predominantly uses multiple hardware sources to acquire input images. This is one of the ways it can be utilized for acquisition of highly scalable and distributed environments. The advantage here is, this method can work with nonrigid registered datasets. Whereas, the limitation here is a very high dependency on hardware underlying for the acquisition of images.

Cheng et al. propose a parallel image fusion algorithm [22] using wavelet packets. In this research work, a set of MxN images is split into P pixel buckets and then supplied to the same number of computers. After applying wavelet fusion finally, all fused images are collected back to a central controlling computer. As the research described here, it is highly hardware dependent work which addresses the scalability issue in the problem domain, as the authors discuss in detail.

Akoguz et al. propose a parallel computing method [5] for data fusion in remote sensing applications. Here, authors apply parallel computing methods to reduce the execution time required for multispectral fusion. The proposed method uses six cores of a Xeon processor for parallel implementation of image fusion algorithms. Due to

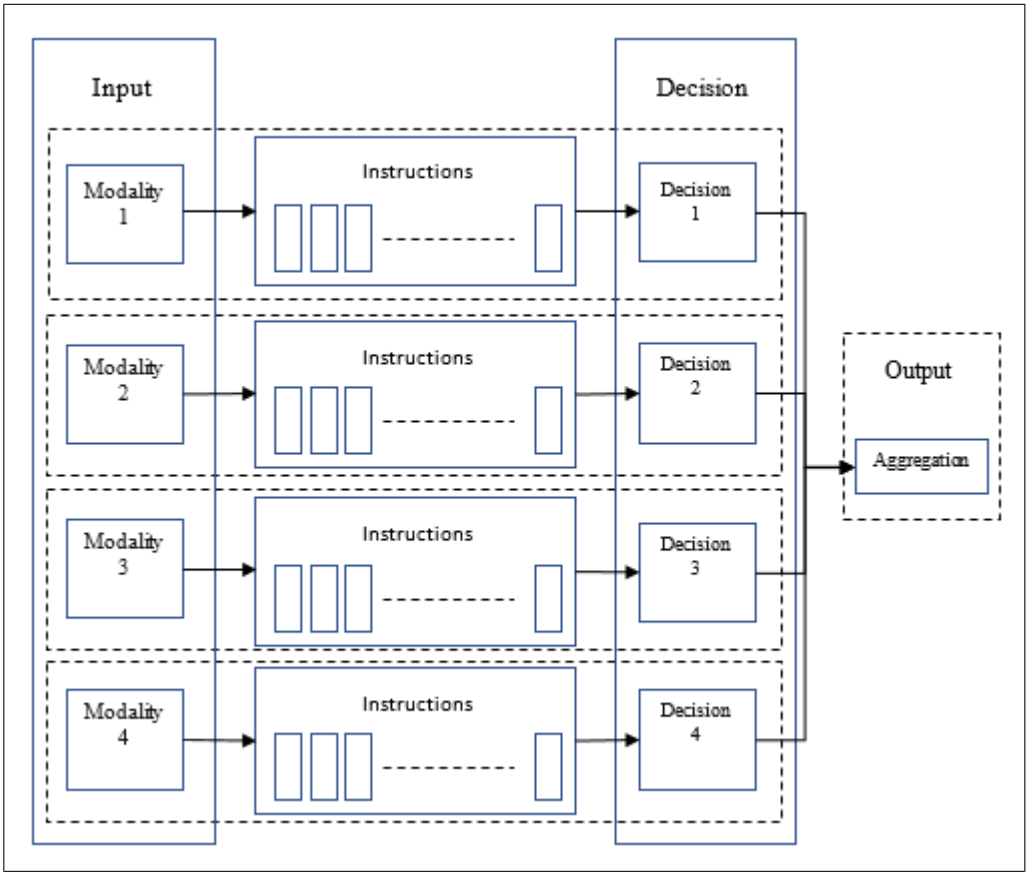


Figure 9: Typical Multimodal Parallel Processing Architecture

concurrent operations of fusion, the proposed system improves fusion performance significantly. In this case, the implementation is specific to the hardware platform used.

Achalakul and Wattanapongsakorn propose a parallel multispectral image fusion approach [2] called a spectral screening principal component transform. This method is developed in parallel and distributed environments. Authors suggest the use of several similar processors to achieve the proposed methodology. Thus, this model improves the performance of fusion operations by using a set of hardware infrastructure.

Rayan et al. propose a parallel fusion computational approach [1] for multimodal image processing. In this work, the authors recommend the adoption of Python programming language libraries in parallel hardware configuration. Authors compare their results against cloud-based and serial implementations. Here, parallel execution is deployed using multiple CPUs with single-core architectures.

Chen et al. review parallel computing methods [19] in remote sensing applications for mosaicking of images. In this study, various approaches to the parallel implementations are discussed which include problem (function) decomposition, data decomposition for image mosaicking parallelisation (master-slave parallel structure), task scheduling (static task schedule strategy, dynamic task scheduling), and Message Passing Interface methods. The authors recognise the limitations concerning portability of these techniques due to variability in underlying hardware platforms.

Zhao and Zhou propose a parallel fusion algorithm [145] based on CPU-GPU combined platforms for remote sensing applications. In this research work, the authors combine the advantages of both hardware platforms to better performance for image fusion activities. Authors propose the usage of CUDA the parallel processing framework for this research work. Authors demonstrate improvement in performance and scalability due to parallel implementation they use. This work is specifically targeted for a CPU-GPU combined hardware platform.

Yang et al. propose a data distributed parallel algorithm [138] for wavelet-based fusion for remote sensing applications. Here, authors advise using a parallel implementation with data load balancing for multimodal fusion using wavelet transforms. Authors claim better scalability as compared to contemporary methods. The proposed architecture is realised using Xeon CPU processors.

Thus, the extensive review conducted for the concurrent data-stream processing methods show they are mostly hardware dependent solutions. In case of a portable

software-driven application, they may not be the suitable choice. But looking at the current trends in microprocessor development these methods will be widely used in the future.

2.2 Multimodal Applications

Depending on the requirements of an application the multimodal sensory detection can accommodate a similar or dissimilar type of signals. A typical example [126] when same signal modalities are used can be a combination of camera outputs in visual and infrared in image forms. These pictures are mostly processed in spatial or spectral quantities, and this combination can be called as homogeneous multimodality. In other cases, when two different forms of a signal are processed the scheme can be put as a composite multimodality. A most commonly observed application of the composite multimodality [40] is using a combination of audio and visual signals. In this combination, the sound signal has a frequency magnitude, and the video is a spatial signal.

As discussed before, the multimodality has its advantages while detecting events and objects in the broader range environments and ambiences. This versatility and flexibility encourage to deploy multiple signal sources in many research studies and commercial applications for detection. This section discusses a variety of applications which use multimodality for signal acquisition. The list of such applications is very long and this discussion tries to accommodate essential and frequent cases. The focus is mostly on the applications within the area of marine, machine-vision and healthcare monitoring. Here, some of the cutting-edge applications are discussed for these two classes.

2.2.1 Marine Monitoring Applications

Nowadays, many computer vision and image processing implementations are used for various marine monitoring applications. Water pollution monitoring, habitat monitoring and event monitoring are the most common examples of marine monitoring applications [88]. Satellite imagery, camera-based detection and visual sensor network are the most known areas of research studies. Multimodality in computer vision applications can assist the water monitoring to enhance reliable sensing in highly reflective and dynamically changing illuminating water surfaces. Here, various multimodal applications are discussed to develop an understanding about their implementations in a marine monitoring space.

O'Connor et al. have developed a multimodal system [88] for detecting events in river water. One of the most significant developments which can be identified includes a river-flooding. Event data is captured by combining an in-situ sensory output with a visual spectrum. Detailed information from in-situ sensors combined with a visual perspective provides very meaningful results in a continuously varying environment of a river. In another study by Kolar et al. combine physical, chemical and biological sensors [69, 70] to monitor the river. Use of a distributed streaming analysis makes their sensory system efficient and high-performing.

Akamatsu et al. apply multimodality to detect dolphins and have tested their method [4] in the Ganges River of India and Irrawaddy of Myanmar. In this study, it proves that a combination of visual and acoustic modalities improves the capability of the detection system significantly. Use of a robotic vehicle for studying an underwater environment using multimodal sensing capabilities can provide very interesting details on the ecosystem there. A study by Rao et al. considers an arrangement of visual and bathymetric modalities [98] for a surveying purpose. This study helps in monitoring and classifying the marine echo-systems efficiently. Kast-

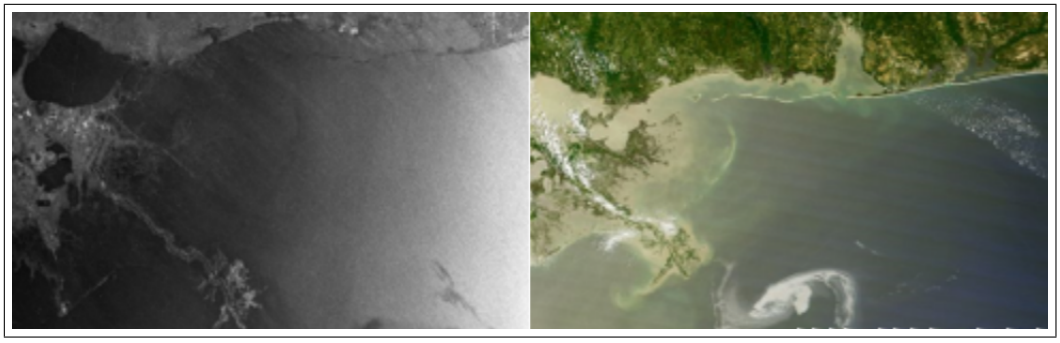


Figure 10: Multimodal Satellite Image for Oil Spill Monitoring

(Image Courtesy: NASA/MODIS Rapid Response Team

https://www.nasa.gov/multimedia/imagegallery/image_feature_1649.html)

ner et al. discuss a sensory platform [63] to monitor whale sharks and planktons. Their study indicates possible future opportunities by using various sensory platforms for marine life monitoring purpose.

In another work by Couture et al. propose a multimodal system [25] to monitor algae in a river. The authors use a feature based detection approach for the detection of the river surface and to find the presence of algae. Sullivan et al. propose an integration [116] on environmental monitoring modalities with decision support systems. This review covers various aspects of the multimodal marine monitoring process including modalities and types of marine environment. Multiple cases using multimodality are discussed by authors to provide a summary. Many related departments from various governments are actively seeking use of this technology for maintaining the cleanliness and detecting events.

Monitoring of oil slicks after an oil spill has become an ongoing activity [142, 113, 114] due to the frequent incidences taking place. Use of satellite imagery in various modalities can provide a certain level of details. The highly scalable data which covers an enormous geographical area, the processing tasks becomes very

challenging. Wan et al. propose an overview for remote sensing method [123] to be used for monitoring of oil spill in the Gulf of Mexico. This study also discusses a successful use of multimodality for detecting trajectory of the oil spill. Kubelka et al. propose a multimodal sensing and fusion [74] of oil spills using a robotic vehicle. The Autonomous underwater vehicle (AUV) sensory system coordinated with the conventional instrumentation is used to survey the sea water. The study mainly tries to investigate a method for organising between the AUV and geospatial sensors for an automated water sample collection. Thus, multimodality with multiplatform approach can develop a co-operative environment to solve this complex task.

2.2.2 Multimodal Biomedical and Healthcare Applications

In biomedical and healthcare application invasive and non-invasive techniques are used to detect human body parameters. Most common approaches for sensing the parameters include the use of imaging techniques (visual, magnetic, infrared, etc.) [111] and physical parameter sensing (vibrations, temperature, liquid level, etc.) [20]. This subsection discusses the most common biomedical applications which practice multimodal sensing technologies.

A malignant tumour is frequently detected by using multimodal detection methods. For enhancing accuracy in a breast cancer detection, Conceicao et al. suggest a multimodal method [24] which combines Positron Emission Mammography (PEM) and Ultra-Wide End (UWE) Radar signals. The purpose of adding the UWE to PEM is to avoid the lapses which happen due to PEM. This method uses a feature based detection for detecting tumour from images and authors claim it enhances the classification accuracy. In another implementation, Romain et al. use a multimodal camera capsule [101] to detect the internals of the human digestive tract. The pur-

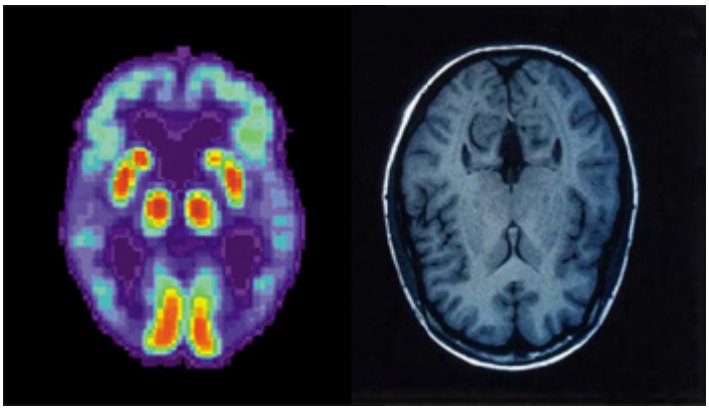


Figure 11: Multimodal Brain image

(Image Courtesy: Israel Institute Technology Image: http://www.cs.technion.ac.il/~mbron/about.html)

pose of this method is to investigate colourectal cancer which is a common cause of cancer-related mortalities in the western world. The system uses a combination of infrared and visual cameras to improve the detection method.

A use of multimodality for frequency domain signals to determine chronic illness can improve the diagnosis results. Challenge in this class of methods is about providing an accurate verdict using waveforms provided in various dimensions, shapes and frequencies. Kenneth et al. suggest a method [65] to detect multimodal cardiovascular signals by capturing a heterogeneous data for finding patient deterioration. The success of this method is described based on the use of an intelligent algorithm to mimic the human decision-making process. Calvert et al. have proposed audio-visual targets [15] for perceiving human brain response in human-computer interfacing applications. The authors analyse the electroencephalographic (EEG) signals against the multimodal target responses.

Other healthcare support applications include the use of technology for addressing a specific health problem or impairment. Use of multimodal sensory technology

to support visually impaired people can improve their lives. Yu et al. propose a walking-assistance system [140] to help sight-impaired people by converting visual information on obstacles in surroundings into voice data. In another example, a system to recognise human emotions using an audio-visual data of subjects [14] is developed by Busso et al. where a smart algorithmic implementation is used to analyze facial expression and voice data. This system provides an automated solution to detect depression levels in a person through the use of technology.

Another most critical example of the multimodal biomedical system is detecting eye disease. A combination of videos of an eye was analysed by Floeth et al. for accurately finding the location of lesions [35]. A registration algorithm was proposed to acquire multimodal video signals with minimum error. Zana and Klein have developed a registration method [141] for a system which obtains fluorescence in and green images of eyes. The purpose of this system is to help in finding a best possible registration strategy to provide accurately registered multimodal data.

Thus, based on this review on multimodal biomedical and healthcare systems, it can be said that multimodality plays a significant role in this domain. The most common aims include enhancing accuracy and introducing or increasing the automation level in the diagnostic process. Since the medical profession demands very high precision, the multimodality is accepted even at the expense of higher computational load. Another critical concern is related to the registration of multiple data sources. The accuracy of the registration, in this case, will determine the success level of the detection process. If the registration error is too high, then the further analysis may lead to the incorrect diagnosis results.

2.3 Summary

This chapter discusses three main categories of multimodal object detection and image processing algorithms. The fusion methods combine multiple modalities with a corresponding effect. Whereas, the limitations of these methods include a high dependency on registration and inability to recognise absence or corruption of incoming data streams. The machine-learning techniques rely heavily on the input dataset representation and training. The training part of these methods consumes plenty of time plus also required a vast number of historical datasets. Finally, the parallel fusion methods are most suitable for computer-hardware based implementations and may not lack portability. Use of adaptive techniques along with fusion and classification algorithms may be able to improve the detection accuracy and performance.

3 Evolutionary Multimodal Computational Models

A multimodal approach is used for concurrently acquiring information from several nodes which can generate a massive amount of data. This data can be analyzed for detection of objects, anomalies and other useful discoveries present in it. But, due to scalability and variability within this information, it is very challenging use traditional deterministic computing methods. To overcome the challenge of variability an adaptive computing model is required which will move its optima according to the changed operating conditions.

Applying algorithms inspired with evolutionary biological concepts can be used for an efficient search and optimisation operation to overcome this variability. Implementation of these algorithms needs special consideration for its software framework to achieve goals related to concurrency and heterogeneousness in datasets. Here, we propose a set of evolutionary algorithms to optimise datasets at modality, feature and data level. Here, a hierarchical structure of multimodal datasets is kept in mind to apply these algorithms at different levels.

This chapter discusses mathematical modelling, algorithm pseudocodes, formal validation methods and limitations of the proposed algorithms. The validation process uses qualitative and quantitative ways to assess performance, scalability and repeatability of the algorithms.

3.1 The Origin and Rationale

3.1.1 Multimodal Evolutionary Approach

Unimodal data acquisition provides datasets for a single type of environment in a limited range of parameters. In the context of computer vision applications, the il-

lumination is one of the most important and dominating factors to recognise objects [89]. In case of limited lighting in a monitored environmental scenario, a unimodal-visual dataset provides limited detection accuracy. This limitation is because the acquired datasets are prone to a noise addition resulting from the illumination variability. In such a scenario one spectrum of data obtained using the only modality may not be enough in many application domains. Thus, expanding another dimension to visible spectrum becomes a need and can be achieved by adding infra-red or some other kind of data acquisition system.

A typical implementation of such a multimodal detection system [125] can be obtained by expanding visibility in highly illuminated and darker conditions by using a combination of passive and active vision. The active vision system can add infrared or any other non-visual frequency and capture its reflection rebounded from surroundings. Combining or selectively processing these parallel acquired modal-datasets can be done for an object detection goals.

Dynamic optimisation problems [137] are defined as problems with moving optimum. In the case of Dynamic Environmental conditions, the patterns recognition process needs to address the variable type of datasets. Tracking moving optimum kind of problems is beyond the capabilities of deterministic mathematical techniques due to their steady-state nature of measurement. Apart from their deterministic tendencies, another critical concern for mathematical approaches is the computational cost – there is a high variability for large size data handling and processing. Stochastic techniques are essential for solving such kind of pattern recognition in large data sizes. Evolutionary algorithms can provide a state of the art solution for moving optima type of problem domains due to their Self-Adaptive operating nature. Meta-modelling approaches can add robustness to Evolutionary algorithms in case of dynamic environment detection using heterogeneous sensory resources.

3.1.2 Evolutionary Stages and Operations

Implementation of Evolutionary approach [26] needs to consider essential concerns like the computational cost of the algorithm or in fitness function and noise in fitness functions. Adaptivity of evolutionary approach to a particular domain depends on specific strategies of evolution for the selected environments. The genetic algorithms applied in case of the dynamic environmental detection using multimodal heterogeneous sensory resources typically employ the following operations:

- **Encoding:**

  In case of the Evolutionary algorithms, a solution needs to be encoded using either in a binary or any other suitable form. This conversion needs to be specific enough to describe the problem metadata using the evolutionary biological concepts like phenotypes, alleles and genetic structures. This whole operation is helpful to represent the metadata of the acquired datasets in a bit-string pattern allowing to apply bit-level manipulations to create further variations.

- **Population:**

  Population initialisation process needs to be developed from the critical perspective that the size of the population has a direct effect on the performance of an algorithm. The ultimate aim of any search and optimisation heuristic approach is providing an advantage over conventional random search methodologies. A large data set (population) tends to influence the solutions (evolutionary or heuristic), so they appear to behave like a conventional search and optimisation algorithms. Thus, some mechanism to select situation specific "horses for courses" best suitable population is needed. Since environmental

dynamism keeps changing elitism definition for the population the initialisation process needs to address this problem. Either repopulating and initialising with changed elitism criteria or recycling of the same population with new elite selection can serve for changed situation.

- **Reproduction Processes:**

  A reproduction process in case of evolutionary algorithms is an operation to generate next generation of solutions or chromosomes by mating initialised or previously mated chromosomes. Depending on an encoding strategy selected, the reproduction probability and frequency can be decided. Frequent operations like mutation may make behaviour and performance of an algorithm synonymous with random search methods. Reproduction strategy is fundamental criteria from algorithm performance point of view. Two major classes of reproduction are crossover and mutation.

- **Fitness Assessment:**

  Every generation of the population needs assessment using a function to evaluate every member for its level of fitness. These fitness values are further used for ranking the population members for selection and then reproduction purposes. Any change in operating environmental parameters gets reflected through the fitness value of an evolutionary algorithm. The environmental variance can either change the optima position or add noise in fitness value. Thus, the variability of optima is one of the significant concerns needs to be addressed.

Knowledge of objects to be detected along with environmental variations can add substantial value to the Computer Vision application domain. Knowledge acquisi-

tion process for a system is performed either by training of system of pre-acquired data sets for similar or same application in the same environment. In case proposed research study, the knowledge acquisition process will include the following steps:

- **Developing Object Knowledge Base:**
  Every object can be distinguished on the basis its noticeable features. Object specification using different visible and geometrical parameters like surface colour, area, edges, centroids and so forth.

- **Defining Environmental Conditions:**
  Environmental conditions can be determined using various perceptible aspects. All possible environmental conditions will be measured and their maximum thresholds will be recorded and essential cues.

- **Object Identification in Individual Modalities:**
  Apart from visible elements, the objects will be assessed in other modalities and images will be stored.

- **Object Identification by Multiple Modalities:**
  Objects will be scanned using all available imaging sensors in ideal environmental conditions.

- **Real-time or Periodic Knowledge Update:**
  In case of dynamic uncertainties, a periodic update is needed for informing changes to the knowledge base.

- **Sensor Prioritizing:**
  Based on historical knowledge and periodic knowledge update the effect on camera sensory resources needs to be judged. This new understanding can be

helpful in prioritising camera sensor signal inputs acquired by the system to decide evolution strategies.

- **Sampling Rate for Knowledge Update and Data Acquisition:** Based on real-life circumstances and domain-specific requirements the signal sampling rate needs to be adapted.

3.1.3 Validation of Multimodal Detection Methodology

In a real-life evolution is a physical and temporal phenomenon. What it means is, the progression of species is a time-dependent function over its genetic structure affecting phenotypic adaptation [3]. Thus, a visible appearance and the underlying structure adapts to a changed scenario which can be either a gradual or random response of the operating environment. Another evolutionary behaviour observed in species can be categorised as a population level adaptation [50] to a changed environmental scenario. Here, to survive through harsh conditions, species can change their reproduction strategies. Finally, we also take into consideration a time-bound nature of evolution [11] where phenotypic characteristics are shared with future generations without changing a genetic structure. Based on these observations, the proposed evolutionary multimodal method is implemented and tested using a family of algorithms. Every evolutionary concept is applied in the form of an algorithm and then tested on a multimodal dataset which provides a simulation of varying environmental scenarios discussed before.

3.1.4 Merits

Evolution of species is a response for adapting to a changed environmental scenario. Capability to respond is an inherent and intelligent mechanism which gets triggered based on the environmental cues received by species. In the case of the proposed evolutionary multimodal method, this adaptation provides multiple advantages.

One of the significant benefits is an improvement in the accuracy of an object classification process due to the selectiveness of feeding datasets to the object classifier. The second advantage is, an improvement in performance since the reduction in datasets reduces the number of input nodes of a classifier. In some cases, it is observed that the precision of the classification can be improved since evolution increases the stability of the datasets fed to a classifier.

Finally, since the same amount of the information can be supplied to classification algorithm with a lesser number of nodes, the overall scalability can be improved with a marginal percentage value. Thus, evolutionary algorithms along with traditional statistical methods can combine their advantages and improve the detection architecture computationally.

3.1.5 Drawbacks and Constraints

The evolutionary algorithms are often criticised for various reasons. One of the most important criticism is, they fail to provide a same and consistent result for a problem solved. In case of genetic algorithms, an outcome for every optimisation cycle can vary with a smaller degree which may not make them suitable for some high precision applications.

Another drawback is that it is a cyclic process which may take an extended period to solve a problem and provide results. In many cases, due to the limited

timeframe, the evolutionary approach may not a practical solution. It can be an excellent offline learning mechanism but may not be an efficient method to be used in real-time systems.

In case of a non-diverse population scenario, the optimisation process may converge prematurely providing an inaccurate result. The most important constraint is to implement an evolutionary computation model is to understand a problem thoroughly and then model its metadata. Thus, the proposed methodology has few limitations which will need to be considered while implementing it.

3.2 Mathematical Apparatus

This part of the chapter describes essential mathematical concepts related to the Multi-modal detection operations. The analytical interpretation of the overall process starting from data acquisition to object classification is described. It is an essential part of the methodology since the mathematical description of the problem space is required to code it using a programming language to provide a solution.

3.2.1 Image Processing and Fusion

Any image I can be mathematically interpreted as a two-dimensional matrix of pixels with every pixel either a singular value or a vector representation of a colour-space.

- Following equation can be one of the representations of an image:

$$I = \begin{matrix} f(x_1, y_1) & f(x_2, y_1) & .. & .. & f(x_m, y_1) \\ f(x_1, y_2) & f(x_2, y_2) & .. & .. & f(x_m, y_2) \\ .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. \\ f(x_1, y_n) & f(x_2, y_n) & .. & .. & f(x_m, y_n) \end{matrix} \tag{8}$$

Where:

I = probability density function of image pixels and (m, n) are pixel coordinates.

- Area of an image or a region on image:

$$A = \sum_{i=1}^{n} P_i \tag{9}$$

Where:

P = a pixel count and

n = number of pixels in an image.

- Object Centroid can be defined as follows:

$$\bar{C}_{(x,y)} = (\bar{x}, \bar{y}) \tag{10}$$

Where:

$$\bar{x} = (\sum_{i=1}^{m} x_i / A) \tag{11}$$

$$\bar{y} = (\sum_{j=1}^{n} y_j / A) \tag{12}$$

Spatial Image fusion techniques are represented either as a mathematical sum or selection of pixel values located at the same location on two or more images.

Mamdani spatial fusion techniques [96] can be represented as below:

- Average Pixel Fusion:

$$I_f = ((C_0 * I_1) + (C_1 * I_2))/(C_0 + C_1) \tag{13}$$

- Maximum Pixel Fusion:

$$I_f(x, y) = max(I_1(x, y), I_2(x, y)) \tag{14}$$

- Minimum Pixel Fusion:

$$I_f(x, y) = min(I_1(x, y), I_2(x, y)) \tag{15}$$

3.2.2 Bayesian Fusion

Bayesian theorem [129] is used for an inference level fusion in the multimodal object detection.

$$P(H/E) = (P(E/H) * P(H))/P(E) \tag{16}$$

Where:

- H= A specific hypothesis which may or may not be a null hypothesis
- E= an observed evidence

- $P(H)$= a prior probability of H
- $P(E/H)$= a conditional probability of E to see if H happens true
- $P(E)$= A marginal probability of E is a priori probability of witnessing the new evidence E under all possible hypothesis
- $P(H/E)$= A posterior probability of H given E and new estimate of probability is estimate of the probability that hypothesis H is true, taking evidence E into account.

3.2.3 Evolutionary Computing

In case of evolutionary computing processes following are the major mathematical concepts used during the thesis.

- *Encoding:* Given a set of input parameters are converted in a binary or character string: $1, 1, 0, 0, 1, 0, 1, 0$ here every bit in this set is called as a gene. Whereas, a group of genes simulates an allele.
- Crossover is a function to switch bits at one or more fixed locations:
  - *Input:* $StringA = 1, 1, 0, 1, 1, 0, 1, 0$ and $StringB = 1, 0, 0, 0, 1, 0, 0, 0$
  - Crossed over product (example by swapping of second and seventh bit from String A and B)
  - *Output:* $StringC = 1, 0, 0, 1, 1, 0, 0, 0$ and String $D = 1, 1, 0, 1, 1, 0, 1, 0$
- *Mutation:* Random swapping of bits in encoded strings: $1, 1, 0, 0, 1, 0, 1, 0 -> 1, 1, 1, 0, 0, 0, 1, 0$

- *Fitness Check:* Either application specific error rate in its classification algorithm

3.3 Adaptations of Evolutionary Strategies

3.3.1 Genetic Algorithm for Feature based Object Detection

Here, an evolutionary multimodal computer vision algorithm for object detection is proposed. In this solution, a genetic algorithmis developed for detecting specific object features[59, 16] in multimodal datasets. The proposed method utilises a simple, robust and high-speed Algorithm to identify objects accurately. In this algorithm, the system is trained with an object database and using the most significant visual features for every class of objects. Images are assessed periodically for detecting the 'Region of Interest (ROI)' within a picture from a dataset. The ROI is defined as an area in which a presence of object features is detected. The ROI detection is achieved by applying a random sampling of pixels and an assessment of colour threshold of every pixel. An object centroid and area are identified as critical features by using the proposed algorithm. The colour intensity as well is assumed as one of the traits for classification that is based on the training data. Then those parameters are classified determine the accuracy of the algorithm.

The process starts with inputting of an object to be searched and the search area or an Image frame. The object and search areas are measured by its dimensions and colour threshold band or range of an object. Then the purpose is divided into 'n' equal sections, each section dimensions proportional to the size of an object. In the next step, a population of pixels is generated in every section ($population_{size} = m * n$). For finding an elite set of pixels the entire population is thresholded at a colour threshold level equal to an object. If the elite count is below $(m * n) * k$ then population generation process is performed again. Else, parent chromosomes

Algorithm 1 Feature Based Object Detection using Genetic Algorithms

1. **begin**
   $DS = Read(Dataset)$
   $DS = \{I1, I2..I4\}$
2. **for all** i $i \leq g$ where g in number of generations
   Generate population $p = \{p1..pn\} = rand(DS)$ where every $p = x, y, s, m$
   $k = size(p)$ size of the population
3. **for all** i $i \leq k$
4. Check fitness of every p where fitness function is $f = sqrt((1 - square(c1^2 - c2^2)/square(c1^2 - c2^2))$ where c2 is a centroid of the detected RoI and c1 is a benchmark centroid.
5. **if** $max(fitness(p) > threshold$ or $generations < max$
6. **if** $temination = 1$ then break
7. else
8. sort(p)
9. ps = select(p)
10. l = size(ps)
11. **for all** j $j \leq l$
    $c_p = crossover(0.8 * ps)$
    $c_m = mutate(0.05 * ps)$
12. **end for**
13. **end for**
14. **end for**
15. **end**
16. PS_{final} = final results or the solution candidates

are created at every elite pixel. Fitness level is evaluated by passing the entire elite population through fitness function. After the fitness check, if the optimisation level is achieved then the process is stopped else process is continued by performing crossover and mutation. After every crossover/mutation operation, the fitness level is re-evaluated and if optimisation level is achieved process is stopped.

3.3.2 Genetic Algorithm for Feature Subset optimisation

Here, a genetic algorithm with an application specific modification[58] is proposed. The algorithmic process involves training and validation modes of operation. The training process requires modelling and optimising of the object feature set. It uses historical image data to investigate patterns within a feature form. Whereas, the validation or work mode confirms the detection accuracy of the algorithm using real-life data. Pseudocode and data-flow are explained in details as bellow(Alg. 2, 3 and Fig. 12).

Algorithm 2 Evolutionary Feature Selection Training Algorithm

1. **begin**

   Input historical image set:$I = I1, I2, I3.....In$

   Extract regions of interests (ROIs) from I using a function Ø:$R = \emptyset(I) = R1, R2....Rx$

2. **for all** i $i \leq g$ where g in number of generations

   Apply labels to every ROI: $L = L1, L2, L3.....Lx$

   Extract y features from R and organize them in a matrix F with size $(x * y)$: $Fx * y$

3. Apply a feature optimisation algorithm to optimize F to $FOx*m$: $FOx*m = Opt(F)$

4. Find f1 & f2 $\subseteq$ FOx*m, with high and low occurrence probabilities p1 and p2 in FOx*m: $p1 = p(f1) and p2 = p(f2)$

5. Find best fs: where $fs = (h * f1) + (l * f2)$, h and l represent high and low quantity selection coefficients;

6. **if** termination = 1 then break

7. Store fs as a training output

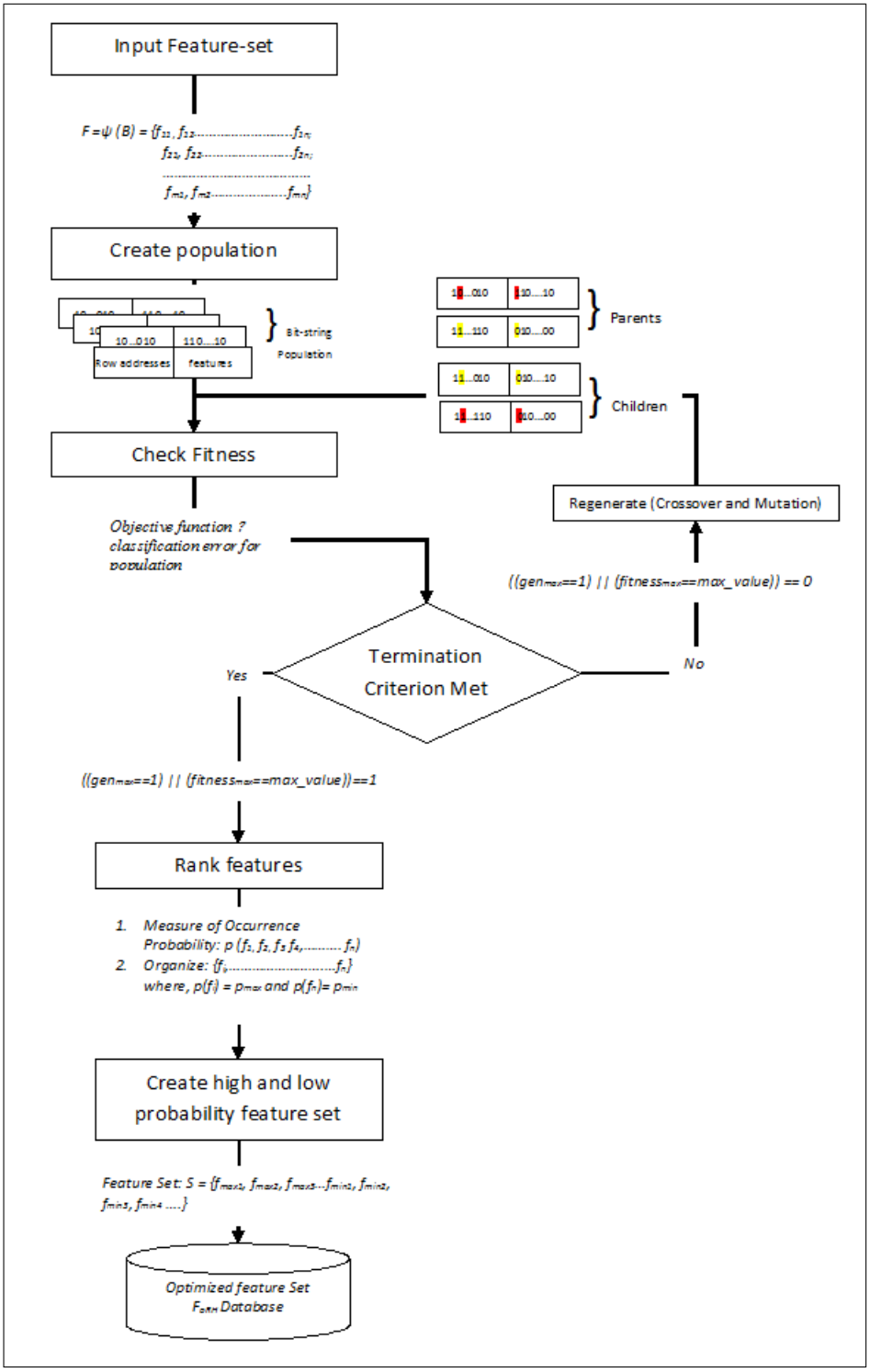


Figure 12: Flow Chart of Enhanced Genetic Algorithm for Feature Selection

Algorithm 3 Validation Algorithm

1. **begin**
2. $V_{DS} = \{I_{1,}I_2.....I_r\}$
3. $[r, c] = size(V_{DS})$
4. $O_c = []$ declaring an empty array
5. **for all** i $i \leq r$

   $RoI = \emptyset(I)$ here, Ø can a thresholding or an equivalent function to extract an RoI

   $f_r = f((RoI)$ in this case f is a feature extraction function

   $O = b(f_r)$ where b is a Bayesian Classifier

   $O_c = [O_c, O]$
6. **end for**
7. $CPM = compare(O_c; O_a)$
8. Stop

Feature Extraction

In this scenario, a feature is defined as an entity which carries valuable information. A combination of elements can help to determine objects which exist in an image. Most commonly used object features can be either in a spatial domain-such as points, edges or shapes or spectral descriptors for colour, shape or images in a spectral domain. From a real-time performance perspective and for reducing the computational cost we have decided to process interest descriptors and geometric features. In our proposed feature extraction method, we use colour pixels

proportionally sampled according to the dimensions of the region interest which are acquired using a blob detection algorithm. We also add shape features in the spatial domain which include area, centroid, fullness and lengths of major and minor axes. The entire feature extraction process is described below:

1. Detect Blobs using chain code algorithm

2. Extract 17 texture/geometrical features: Area, Centroid, Fullness, Major and Minor Axis, etc. and organize as shown in equation 17.

$$f_g = A, C, ful, Ax_maj, Ax_min, ... \tag{17}$$

3. Based on blob dimensions, locate region of interests (RoIs) on the colour image

4. Calculate length l and width w of bounding boxes

5. Sample points proportional dimensions of the blob (next subsection provides further deatials about this method)

$$f_p = p_1, p_2.....p_a \tag{18}$$

6. Arrange extracted points in matrix form

$$FM = [f_1; f_2...; f_n] \tag{19}$$

Where every f is

$$f_i = \{f_g, f_p\} \tag{20}$$

Pixel Sampling Algorithm

The pixel sampling algorithm involves extraction of colour samples from a detected RoI. It is an adaptive process which adjusts the sampling rate according to the variability in length and width of the RoI. Figure 13 provides an overview of the colour sampling matrix. The sampling process is explained as below:

1. Based on blob dimensions and locations, make a bounding box on the colour image,
2. Calculate length l and width w of the bounding box
3. For required $totalsamples$ acquire a pixel periodically over the length and width of the blob, and
4. Remaining samples ($totalsamples$–$(P_{length} * P_{width})$ are sampled randomly.

3.3.3 Bet Hedging to Reduce Risk in Uncertain Environmental Conditions

Adapting to uncertainty is a challenging task in multimodal feature-based supervised learning algorithms. Current optimisation schemes can provide relatively limited success since the focus is to reduce redundancy in datasets. The reduced redundancy allows improving classification accuracy but does tend to make learning specialised to specific operating environments. Such a scenario requires a diverse enough genetic pool make the population adaptive to multiple working environments. Thus, a particular method for adapting to environmental variability is needed. Here an

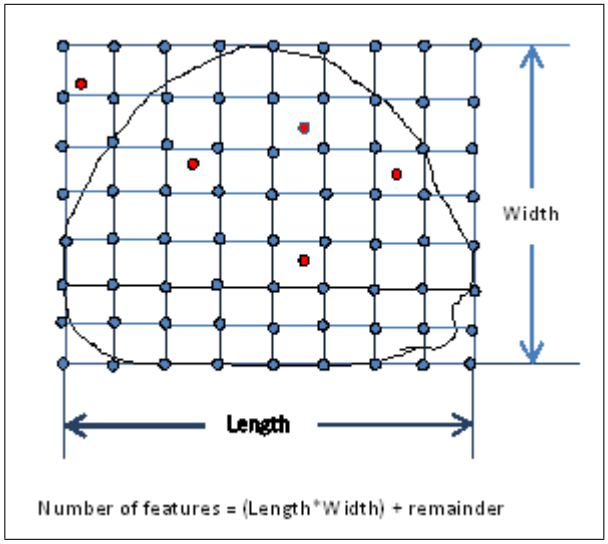


Figure 13: Pixel Sampling Matrix

approach[58, 60] inspired by an evolutionary biological principle called 'Bet Hedging' is proposed. This evolutionary concept is applied by species to improve their sustainability over different environmental conditions. It is achieved by compromising fitness level in stable environments to make species more generalist than environment specialist. In the proposed algorithm datasets addressing different situations are used to create a combination of a generic feature pool. The hedging principle is applied to generate a feature-set which discusses the most probable environmental scenarios. Attaching feature values from uncertain environmental conditions compromises its traditional classification accuracy but improves the adaptivity in random situations.

1. Detect and optimize featuresets $A, B \ldots, N$ as representatives of environments $a, b, \ldots, n$

2. Based on knowledge on environments assign one of the featuresets as a main or a master status, here assume A

3. Find differences between all featuresets with $A : d1, d2 \ldots dn$: $d1 = B\ A$, $d2 = C\ A. \ldots .dn = N\ A$

4. Calculate classification error of A when training is A and when training is performed using B, C ...

5. Establish error levels when $d1, d2 \ldots dn$ are systematically added to $A : e1, e2 \ldots .en$

6. Calculate Hedging coefficient:

$$h_c = 1 - (s/(a.s + u_i)) + (u_s/b.u_s + 1) \tag{21}$$

Where:

h_c = hedging coefficient (value from 0 to 1)

s= feature set size

a= feature set ration coefficient

u_i =uncertainty measure or classification error in percentage

u_i =number of environmental conditions addressed

3.3.4 Phenotype Switching for Adapting to Unpredictable Environmental Variations

Many living species are continuously living in a variable and unpredictable living conditions. The natural ability to respond to the changing circumstances using an in-built adaptive mechanism helps them to survive over the range of environmental states. Adaptivity in an individual organism can be achieved by varying its particular phenotypic traits [84]. Phenotypes are observable characteristics of species and have an effect of surroundings over its magnitude. However, genotypes form a genetic makeup which may have one or several phenotypic representations. The

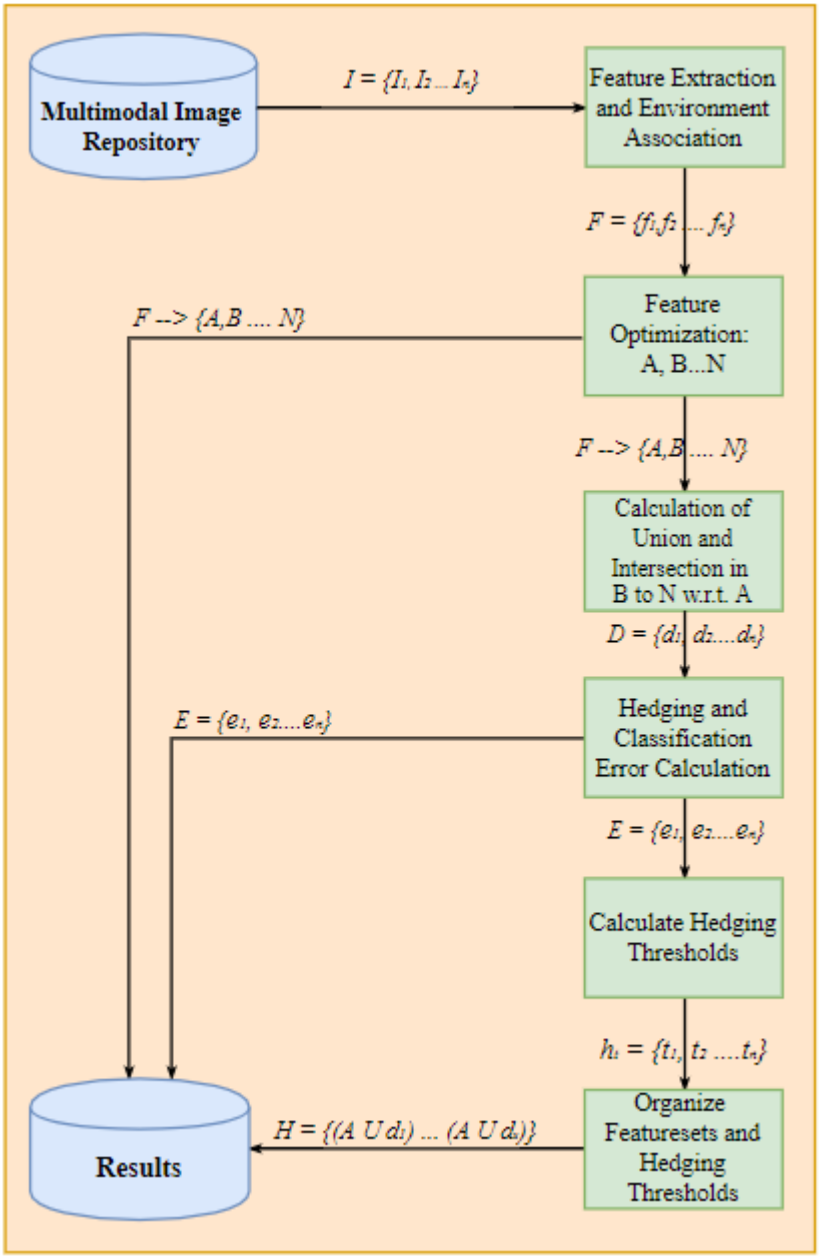
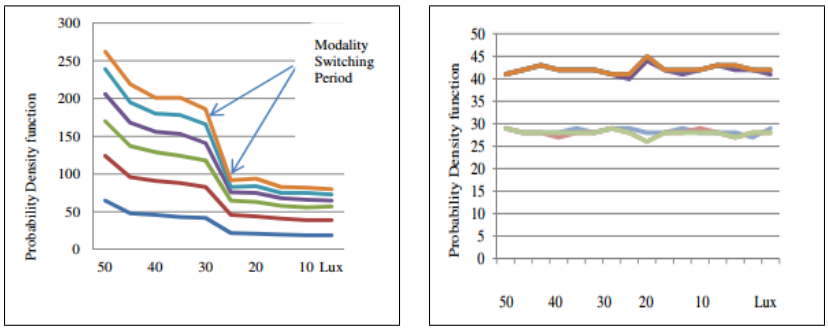


Figure 14: Bet-Hedging Algorithm



(a) Visual Spectrum (b) Infra-red Spectrum

Figure 15: Reaction-norms for Modality Switching using RGB colour Properties

association between phenotypes and genotypes can be represented using an equation given below:

$$P = G + E \tag{22}$$

Here, the phenotype P depends on its Genetic makeup, and environmental states depend on its surroundings. One of the best examples of phenotypic adaptivity to the environment is a difference between aerial and aquatic leaves of the plants. Evolutionary biologists commonly interpret a phenotypic variability as phenotypic plasticity [28]. Phenotypic plasticity has range and costs associated with it. Phenotypic values between its persistence and extinction limits can be called a range of Plasticity. The functional relationship between an environment and a Phenotypic Plasticity of a single genotype can be plotted and is named as Reaction Norm or Norms of Reaction. Thus, this interrelationship plot can be used as a cue to predict the environmental state. This inspiration is utilised to develop an algorithmic method [57, 60, 62]which allows switching between modalities.

Deployment of more than a visual modality to detect objects in a dynamic environment is a common practice. Here, every monitoring modality senses and extracts

information which addresses different environmental states. As an example, a fusion of visual and thermal or infra-red spectrums can be deployed and used in variable lighting conditions. Due to the massive size of information acquired, processing may become expensive and sometimes lead to inaccurate results. Thus, improvement of performance and robustness needs particular attention. A feature-based detection method reduces information size to be processed by representing every modality as a feature set. The matrix in equation 19 depicts a typical representation of the multi-dimensional numeric data.

Proposed Feature-set Selection Scheme

1. Training Process: Reaction norms of all features of all modalities over the environmental range are plotted.
2. Same features acquired from different modalities are compared for finding the highly plastic features with respect to environments.
3. Select highly plastic features as environmental cues.
4. For any geometric variation representing environmental change is detected the corresponding modality is selected.

Feature Subset optimisation

The feature set extracted in the previous step includes excessive, redundant features and thus implies an effect on classification error and computational cost. The genetic algorithm (GA) method proposed for feature optimisation can be fine-tuned through encoding plus an addition of another mathematical operator. This subsection discusses the proposed feature optimisation technique with all internal processes in detail.

1. **Chromosome Encoding:**

   We propose an encoding which assumes bi-directional search in the feature matrix. Two directions of search include rows and columns of the feature matrix to explore feature space thoroughly. Every chromosome gene represents a row number or feature number, along with the chromosome encoding.

2. **Crossover and Mutation:**

   The crossover and mutation operations are operations to create children population by swapping genes at fixed and random locations. In the case of crossover operation, we propose swapping at row alleles and feature alleles at multiple points to create a new diversified population. Whereas, in the case of mutation, the swapping will happen at a random location within an allele.

3. **Fitness Check:**

   We propose using a traditional parameter for the fitness check which is an error in the feature classification. Here, we exclude the other parameters including the number of features because we are proposing the predetermined length which will much lesser than the extracted one.

4. **Selection:**

   We propose a tournament selection approach to select most fit population for the regeneration operations. After every cycle of the regeneration, the topmost candidates are added to the selection pool.

5. **Termination Criterion:**

   The termination criterion, in this case, will be a threshold value representing the count of generations to spend enough time before the fitness saturation.

Feature Ranking and Hedging

The feature ranking and hedging process investigate the occurrence or frequency of presence of the features found after optimisation using GA. In this case, a routine to count the event is used to separate the most commonly and rarely occurred functions. The outcome of this process is used to develop a set of features representing both frequently and infrequently occurred features in different proportions. Various combinations of both types are tried in this set to find a best possible combination. The purpose of this entire exercise is to mix those rarely occurred features as compensation for adapting to the environmental fluctuations.

Feature Classification

The focus of this work is on the feature optimisation and selection process we propose a most commonly used classifier for such applications, i.e. an SVM classifier. The proposed classifier provides the following advantages: Choice to select different kernels to make classification more flexible, Suitable for nonlinear feature combinations, robustness against partially biased training sets.

3.3.5 MEgA: A Multimodal Epigenetic Algorithm

An innovative approach is proposed to improve the robustness of the multimodal detection problem. The acronym MEgA stands for a Multimodal Epigenetic Algorithm. It is heavily inspired by the theory of Epigenetics and uses concepts like epigenetic modelling [91]. This approach which tries to improve the reliability of detection in a multimodal data-stream is proposed.

DNAs have an overriding master-code formed by Carbon Tags CH_3which decide the availability of genes or group of genes for transcription. The outline of this attachment is called as methylation and it characterises superseding mask. Genetic

transcript formed due to methylation allows availability of only specific proteins at a given time. Bees are one of the examples where a genetic difference between the fecund queen and female workers is determined by methyl tags based on their genetic transcript. These tags are formed due to a type of food they are fed when they are larvae. Those larvae one of which are to be specialised to be a future queen are supplied with a particular food which is not fed to other worker larvae. In nature, many viruses, bacteria and chemical pollution can influence the genomic pattern to create a new transcript. Other influential entities may include the human body's immune system, response towards stress, the process of ageing as well can affect genetic representation due to methyl tags.

The growth of a species is not only dependent on a genetic structure responsible for creating protein but also depends on environmental factors [99]. Environmental factors do influence genetic code and its translation into physical properties. Epigenetics is a system which overrides on the genetic structure and controls the expression of the genetic information by switching on and off specific genes. In case of the epigenetic representation, the example where particular genes in a child are active due to the genetic information passed by the maternal and paternal sperms or eggs.

This process of stamping over selected genes due to external influence is called as 'genetic imprinting' [78]. The genetic transcript is thus an inherited from maternal and paternal genes. Imprinting can be a reversible process with next generations. Faulty genes due to mutation influences expression of a genetic transcript since it overrides on the genetic structure. Information genes carry to convert a group of genes into protein is called the genetic code.

Algorithm 4 Epigenetic Algorithm

1. **begin**
2. $A = M1, M2.....Mn$
3. Divide every M into smaller matrices: $M->m1, m2.....mx$
4. Acquire fixed number of pixel samples from every 'm' where p has fixed number of pixels $m_x->p1, p2.....pz$
5. Convert mx into colour, texture and geometric features $m_x->p1, p2.....pz->\{a_1, a_2...a_n\}, \{f_1, f_2....f_n$
6. **while** $gen_{count} \leq maximum$
7. Encoding:

   a) Matrix Encoding: $F_i = f_{11}, f_{21}....f_{n1}; f_{12}, f_{22}....f_{n2}..f_{1m}, f_{2m}....f_{nm}$

   b) Serial(DNA) Approach: $F_i = \{f_{12}, f_{22}....f_{n2}.., f_{2m}....f_{nm}\}$

8. Population: $F = \{F_1, F_2.....F_k\}->\{g_1, g_2....g_t\}$ where it is either matrix or DNA string of Fi
9. $while(gen_{count} \leq x)$
10. Epigenetic Masks:

    Create bit string patterns of size n: $\{1, 1..0..1.0\}, \{1, 0..1..1.0\}.....\{0, 0..1..1.0\}$

11. Randomly extract x Genes from chromosome: $Population_{count}\{g_1, g_2....g_x\}$
12. Create a time series of genetic responses
13. Calculating genetic responses on micro and macro level

    a) max = allowed maximum value of a feature; min = allowed minimum value of a feature

    b) Cs = start feature magnitude value of a current time series;

    c) Cf = final feature magnitude value of a current time series

    d) S = total data-points in current time series Classification Strategy;

    e) Ø = aggregate slope of the current time series = (Cf-Cs)/S;

    f) α = Cf/max

14. Select environment specific genetic transcript/mask by silencing highly sensitive genes: if Ø close to 0 and select the gene.
15. $e = \{classify(g), training\}$
16. **end while**
17. $g = \{g_{final}\}$

Here, we propose an algorithmic model which uses inspiration from the theory of epigenetics. Concepts like genetic imprinting are used as a switching mechanism to control data flow. The whole idea operates over metadata of a multimodal data-stream acquired from cameras and sensors. Here, a process starts with data-acquisition from multiples sensing nodes. This data is then reduced to feature sets and organised in a matrix form. Next computing operation applied is to convert the feature-set into a genetic-string like structure. This genetic-string or DNA like feature structure is then assessed for its response towards environmental conditions. Based on this environmental response, features which are having high slope are rejected and those with the inclination close to zero are retained. This selection process applied here is inspired by a genetic imprinting process in epigenetics. Once selection the process is completed then a classification approach [104] is applied by setting selection rules according to feature values and its relation to objects present in the environment.

3.4 Validation and Evaluation

Validating the proposed methodology is very important to assess its performance, scalability, accuracy and precision against benchmark methods. Formal methods for validation exercise can help to estimate and predict error rates. The validation methods also help in building the best fit model for the proposed methodology. Here, we discuss formal validation methods implemented in the contest of the proposed evolutionary multimodal approach. In the proposed validation methods an application specific process is performed by enhancing the formal validation models. Purpose of the validation exercise is to measure error rates in the proposed methodology concerning benchmark methods. A comparative analysis is performed to assess the superiority of the proposed method from various perspectives. Most

relevant assessment parameters are performance, scalability, accuracy, precision and reliability. Formal methods for validation exercise can help to estimate and predict error rates. Here, we discuss formal validation methods [68, 100] implemented in the contest of the proposed evolutionary multimodal approach. In the proposed validation methods an application specific strategy is applied by enhancing the formal validation models.

Following Cross-validation Methods with application-specific enhancements are applied for validating the proposed Evolutionary Multimodal Approach.

3.4.1 Hold-Out Method

This method is most commonly used in the research community where the set of observation samples or dataset is divided into training and testing or holdout sample subsets. The most frequently observed division in case of holdout method is 2/3 portion of data for training and remaining 1/3 for testing. The holdout validation method randomly selects mutually independent datasets for training and testing. Often this technique is criticised for sample biased method. In the case of the evolutionary multimodal approach, this procedure is further enhanced to get a better validation insight into scenarios where datasets are highly scalable.

Here, use of the training data at varying levels from 70% to as a low as 2% is proposed. The rationale behind this is, the reduction in the training datasets allow reduced training time in case of highly scalable datasets. The proposed hold-out model works cyclically to assess a certain percentage of datasets for certain n number of iterations and then reduces the training dataset size by 4% for the next lower level. This cyclic activity keeps happening till the lowest threshold level is not reached. For every validation sub-cycle error rates and time required is measured.

Finally, an average value with deviation in these results in positive and negative directions is stored. This validation method is particularly useful to validate accuracy, performance and scalability aspect of the proposed Evolutionary Multimodal Method.

Another essential validation strategy applied using Hold-Out method has a test subset validated with every sample separately. For every sub-cycle, the error and processing time are accumulated to find respective values for the entire cycle. This variation of the hold-out method is particularly useful for assessing the performance and accuracy of the method to be validated. Thus, the Hold-Out method with application-specific modifications provides accuracy checking mechanism with scalability and performance as well is validated.

3.4.2 K-Fold Validation

K-Fold Validation: In case of K-Fold validation method, the dataset is divided into K sets or folds of subsamples with an equal number of data-points or observations. Starting with one among those K subsets is treated as a validation or test set and remaining as training sets. Every training set is used K times individually to keep training and then to test the computation model under validation. Every validation cycle generates a prediction error which is averaged by dividing their sum by K. Normal practice is to use K=5 to 10.

$$P_{Error} = Av_{Error} = \sum_{i=1}^{k} error_i \tag{23}$$

In this case, the K-Fold approach is used by the varying value of K from 2 to a maximum count below or equal to 10 from where results start saturating. In the case of the proposed methodology, K-Fold validation can be applied to test precision

and stability.

3.5 Summary

This chapter outlined adopted and designed methodologies for validating the proposed hypothesis. The methodology provided mathematical and logical details of every proposed algorithm. Evolutionary biological concepts were discussed to explain the origin of the inspiration behind these algorithms. The further action is to develop a validation framework to test the proposed set of algorithms. The following chapter discusses the tools and techniques used during the experimental work.

Part II

Research Contributions and Validation

4 modelling and Design of Experimental Framework

This chapter outlines design for the experimental framework required to validate the proposed methodology. Since the intended methods are applied to a variety of operating environments, the experimental setup and datasets as well need to be developed and modified accordingly. Environmental dynamism can deteriorate the multimodal data and render selected modality ineffective. Thus an adaptive modality selection mechanism is required to overcome this issue.

An experimental setup is developed to validate primarily multimodal and evolutionary algorithms operating in diverse conditions. The multimodal characteristic of datasets is implemented using a parallel image acquisition design. This design is realised using a set of hardware and software tools. The datasets are modified using domain-specific modelling techniques to emulate degradation due to the changes in surroundings. This modelling process creates a vast range of expansion over every instance in the dataset. Thus, the design of the proposed framework principally focuses on a concurrent data-acquisition. Whereas, data modelling is performed to produce a broad variety of data samples to validate the adaptivity of the proposed algorithms.

The following subsection starts with explaining the purpose, overview and scope of the setup. Then it discusses various considerations and rationale for this framework. The next subsection presents the system architecture with high-level software design. Design includes subsystems required for different data processing activities. The human-machine interfaces used during multiple experimental operations are discussed to provide details on the usability of the system. Finally, the data modelling process with all related computational building blocks is presented.

4.1 Purpose, Scope and Overview

Purpose of the experimental platform discussed is to provide an environment to validate the hypothesis explained in the introduction chapter of this report. The proposed hypothesis has the following keywords which are taken into consideration when this platform is developed:

- Environmental Dynamism,
- Multimodality, and
- Evolution

Environmental Dynamism may have a two-fold effect on a monitoring system, first is on the acquired datasets and second on the computing model processing it. Due to the changes in illumination, reflectivity and so forth, the acquired datasets start receiving noisy and occasionally even corrupted datasets. Whereas, the noise impacts over a computing model if it is not trained at sufficient level to adapt itself to the variations.

In case of the multimodal datasets, a parallel stream of data with multiple channels is processed by its detection algorithm. This concurrency needs to be carefully managed using techniques to acquire and store dataset instances in an organised manner. A design to accommodate these concerns is required.

Evolution is a phenomenon commonly used in many domains including biology to reflect an ability to adapt to changes in environments. In the case of biological systems, nature has gifted evolution in many forms to overcome or survive through gradual and sudden environmental changes. As an example, Bet Hedging is a concept allows species to overcome sudden changes by increasing it breading quality

with a lowered fitness of newborns. Whereas, phenotypic plasticity enables species to survive through new environments by varying their physical properties within a specific range. In this case, we use our experimental setup which allows testing evolutionary computing concepts we discussed in the methodology section during the last chapter. A feature level processing inside an algorithm is applied to reduce the scale and then various feature optimisation and classification techniques are employed to find an accurate solution.

The scope of this design for the experimental platform is mainly limited to provide a testbed for multimodal imaging applications. The proposed setup restricts itself to work with spatial domain data and intends to expand in future to deal with spectral data to make it a hybrid data processing system. The proposed method also intends to work using various image fusion, feature extraction techniques and machine learning algorithms to process multimodal imaging datasets. Currently, it is limited to work on registered or unregistered images and not offering to work with any other datatypes.

4.2 Requirements and Considerations

Various non-functional requirements were considered for developing the experimental platform as below:

- **Concurrency:**

  Capable of acquiring real-time data inputs from multiple modalities which include video-cameras and image repositories. A possible future requirement to include wireless sensing nodes as well needs to be considered.

Performance:

Ability to acquire images from a maximum frame-rate of 30 frames/second to a minimum value as per the experimental requirement.

- **Dimensionality Reduction:**

  Capable of converting raw data into dimensionally reduced feature form using various techniques like point-features, textures, colour-descriptors, etc.

- **Storage Capabilities:**

  Capability to store the acquired data, final or intermediate level results on a local or a remote computer. It means the proposed software should be able to connect with RDBMS databases like SQL-Server or spreadsheets like Microsoft Excel. This feature is essential to achieve data in various forms at various stages and retrieve in the future as required.

- **User Interface and Presentation:**

  Right from data acquisition to final results, data needs to be presented in a form to make it readable to future readers when it is published. An effective presentation of findings is vital to convey the message of the experimental activity to a research community. The proposed system has presentation modules to view data in raw, tabular and plotted forms.

- **Reusability:**

  Since many algorithms tend to use standard components, the software coded in this platform using procedural blocks shall make them reusable. The reusability principle allows a reduction in algorithm design and development time to make both processes faster.

- **Modularity:**

  Being programmed in smaller code blocks the proposed architecture shall allow maintaining modularity. It enables to test and apply individual components relatively faster and more straightforward. A combination of little modules supports developing of a variety of scenarios and case studies in a relatively quick manner.

- **Hardware Interfacing:**

  Proposed experimental setup shall be capable of interfacing with external cameras and sensing modalities to acquire data from real-life environments. Purpose of adding this capability is to create a way for developing applications based on this software platform.

- **Portability:**

  Proposed system shall be capable of being implemented in multiple software libraries including procedural and object-oriented programming paradigms. Due to recent research in the analytics and big-data applications, there are many software libraries and tools readily available. Having a portable design would make very easy to apply the same solution in multiple domains and then compare for performance, reliability, scalability and so forth.

- **Scalability:**

  Few of the experiments may deal with highly scalable datasets and this fact will have to be considered here to make the system capable of handling it. What it means is, the system shall be designed in a way to operate large volumes of data during the experimental activities. Another aspect of the scalability is the ability to acquire information from multiple channels. This framework has a capability of up to obtain data from four image acquisitions

and the same number of wireless sensor acquisition channels.

- **Expandability:**

  The proposed architecture can expand itself from the current number of image acquisition channels to higher numbers as per the future needs. It can also allow accommodating other variety of modalities to provide a versatile system.

Summary of the functional aspects considered for the proposed architecture are as described below:

- The input data in raw and feature form will be presented on a graphical user interface. This feature of the system will allow visualising of acquired information at various stages.
- Data acquisition and experimental processes shall be started, paused and stopped using control features on a user interface. Many time lasting and lengthy experimental activities shall have an ability to start, pause and stop itself as required.
- Detected objects shall be marked using a bounding box with distinct coloured edges to make them visible.
- Numerical information acquired and calculated during the experimental activities will be presented using tables and plots. Quantitative and qualitative results of the validation process shall be made available in a conventional scientific form.
- Ability to control the number of active input channels by using user-interface. Depending on the experimental requirements and datasets used the number of

input data channels needs to be switched on and off. Provision of a user-control to allow this switching shall make the experimental process more flexible. The same setup shall be used for different types of experimental activities.

4.3 HyMuDS System Design

A multimodal detection system called 'Hybrid Multimodal Data Acquisition System' [61] or in short form HyMuDS is proposed. The system has an architecture, which to addresses concerns associated with concurrency, and the discreteness of the data acquired from multiple sources. The problem of concurrency is resolved by a multi-threaded approach to poll a data coming from different nodes interfaced to hardware ports. Details of the software and hardware design are discussed in the following sections.

4.3.1 Hardware Architecture

Hardware architecture includes a scheme to address practical challenges relating to the deployment of the sensing modalities to acquire accurate information of a scenario. Major hardware concerns related to the HyMuDS system are:

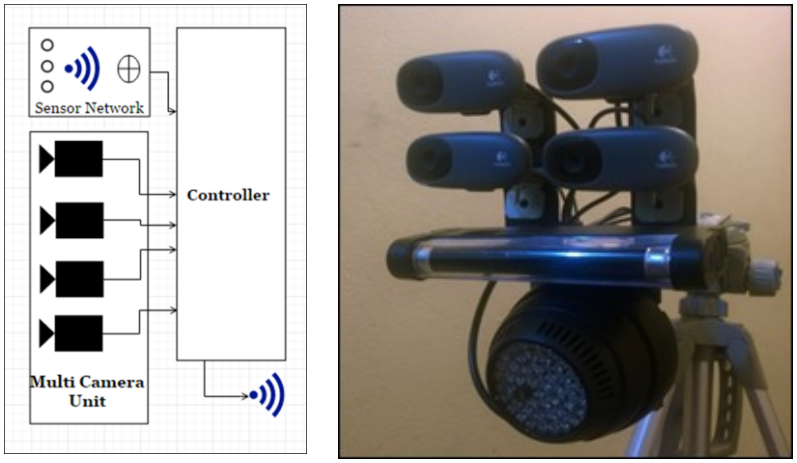
- Bigger bandwidth is required for the imaging modalities to increase detection capability in various environmental conditions,
- The system needs significant flexibility in the deployment of microsensors in order to cover larger distances and distributed coverage within the waterway under observation,

- Portable and maintainable setup is required to detect signals in various scenarios and geographical locations,
- Camera modalities need a flexible mounting mechanism to cover a detection at various angles and heights, and
- Camera modalities need a flexible mounting mechanism to cover the video detection process at various angles and heights.

As shown in Figure 16, the hardware design is implemented. It allows addressing all concerns discussed previously and provides a solution to achieve architectural objectives. Four cameras with a serial interface are mounted to acquire imagery in parallel. Two of these cameras are equipped with ultraviolet and infrared pass lenses to allow the specific spectrum to be captured. Whereas, the other two cameras are equipped with filter lenses to disallow IR and UV lights to pass through to camera sensors. A provision needed to be made for a wireless sensor network to acquire analogue parameters related to macro-level properties of the water. In this case, a gateway connected with a USB interface interacts with the nodes deployed in the area under observation.

4.3.2 Software Framework

Based on the background study and problem specific concerns, the software framework is proposed. The proposed structure of the implementation uses an object-oriented programming language for modularity. The architecture is divided into several layers focusing various concerns within the problem-space.



(a) Schematic Diagram (b) Implementation

Figure 16: HyMuDS Data Acquisition Hardware

- **Data Acquisition:**

  The topmost layer represents data acquisition layer to acquire real-time environmental details. Here, multimodal datasets are obtained using a diverse set of modalities. The connectivity of these cameras can be established using various hardware interfaces. A provision is made in this layer of the framework for WiFi, Bluetooth, USB and Serial communication protocols. The WiFi and Bluetooth connectivity are provisioned for connecting to wireless camera sensors. This layer has a multithreaded programming implementation to handle concurrency.

- **Presentation:**

  To display data in various stages using a variety of presentation modules. Raw data is presented in a pictorial form using video-players whereas a provision is made to show features in a tabular form. Other commonly used controls are utilised for controlling the system operations. Plotting user-controls are applied to present intermediate or final results of test cycles.

- **Pre-Processing and Data Conditioning:**

  The acquired data needs to be preprocessed through data conditioning algorithms to perform noise-elimination, normalisation and dimensionality-reduction. In the proposed research context, we use feature extraction techniques to reduce a raw multimodal dataset at a lower dimension.

- **Data optimisation, Fusion and Supervised Learning:**

  Data optimisation and supervised learning: A dataset in its original or a derived form needs to be optimised further by applying a supervised machine learning technique. This layer is the most crucial set of computational elements for validating the proposed methodology. Within this class of algorithms, evolutionary selection and classification of features and modalities are performed. The data optimisation techniques include evolutionary algorithms to implement the proposed hypothesis. Whereas, the classification algorithms include a set of commonly practised algorithms like Artificial Neural Networks, K Nearest Neighbors, Support Vector Machines, etc. The data fusion algorithms including wavelet, Principal Component Analysis and primitive techniques (min, max, etc.) are implemented. These methods are very commonly used in the multimodal data processing applications.

- **Repository or Data Layer:**

  The data layer is a repository for saving data and results in spatial, numerical and address forms. A well-organised structure with separate compartments for types and instances of the datasets is developed.

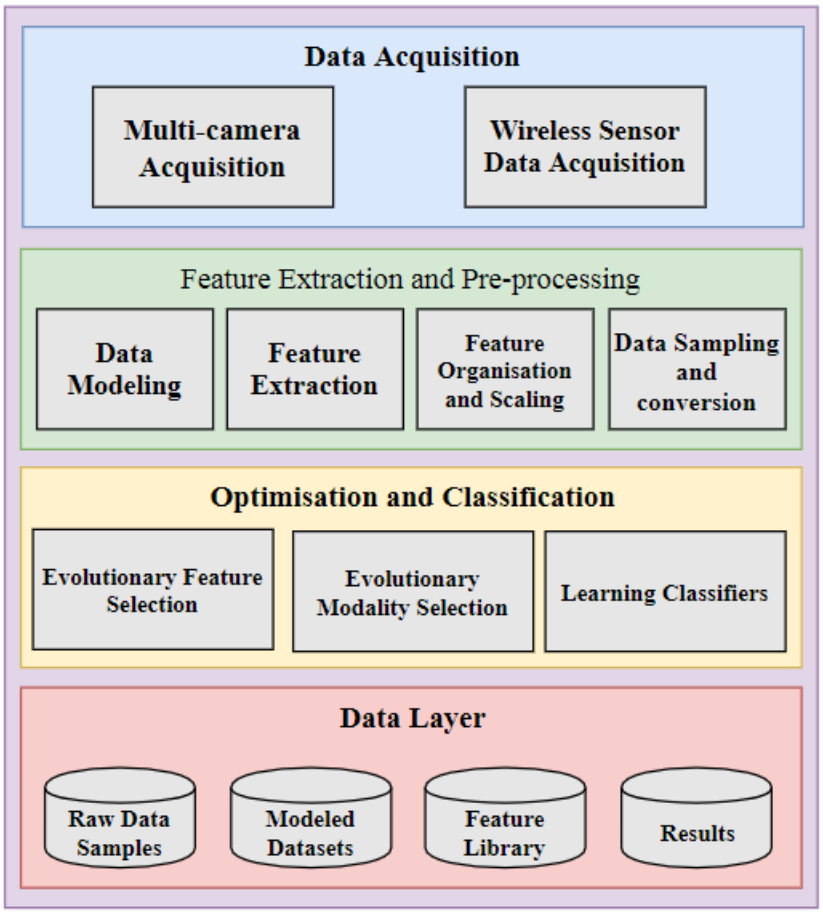


Figure 17: Software Framework for Proposed Set of Algorithms

4.3.3 Component Level Design

This subsection explains a component level design of most essential modules of the software architecture.

Data Acquisition

The data modelling process (Alg. 5) describes how data acquisition and preparation is done for the experimental activities. Here, a dataset or acquired data is accessed and then the even modal set is assigned to a separate variable. After that,

Algorithm 5 Data Modelling Process

begin

1. $DS = Read(Dataset)$ where $DS = \{I1, I2..I4\}$
2. Calculate size of a modality: $[r, \sim] = size(I1)$
3. k =number of noise levels
4. **for all** $i \leq 66$
5. **for all** $j \leq r$
6. $M_{i,j} = imnoise_i(I_j)$
7. **end for**
8. **end for**
9. Save modelled dataset: $M = \{M_{1,1}, M_{2,1}..M_{i,j}\}$

the size of a significant modality is measured to define a number of iterations required to perform for modifying every element in the set. Based on the scientific techniques every modal element is expanded into multiple variations to simulate different environmental conditions. Here, 'imnoise' functions are typically representing environmental conditions emulated by adding different types and levels of image noises. Finally, a derived dataset M with modal representations$\{M_{1,1}, M_{2,1}..M_{i,j}\}$ is saved which is used an input data for other software components. Especially, Data Fusion and Feature Extraction modules will be extensively using M as their input.

Algorithm 6 Image Fusion Algorithm

begin

1. Input Multimodal Dataset: $M = \{M1, M2, M3, M4\}$
2. Calculate Size of a Modality:$[rows, cols] = size(M1)$
3. Declare empty arrays for fused images: $fusedmin = []$, $fusedmax = []$, $fusedavg = []$, $fusedwave = []$
4. **for all** 1 :$i \leq rows$

   $M_{i,j} = imcrop(M_{i,j}, [r, c])$

   $fusedmin_i = minpixel(M_{1,i}, M_{2i}, M_{3i}, M_{4i})$

   $fusedmax_i = maxpixel(M_{1,i}, M_{2i}, M_{3i}, M_{4i})$

   $fusedavg_i = averagepixel(M_{1,i}, M_{2i}, M_{3i}, M_{4i})$

   $wvfusion_i = waveletfusion(M_{1,i}, M_{2i}, M_{3i}, M_{4i})$

   $pcafusion_i = pcafusion(M_{1,i}, M_{2i}, M_{3i}, M_{4i})$

   $fusedmin = [fusedmin; fusedmin_i]$

   $fusedmax = [fusedmax; fusedmax_i]$

   $fusedavg = [fusedavg; fusedavg_i]$

   $fusedwave = [fusedwave; wvfusion_i]$

   $fusedpca = [fusedpca; pcafusion_i]$

5. **end for**

Data Fusion Module

The data fusion module has the programming routine as shown in Algorithm 2. It starts with measuring sizes of every modality and then uses those measurements

inside a for loop to restrict sizes of all images and make them uniformly sized for fusion operations. Here, four different fusion algorithms operate within the for loop and fuse all modalities together to create combined images for future operations.

Algorithm 7 Feature Extraction and Storage Algorithm

begin

1. Input Multimodal Dataset: $M = \{M_1...M_n\}$
2. Count Images in a modality: $[rows, \sim] = size(M1)$
3. Declare feature matrices: $[feature_{m_1}....feature_{m_n}]$
4. Calculate size of every image: $[r_1, c_1] = size(M_1)...[r_n, c_n] = size(M_n)$
5. Define container image size: $r = min(r_1...r_n)$, $c = min(c_1....c_n)$
6. **for** $i \leq rows$
7. **for** $j \leq n$
8. $M1i = imcrop(M_{ji}, [r, c])$
9. $fM_{ji} = featureextract(M_{ji})$
10. $feature_{m_j} = [feature_{m_j}; fM_{ji}]$
11. **end for**
12. **end for**

Data optimisation

The feature optimisation layer in the proposed architecture includes set of algorithms proposed in the methodology section of this report. The optimisation process

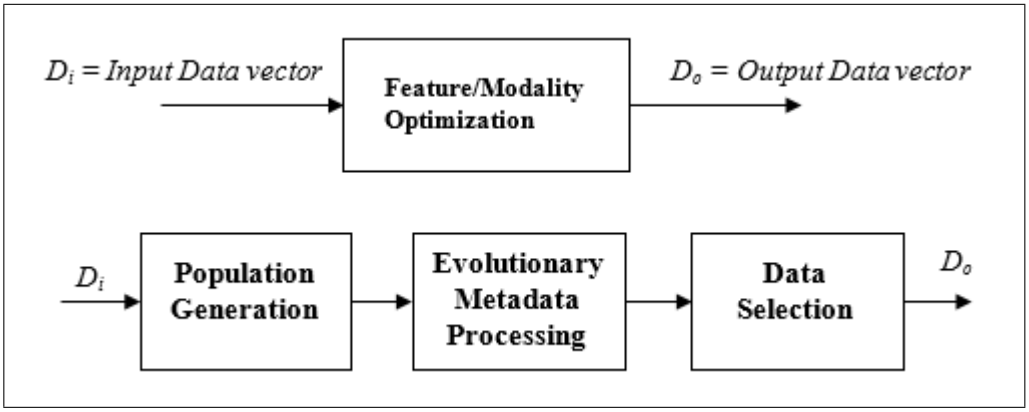


Figure 18: Generalization of optimisation Process

is applied on an input data in feature and modality form. For every component in this layer, the type of input and output data types are the same. The optimisation process in the proposed methodology is a population-based search and optimisation approach. In this method, metadata of the input data vectors is iteratively processed to find an optimum solution. Three main building blocks of this method include population generation, metadata processing and data selection. In case of feature level optimisation, a classification algorithm is integrated to calculate intermittent fitness values during metadata processing cycles.

Feature Classification Functional blocks of the most commonly practised classification algorithm for multimodal object detection are implemented. In this research study, this set of algorithms are essential since objects detected are labelled using classification algorithms. The proposed evolutionary set of algorithms either can be used along with these classifiers to find their effect on the classification process. As described in figure 19, a set of features in a vector form is supplied to receive singular or multiple classification results.

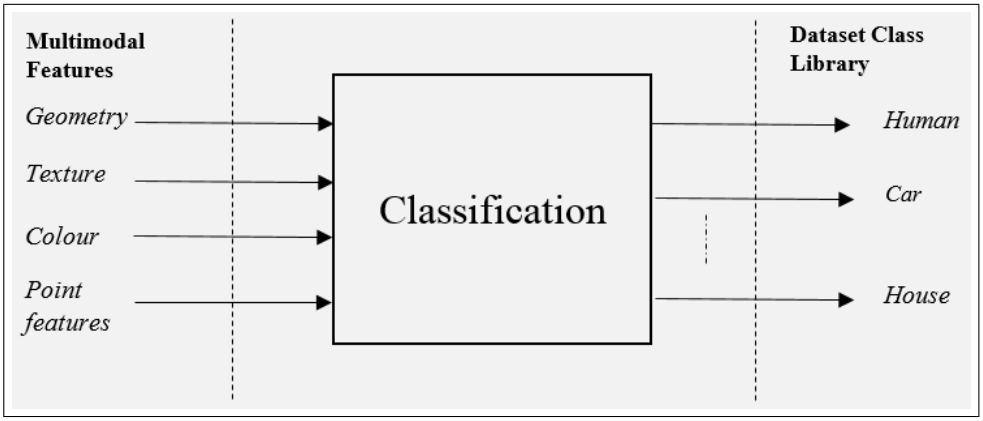


Figure 19: Classification Block Diagram

4.3.4 Human Interface Design

A Human-Machine Interface or HMI for object detection process was developed as shown in diagram 6. The HMI was designed by following the modularity principle to accommodate various experimental requirements. Data visualization for four imaging modalities is provided with the ability to select one image sequence at a time.

A provision to vary the settings and parameters of the algorithms tested is provided using text boxes. The validation process can be controlled using radio and push buttons. Analytical results are visualised using line graphs. Thus, a modular user interface is created to conduct the experimental analysis.

4.4 Data modelling Process

Since proposed research study requires investigating effects of environmental fluctuation in gradual and sudden ways, it is essential to design datasets used for validation to accommodate those variations. In this case, we consider three classes of possible variations with abrupt and gradual changes in them:

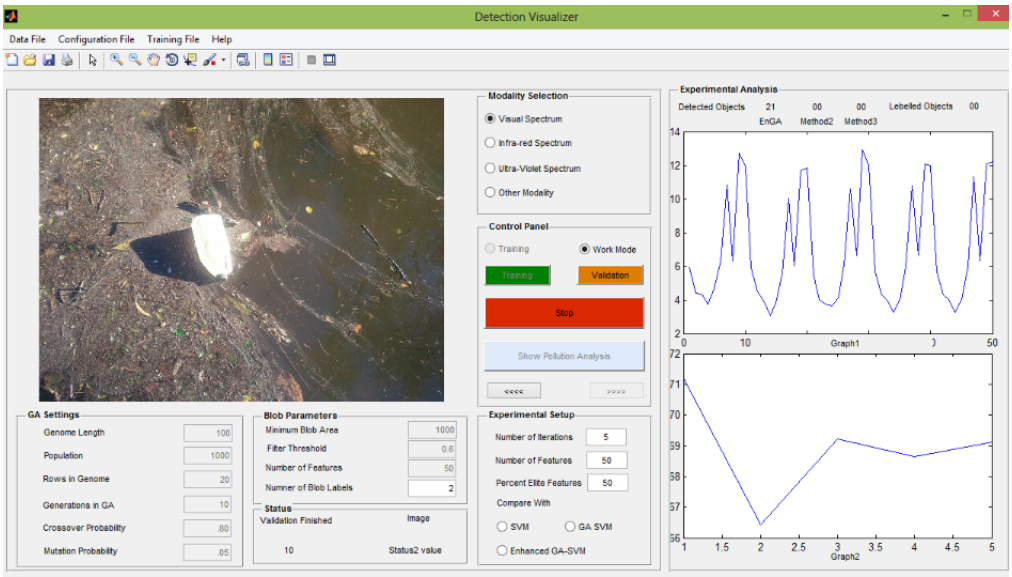


Figure 20: Object Detection and Analysis HMI

- **Illumination:**

  This is usually observed as a gradual phenomenon on a typical, bright sunny day. It becomes rapidly degrading for visual spectrum when twilight or other causes stopping the sunlight reaching to the object under observation.

- **Fog:**

  It is commonly observed during winter that fog may create an occlusion of the object under observation and this can be a gradual or sudden addition to the environment.

- **Rain:**

  Rainy conditions may impose limited light with some sort of noise due to raindrops which will make the observed scene hazy. It can also add micro-occlusion to hide important features.

  Considering all the above possible scenarios, either multimodal datasets already having those environmental adversities embedded in them or simulation

to create such datasets is required. Following paragraphs contemplate on both of these possible options to develop a strategy to build datasets for either representing or simulating gradual and sudden environmental changes.

Another aspect relating to variability in datasets considered for proposed experiments is a number of classes in datasets. Since object detection can be a subjective study concerning its area of application, the number of object classes within may vary. Here, we conduct experiments to test the proposed algorithms from binary to multiple classification models. Such testing will validate merits of these algorithms to work on a variety of applications. Thus, we require datasets to accommodate this concern as well.

Validation of the performance of the proposed methodology, it will be required to use datasets with a certain level of scalability. Having a minimally sized dataset may not be enough to measure time consumed by a particular algorithm to detect objects and anomalies within. A dataset with too high scalability may prove a time-consuming exercise to run the proposed experiments. Here, we try to accommodate both these concerns by using a dataset with moderate size added with variations to simulate environmental dynamism. Due to such a strategy, both scalability and variability concerns will be addressed. Following subsections provide detailed information on the datasets used.

4.4.1 Proposed Datasets

According to the experimental requirements, various datasets were investigated to test multimodality along with environmental variability features. In general, the datasets are available either purely of unimodal visual objects with illumination

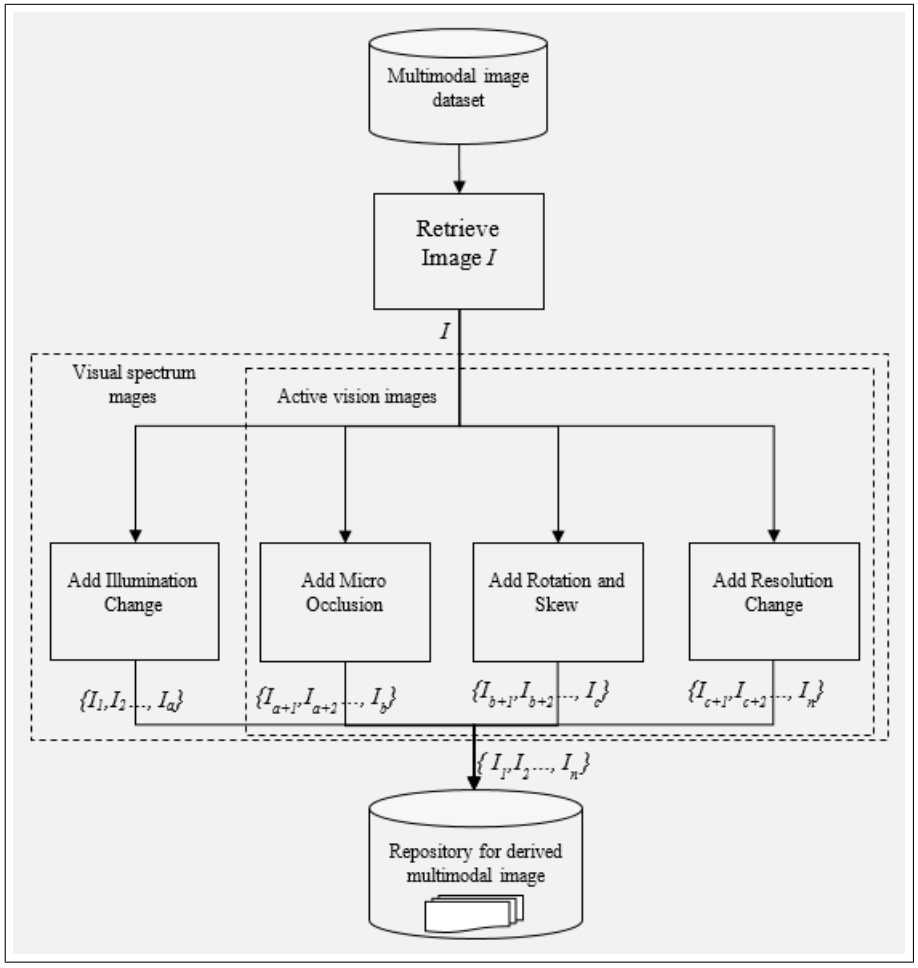


Figure 21: Data modelling Process

variations or multimodal datasets captured in stable environmental scenarios. Due to these limitations, we chose to expand a base dataset added with extra noise for modelling fog, micro-occlusion and lighting variations. This modelling process provides a simulation of a broader environmental range to validate uncertainty and variability in environmental conditions. The subsection 4.4.2 provides a detailed description of both modelling process and dataset.

4.4.2 Data modelling

The data modelling is performed to simulate environmental variability and its effects on the multimodal imagery. Here, we consider four commonly observed parameters to model the environmental dynamism. Following paragraphs explain every parameter to a full extent.

Variability in Lighting Conditions

Estimating illumination changes in stochastic conditions is a very challenging process. Adversities added to visual spectrum images due to the disparate illumination adds various noises to reduce the visibility of objects present in it. Main effects due to noise include visibility dilution and loss of sharpness in colour reflectance. Gaussian noise is a most commonly observed noise due to poor illuminating conditions. Since computer vision applications mostly operate in really distinct environments, it is challenging having a common strategy to model illumination noise. Here, based on inspiration few research studies [93, 73] we propose using addition of the Gaussian noise [53] on visual spectrum images. The noise addition is done systematically and randomly to simulate gradual and sudden changes into the imaging platform environment.

Micro Level Occlusion

Adding micro level occlusion provides a close equivalence to the simulation of rain. In the current investigation, we give a limited attempt to create a scenario resembling to rain by adding microparticles using Salt and Pepper noise [9]. The density of this type of noise as well varied systematically and randomly. This type of the noise is added over all modalities experimented in a validation activity.

Adding Rotation and Skew

All modalities are added with rotation and skew to create a scenario where objects are observed either at an axial and planar angle. Rotating of images [37] allow

looking at the same dataset with another environmental scene when objects are seen from an oblique perspective.

Simulating Scale Variation

Modelling of the scale variance [103] is attempted by varying image resolutions to dilute pixel densities at various levels. By applying this technique objects present in those images lose their resolution to create a sense of increased distance concerning the capturing device or a camera. This type of variance is applied to all modalities since it is a typical behaviour across the spectrum for all camera.

TNO Image Multiband Datasets

Table 1 describes the TNO datasets [119] used for the final experimental analysis. The entire dataset was divided into three smaller sets based on a number of modalities. Every modality added with the variations as discussed in the previous paragraphs of this subsection. An example showing the effect of variability added to one image is displayed in Figure 22.

Table 1: Dataset Description

ID	**Number of Modalities**	**Number of Images**	**Variations Added**	**Synthetic Images**	**Image Sizes**	**Number of Classes**
DS1	Two	21	66	1386	64KP ∼ 0.25MP	14
DS2	Three	22	66	1452	64KP ∼ 0.25MP	12
DS3	Four	17	66	1122	64KP ∼ 0.25MP	12

Figure 22: Typical Example of Modeled Dataset for an Image

4.5 Summary

Thus, this chapter outlines a systematic approach towards various software and hardware tools used in the validation process of the proposed methodology. A hardware setup capable of capturing multimodal images from the range of environmental conditions is used in the experimental work and case studies followed in this report. Whereas, the data modelling process used is used to simulate various environmental scenarios by modifying a multimodal imaging dataset. Following chapters will present the applicability of these tools for validating the hypothesis.

5 Validation and Evaluation of Evolutionary Multimodal Object Detection Methods

5.1 Background

This chapter presents the experimental results of the proposed hypothesis during the chapter one. Here, the emphasis is given to the following keywords while validating the hypothesis:

- **Multimodality:**
  Every experimental cycle is performed using multimodal imaging datasets to validate its effect in dynamic environmental scenarios when evolutionary optimisation techniques are used.

- **Accuracy and Performance:**
  An accurate prediction of datasets is assessed in various abstraction levels of the dataset which include modality, feature and inference. Another aspect of accuracy check consists of the presence of number objects or classes in datasets experimented. Finally, the number of features extracted are varied from minimum level to maximum to see their effect on the object detection accuracy.

- **Robustness:**
  Precision level of detection is assessed in various scenarios as discussed in accuracy check section. Here, importance is given to the variability of detection accuracy from a mean value. This experimental aspect measures fluctuation with respect o this mean value.

Following subsection provide a detailed explanation of experimental results on validating accuracy, performance and robustness.

5.2 Experimental Results

This subsection presents experimental results for every algorithm in the proposed methodology. Every algorithm was developed to overcome a different kind of improvement in the detection process. The Enhanced GA was designed to improve accuracy using a pre-learned feature-template. The phenotypic switching method mainly is used for optimising modality level data. The bet-hedging and epigenetic approach use feature level details for selection to improve robustness and accuracy. This subsection discusses the experimental results for every algorithm and compares them with conventional methods.

5.2.1 Enhanced Genetic Algorithm

The enhanced-GA operates on the multimodal features to extract most relevant features to the range of environments. The experimental work mainly assesses the algorithm from accuracy, robustness and performance perspectives. Every assessment parameter was further tested by varying the input dataset at modality, feature and object quantities. Another essential validation strategy includes assessment of fused and unfused modalities. The optimisation algorithm was run for 100 generations to get a stable output. The following data in Figure 24 shows an example plot with the number of generations vs. fitness value.

Accuracy

The proposed GA version was assessed for fused and unfused images. The classi-

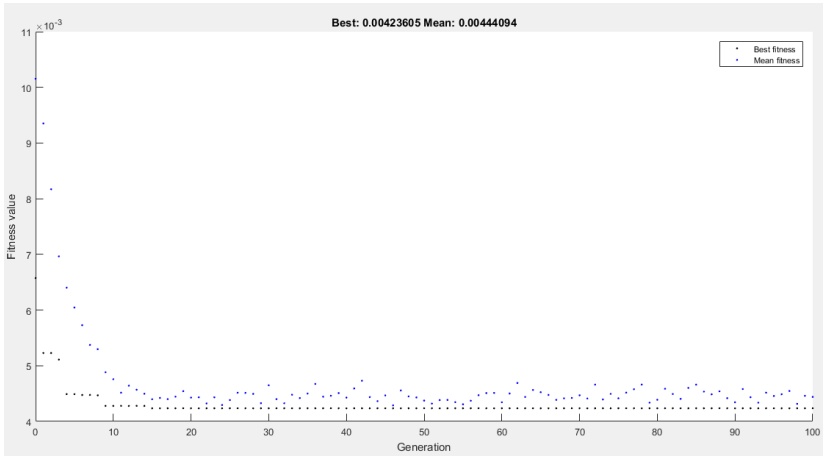


Figure 23: Genetic Algorithm Generation Count Vs. Fitness

fication process was performed using an SVM classifier with a linear kernel. The following data in Table 2 shows the results for the test cycles ran to validate the accuracy and performance of the proposed algorithms.

Class Label

In this experimental cycle, the number of class labels were varied to assess the accuracy of the algorithms tested. The results are summarised in the following data in Table 2. The classification accuracy for unoptimised featureset was compared against the proposed method. The maximum number of class labels is equal to total types of image-sets available in a subset of the dataset. The accuracy column shows the range of the accuracy level observed for every subtype of the assessment.

Feature Count

Another scenario assumed in the validation process included a varying number of features for every type of dataset. It was considered the number of features may

Table 2: Accuracy Comparison of Feature Classification Methods when Class Labels Quantity is Varied

Dataset	**Number of Class Labels**	**Unoptimized Feature Classification Accuracy**	**Optimized Feature Classification Accuracy**
DS1 (Bimodal)	2	43 $\sim$ 49 %	44 $\sim$ 47 %
	4	44 $\sim$ 47 %	44 $\sim$ 49 %
	8	40 $\sim$ 44 %	42 $\sim$ 48 %
	10	32 $\sim$ 38 %	36 $\sim$ 43 %
	14	18 $\sim$ 29 %	28 $\sim$ 41 %
DS2 (Trimodal)	2	67 $\sim$ 73%	69 $\sim$ 76 %
	4	64 $\sim$ 69%	63 $\sim$ 72 %
	8	61 $\sim$ 66%	62 $\sim$ 71%
	12	51 $\sim$ 59%	61 $\sim$ 67%
DS3 (QuadModal)	2	63 $\sim$ 67%	66 $\sim$ 71%
	4	57 $\sim$ 61%	60 $\sim$ 71 %
	8	48 $\sim$ 53 %	56 $\sim$ 62 %
	12	45 $\sim$ 51 %	55 $\sim$ 59 %

affect performance and accuracy for every method. The assumption was based on a consideration that if a low percentage of object details without any optimisation can compromise then the accuracy and improve the performance.

Table 3: Accuracy Comparison of Unoptimized Feature Classification Methods when Feature Quantity Varies

Dataset	**Number of Features**	**Unoptimized Feature Classification**	**Optimized Feature Classification**
DS1(Bimodal)	$10 \sim 56$	$18 \sim 29$ %	$28 \sim 41$%
DS2 (Trimodal)	$30 \sim 78$	$23 \sim 32$ %	$31 \sim 43$%
DS3 (Quad modal)	$40 \sim 102$	$24 \sim 33$ %	$32 \sim 43$%

Fusion Methods

Three most commonly practices fusion methods were tested for the accuracy when the fused images were applied with and without a feature subset optimisation method. These test cycles were applied over complete datasets with the maximum number of class labels.

Table 4: Accuracy Comparison when Feature optimisation is Applied to various Fusion Methods

Fusion Method	**Dataset**	**Unoptimized**	**Optimized**
DWT	DS1	28 $\sim$ 40 %	31 $\sim$ 42 %
	DS2	48 $\sim$ 58 %	53 $\sim$ 65 %
	DS3	49 $\sim$55 %	52 $\sim$ 62 %
Mamdani (Min)	DS1	24 $\sim$32 %	29 $\sim$ 35 %
	DS2	43 $\sim$ 49 %	48 $\sim$56 %
	DS3	45 $\sim$51 %	52 $\sim$ 59%
Momdani (Max)	DS1	21 $\sim$ 33 %	27 $\sim$ 32 %
	DS2	40 $\sim$ 46 %	44 $\sim$ 51 %
	DS3	42 $\sim$ 49 %	49 $\sim$ 56%

5.2.2 Classification Methods

Mean accuracy for every method which was applied for fused, unfused and optimised feature sets. Seven most commonly practised classification algorithms were tested using the proposed method. The list of the classifiers with their acronyms is as below:

- SVM: Support Vector Machines,
- DT: Decision Trees,
- NB: Naive Bayes,
- RF: Random Forests,
- DC: Discriminant Classifiers,

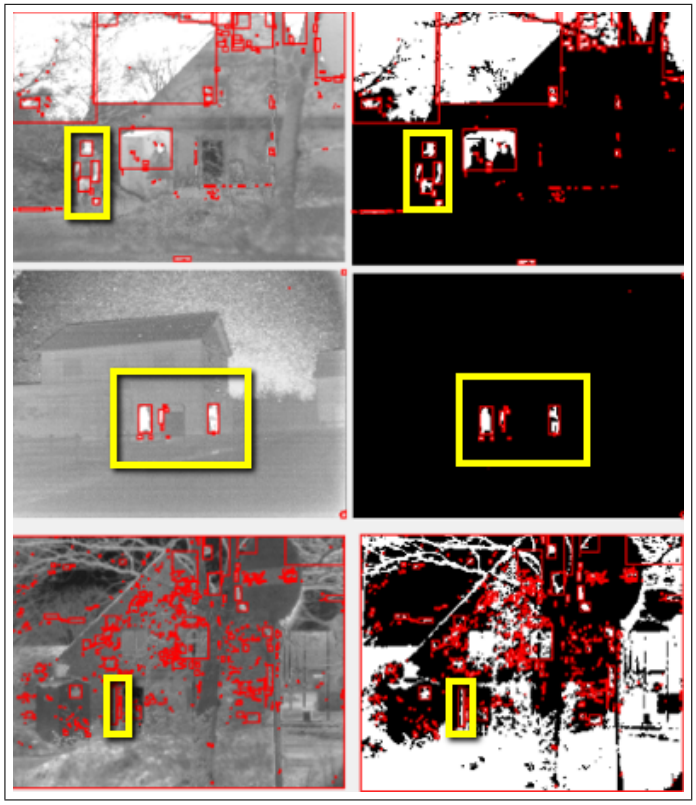


Figure 24: RoIs for Detecting Human Presence near House

- KNN: K Nearest Neighbors, and
- ANN: Artificial Neural Networks

Following the data in Table 5 summarises the accuracy comparison between all classification algorithms with and without fusion or optimisation. The number of features per unoptimised modality is equal to 26. In case of the optimised features, the feature count may vary and the results described here are the most commonly observed values.

Table 5: Accuracy Comparison of Feature Classification Methods

Modality Count	**Number of Outputs**	**Feature Type**	**SVM**	**DT**	**NB**	**RF**
Two	14	Unfused	22.28%	18.58%	15.79%	21.51%
		Fused	33.15%	30.46%	55.01%	32.05%
		Optimized	35.27%	31.58%	59.53%	35.63%
Three	12	Unfused	58.57%	55.42%	64.80%	59.54%
		Fused	59.17%	61.54%	70.95%	62.73%
		Optimized	62.71%	62.41%	72.41%	64.34%
Four	12	Unfused	49.45%	59.01%	61.70%	59.71%
		Fused	52.65%	63.84%	64.92%	66.04%
		Optimized	55.47%	65.88%	66.42%	68.37%

Modality Count	**Number of Outputs**	**Feature Type**	**DC**	**KNN**	**ANN**
Two	14	Unfused	22.26%	21.93%	57.01%
		Fused	35.81%	32.01%	66.27%
		Optimized	39.62%	40.12%	82.83%
Three	12	Unfused	57.38%	58.37%	74.30%
		Fused	61.54%	59.86%	78.19%
		Optimized	63.66%	60.55%	89.45%
Four	12	Unfused	60.77%	61.61%	85.34%
		Fused	65.81%	65.67%	87.07%
		Optimized	68.43%	66.08%	91.97%

Performance

The performance comparison for every dataset with the maximum number of classes was tested. The results are shown in Table 6. In this validation, all classification algorithms listed in Table 5 were tested. The following results are for the SVM classifier since feature accuracy results in Table 2, 3 and 4 were achieved using the same.

Table 6: Performance Comparison of Feature Classification Methods

Dataset	**Number of Class Labels**	**Unoptimized Feature Classification**	**Optimized Feature Classification**
DS1	14	$0.0392 \sim 0.618$ s	$0.0346 \sim 0.0521$ s
DS2	12	$0.0406 \sim 0.3847$ s	$0.0378 \sim 0.0432$ s
DS3	12	$0.0406 \sim 0.3847$ s	$0.0378 \sim 0.0432$ s

Robustness

The robustness or stability of the proposed method is compared with the existing feature classification algorithms. The testing was conducted for 100 cycles for every algorithm to vary the fluctuations in their accuracy levels. The following data in Table 7 shows a comparison between the percentage range of the accuracy fluctuation for SVM classified feature sets belonging to the image sets which are fused or unfused. Thus, the accuracy, performance and robustness were tested and compared with the traditional methods. The results are indicative of significant and marginal improvements which will be discussed in the conclusion section of this chapter.

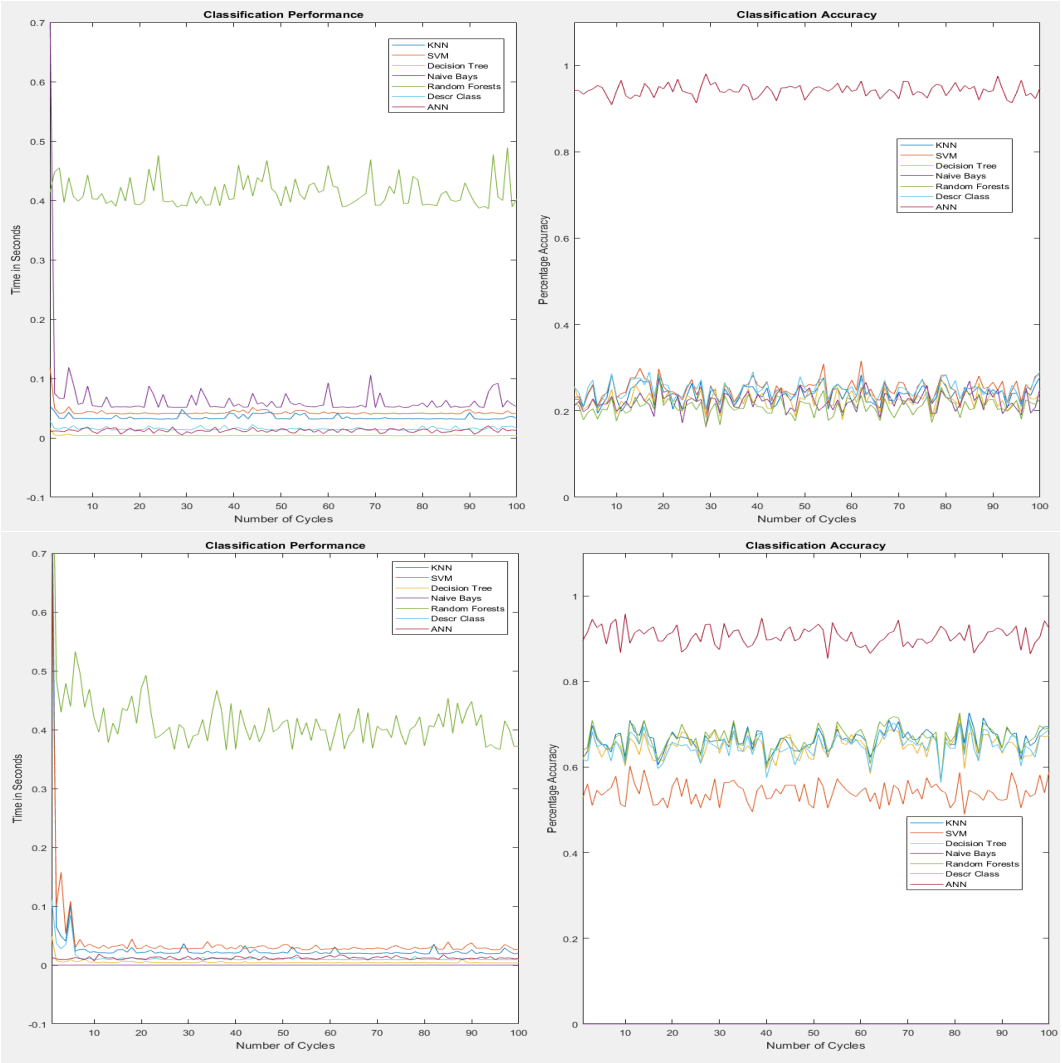


Figure 25: Performance and Accuracy Comparison for Classification Methods (Part 1/2)

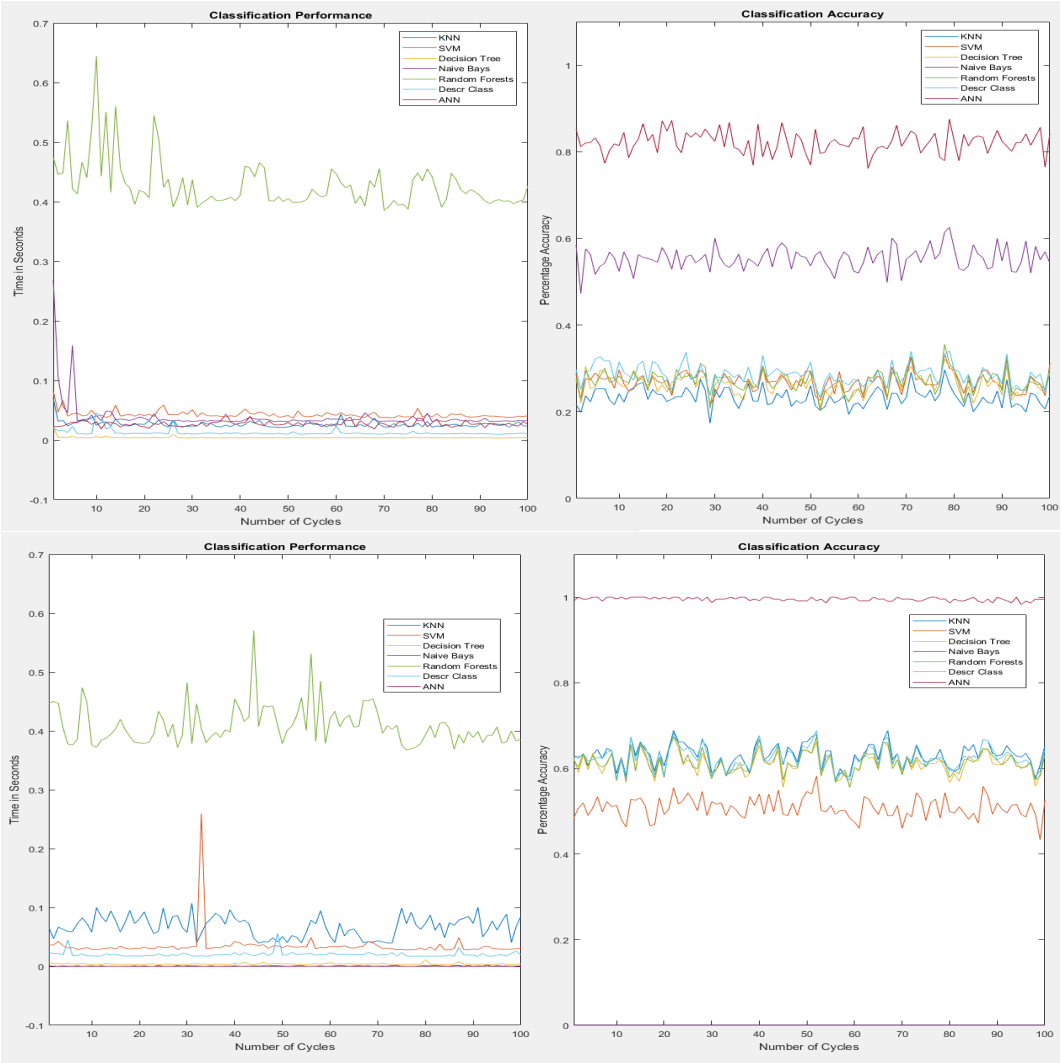


Figure 26: Performance and Accuracy Comparison for Classification Methods (Part 2/2)

Table 7: Robustness Comparison of Unoptimized Feature Classification Methods when Feature Quantity Varies

Data-set	**Feature Count**	**Class Label Count**	**Unoptimized Feature Classification**		**Optmized Feature Classification in %**
			Fused (Wavelet Method) in %	**Without fusion in %**	
DS1	56	14	-6.7 $\sim$ 4.6	-6.9 $\sim$ 5.3	-6.2 $\sim$ 4.3
DS2	78	12	-4.3 $\sim$ 3.9	-5.4 $\sim$ 4.8	-4.0 $\sim$ 3.3
DS3	102	12	-5.2 $\sim$ 4.6	-5.8 $\sim$ 6.6	-5.1 $\sim$ 4.3

5.2.3 Phenotype Switching of Modalities and Bet-Hedging of Features

The phenotype switching was applied at modality level to reduce effect redundant modalities on the detection process. In the proposed phenotypic switching also combines the Bet-Hedging method. Thus, the phenotypic approach optimises at modality level whereas the Bet-Hedging technique works on the detailed level to improve the robustness. Here an experimental validation was conducted to find the effect of datasets with two, three and four modalities. In a first trial run, common fusion methods were applied to the datasets to find their accuracy, robustness and performance. Another set test cycles were run to observe the effects of the learning

classification algorithms used for feature based detection methods. Table 8 shows the comparison between various fusion methods when the proposed method was applied.

Table 8: Accuracy Comparison with Fusion Methods

Fusion Method	**Dataset**	**Label Count**	**Unoptimized**	**Proposed: Pheno-Hedging**
DWT	DS1	14	28 $\sim$ 40 %	36$\sim$ 42 %
	DS2	12	48 $\sim$ 58 %	56 $\sim$ 61 %
	DS3	12	49 $\sim$ 55 %	58 $\sim$ 62 %
PCA	DS1	14	24 $\sim$ 32 %	31 $\sim$ 36 %
	DS2	12	43 $\sim$ 49 %	54 $\sim$ 59 %
	DS3	12	45 $\sim$ 51 %	55 $\sim$ 59 %
Mamdani	DS1	14	21 $\sim$ 33 %	29 $\sim$ 33 %
	DS2	12	40 $\sim$ 46 %	46$\sim$ 50 %
	DS3	12	42 $\sim$ 49 %	49 $\sim$ 53 %

Table 9 shows a comparison between classification accuracies when the feature-sets were extracted after the phenotypic approach was applied.

Table 9: Accuracy Comparison of Feature Classification Methods when Feature Quantity Varies

Dataset	**Number of Features**	**Number of Class Labels**	**SVM**		**DT**		**NB**	
			Un Opt	**Phen**	**Un Opt**	**Phen**	**Un Opt**	**Phen**
DS1	52	14	22	24	18	19	15	16
DS2	78	12	58	61	55	57	64	66
DS3	102	12	49	51	59	61	61	63

Dataset	**Number of Features**	**Number of Class Labels**	**DC**		**KNN**		**ANN**	
			Un Opt	**Phen**	**Un Opt**	**Phen**	**Un Opt**	**Phen**
DS1	52	14	21	23	22	24	57	60
DS2	78	12	59	61	57	59	74	76
DS3	102	12	59	62	60	62	85	87

Table 10 provides cumulative times consumed as a performance comparison between fusion and proposed methods. The comparison was performed to find the impact of the hedging and optimisation over the performance of the fusion techniques.

Table 10: Performance Comparison of Fusion Methods

Fusion Method	**Modality Count**	**Unoptimized**	**Proposed: Phenotype Switching**
DWT	Two	0.0346 $\sim$ 0.0521 s	0.0315 $\sim$ 0.0467 s
	Three	0.0378 $\sim$ 0.0432 s	0.0334 $\sim$ 0.0410 s
	Four	0.0406 $\sim$ 0.0484 s	0.0371 $\sim$ 0.0452 s
PCA	Two	0.0403 $\sim$ 0.0739 s	0.0356 $\sim$ 0.0548 s
	Three	0.0434 $\sim$ 0.0884 s	0.0374 $\sim$ 0.0732 s
	Four	0.0505 $\sim$ 0.0987 s	0.0451 $\sim$ 0.0772 s
Mamdani	Two	0.0279 $\sim$ 0.0421 s	0.0234 $\sim$ 0.0401 s
	Three	0.0367 $\sim$ 0.0647 s	0.0334 $\sim$ 0.0577 s
	Four	0.0409 $\sim$ 0.0747 s	0.0378 $\sim$ 0.0652 s

Table 11 gives a comparison between the robustness of the proposed approach applied to various image fusion methods.

Table 11: Robustness Comparison of Fusion Methods

Fusion Method	**Modality Count**	**Unoptimized**	**Phenotype Switching**
DWT	Two	-6.9 $\sim$ 5.3 %	-6.7 $\sim$ 5.1 %
	Three	-5.4 $\sim$ 4.8 %	-5.1 $\sim$ 4.6 %
	Four	-5.8 $\sim$ 6.6 %	-5.4 $\sim$ 6.3 %
PCA	Two	-6.7 $\sim$ 5.2 %	-6.3 $\sim$ 5.0 %
	Three	-5.2 $\sim$ 4.2 %	-5.0 $\sim$ 4.2 %
	Four	-5.4 $\sim$ 6.3 %	-5.1 $\sim$ 6.0 %
Mamdani	Two	-6.7 $\sim$ 5.1 %	-6.3 $\sim$ 4.5 %
	Three	-5.2$\sim$ 4.6 %	-4.6 $\sim$ 4.1 %
	Four	-5.4$\sim$ 6.2 %	-5.2 $\sim$ 6.0 %

5.2.4 Gene Silencing

Table 9 provides the cumulative times consumed as a performance comparison between fusion and proposed methods. The comparison was performed to find the impact of the hedging and optimisation over the performance of the fusion techniques. The Gene-Silencing or an epigenetic approach affects a temporal sense. The continuous process of acquiring and detecting needs real-time adjustments. This method is capable to periodically analyse the datasets obtained for desired environmental changes and adjust the detection mechanism accordingly. This subsection mainly compares the effect of gene-silencing over the classification results from a robustness perspective. Table 12 shows a comparison between sensor fusion methods with and without Gene-Silencing applied. The values in columns indicate the minimum and maximum percentage of the deflection with respect to the mean accuracy.

Table 13 shows the effect on the classification algorithms when the Gene-Silencing is applied. The values in Table 12 and 13 indicate the sum of the percentage of deflection from mean accuracy for every classification algorithm with and without gene silencing.

Table 12: Robustness Comparison of Fusion Methods

Fusion Method	**Modality Count**	**Unoptimized**	**Proposed: Gene-Silencing**
DWT	Two	-6.9 $\sim$ 5.3 %	-3.7 $\sim$ 3.3 %
	Three	-5.4 $\sim$ 4.8 %	-4.1 $\sim$ 3.9 %
	Four	-5.8 $\sim$ 6.6 %	-4.3 $\sim$ 5.2 %
PCA	Two	-6.7 $\sim$ 5.2 %	-5.3 $\sim$ 4.1 %
	Three	-5.2 $\sim$ 4.2 %	-4.4 $\sim$ 3.1 %
	Four	-5.4 $\sim$ 6.3 %	-4.0 $\sim$ 5.1 %
Mamdani	Two	-6.7 $\sim$ 5.1 %	-4.3$\sim$ 2.9 %
	Three	-5.2 $\sim$ 4.6 %	-3.7$\sim$ 3.5 %
	Four	-5.4 $\sim$ 6.2 %	-4.4 $\sim$ 5.2 %

Table 13: Accuracy Comparison of Feature Classification Methods when Feature Quantity Varies

Dataset	**Number of Features**	**Number of Class Labels**	**SVM**		**DT**		**NB**	
			Un Opt	**GS**	**Un Opt**	**GS**	**Un Opt**	**GS**
DS1	52	14	12%	7%	13%	7%	15%	11%
DS2	78	12	11%	8%	12%	6%	13%	10%
DS3	102	12	12%	7%	13%	7%	11%	9%

Dataset	**Number of Features**	**Number of Class Labels**	**DC**		**KNN**		**ANN**	
			Un Opt	**GS**	**Un Opt**	**GS**	**Un Opt**	**GS**
DS1	52	14	13%	9%	11%	8%	7%	5%
DS2	78	12	12%	9%	10%	7%	6%	4%
DS3	102	12	12%	9%	11%	8%	5%	3%

5.3 Summary

The experimental work in this section is performed on qualitative and quantitative levels. Every algorithm was assessed from various perspectives to validate either one or more parameters from accuracy, performance and robustness. Since every algorithm is intended to serve a different purpose, the tests were carried out accordingly. The GA implementation and phenotypic algorithm were developed mainly for improving accuracy and performance. Whereas, the Gene-Silencing approach was

primarily to improve robustness to acquire more stable results. The GA implementation was analysed for its usage find accuracy when applied with different fusion and classification algorithms. The results are indicative of increase in accuracy (6 to 11%) and performance (5 to 9%). Whereas, the robustness was improved relatively lower magnitude (4 to 7%).

6 Action Research

This chapter presents case studies based on the investigated methodologies. Applicability of the multimodal evolutionary algorithms was tested over three areas applications. These case studies are to capture applications for water quality and surface monitoring, healthcare and machine vision for robotic applications. The environmental monitoring case study is called as iMuDS which is an acronym for 'Internet of Multimodal Data Acquisition Systems'. In case of the healthcare, a solution was developed to use multimodality for patient activity monitoring. A further expansion of this solution is currently under development where evolutionary methods are used for optimising data and decisions. The machine vision algorithms were used for detecting objects by extracting and analysing their most prominent features. The following subsections of this chapter provide a brief overview of the action research conducted.

6.1 iMuDS: IoT Application for Water Surface Monitoring

Due to a continuous presence of a human population, the urban waterways require constant monitoring to detect the presence of contaminants. Many research studies show the litter discharging in the rivers and lakes is a serious threat to species who rely on them. Furthermore, this water as well may become useless for human use. As per few of the alarming statistics, globally two million tons of waste is discharged into urban waterways. Implications of the polluted water include millions of cases of Diarrhea, a threat to the survival of 24% of mammals and 12% of birds who depend on inland waterways.

Many attempts are made to diagnose water quality on a periodic basis using

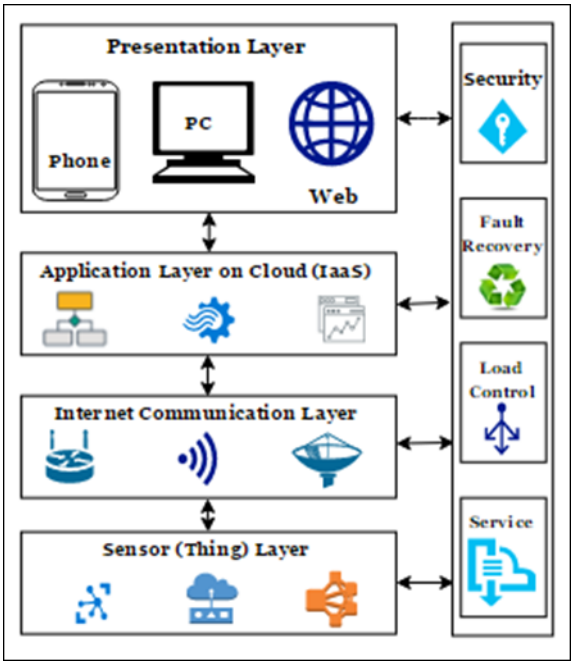


Figure 27: Generalization of IoT Architecture

either manual intervention or by an automated approach. Use of the in-situ sensors to diagnose micro-level impurities is a commonly practised method. Use of video streaming water surface as well provides another level of detail. In this case study, an IoT architecture iMuDS an acronym for `Internet of Multimodal Data Acquisition Systems' is proposed. The iMuDS uses a combination of cameras with in-situ sensors connecting to a cloud-based server using a mobile communication network. The evolutionary intelligence is applied to provide an adaptive detection process. The evolutionary algorithms help to optimise, select and accurately classify object and anomalies.

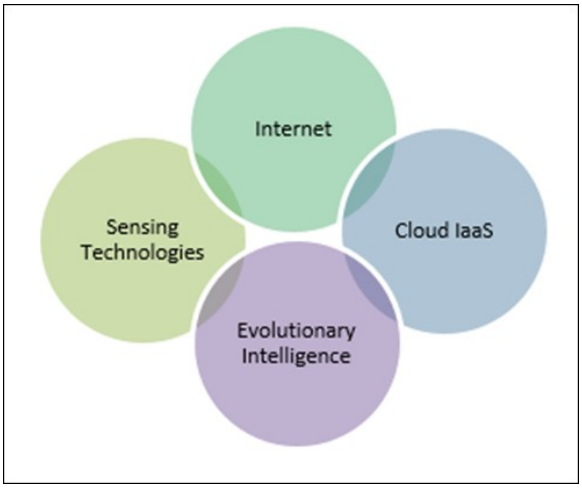


Figure 28: Technological Intersection in IMuDS

6.1.1 Development Rationale

The motives behind iMuDS architectural development using the IoT and evolutionary concepts are as below:

- IoT is an industry standard with proven usefulness in many applications. Due to its growing recognition, many `o the shelf' system components are available.
- Evolutionary computational intelligence can help to improve the efficiency to transfer data packets through internet transmission. Evolutionary algorithms can allow selective transportation of datasets from high dimensional sensing modalities like video cameras. Furthermore, evolutionary intelligence also provides an essential mechanism for data optimisation to improve detection accuracy.
- The Cloud platform provides a subscription-based server infrastructure with relatively moderate cost. Due to readily available cloud environments which

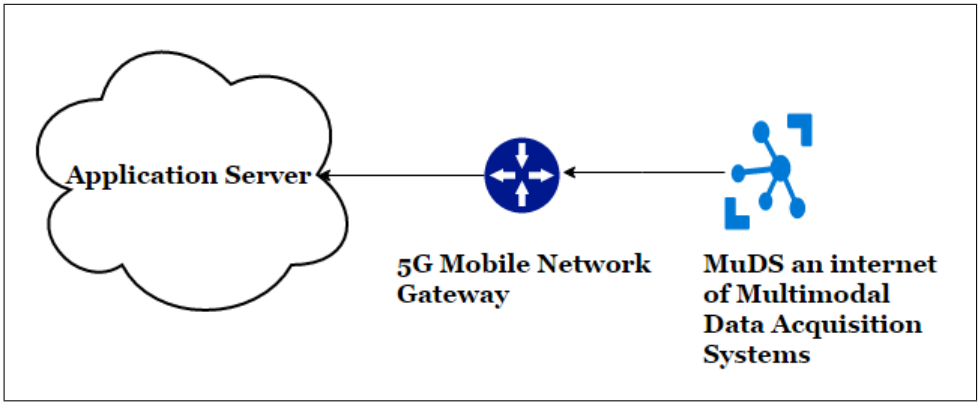


Figure 29: iMuDS Schematic Diagram

match most of the industry specifications, the delivery time for infrastructure development can be reduced substantially.

6.1.2 High-Level Architectural Perspective

The iMuDS architecture has three main elements as follows:

- **Application Server:**
  A cloud-hosted application server with data acquisition, storage and analytic capabilities.
- **MuDS:**
  A multimodal subsystem to combine wireless-sensing and camera modalities to detect a targeted scenario.
- **Mobile Network:**
  Providing internet connectivity to bridge the application server and multiple MuDS sensory nodes.

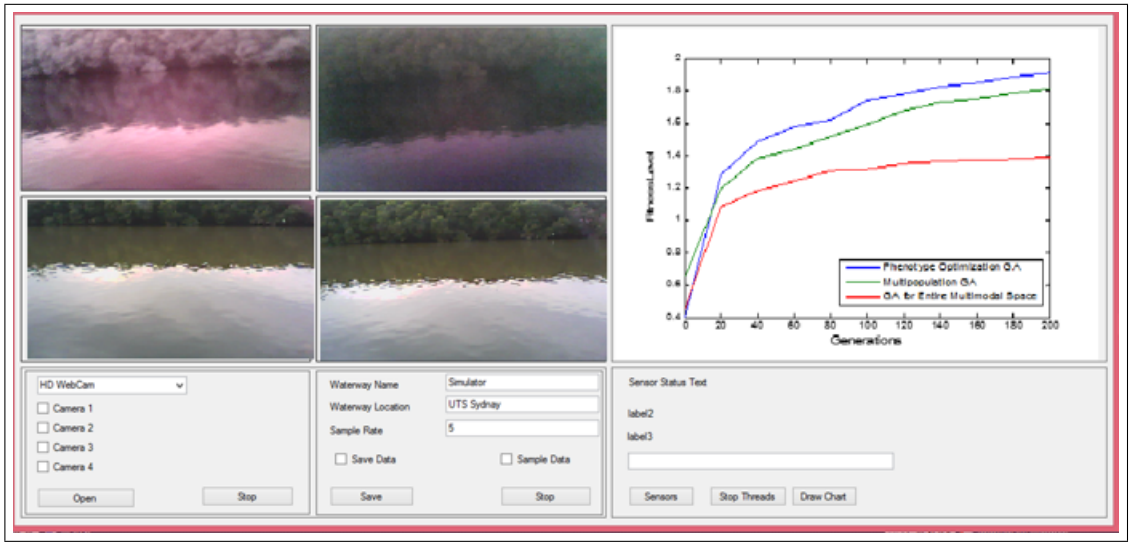


Figure 30: MuDS GUI

6.1.3 MuDS: Multimodal-Data Acquisition System

The MuDS uses two main types of modalities, cameras and wireless sensors. The cameras acquire a water surface level activity whereas the wireless-sensors detect detailed or micro level parameters including turbidity, salinity and so on. Thus, two different modalities are coupled to monitor macro and detailed level contamination in water.

The wireless sensors are clustered and connected to a wireless gateway device. The gateway is a bridging device between the sensor-nodes and the controller unit. Whereas, camera sensors are directly attached to the controller using Universal Serial Bus (USB). The controller device uses a mobile wireless gateway for transmitting acquired data to the application server. Thus, a hardware arrangement for obtaining and dispatching real-time data of a waterway is developed.

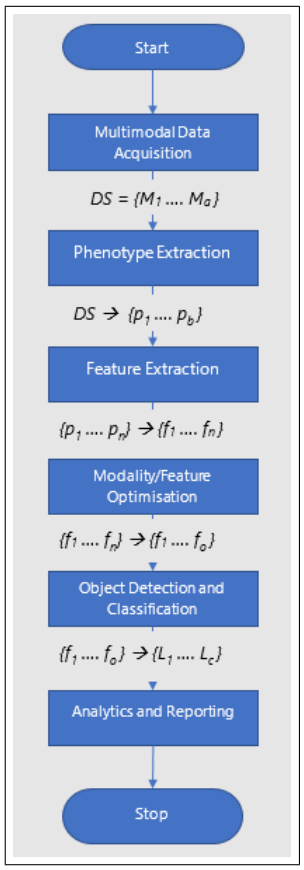


Figure 31: Modality and Feature optimisation Process

6.1.4 Feature Selection and Classification

The data-sets acquired from every MuDS are sampled and converted into a feature form. Every set of modalities and features are optimized at a node and application server. The evolutionary algorithms act here as a switching mechanism which periodically runs an assessment to track the environmental changes and their effects on imaging modalities. Thus, the evolution process here reduces redundancies which improve the efficiency of transmission and accuracy of the detection process. Figure 31 presents a process involved in feature and modality selection.

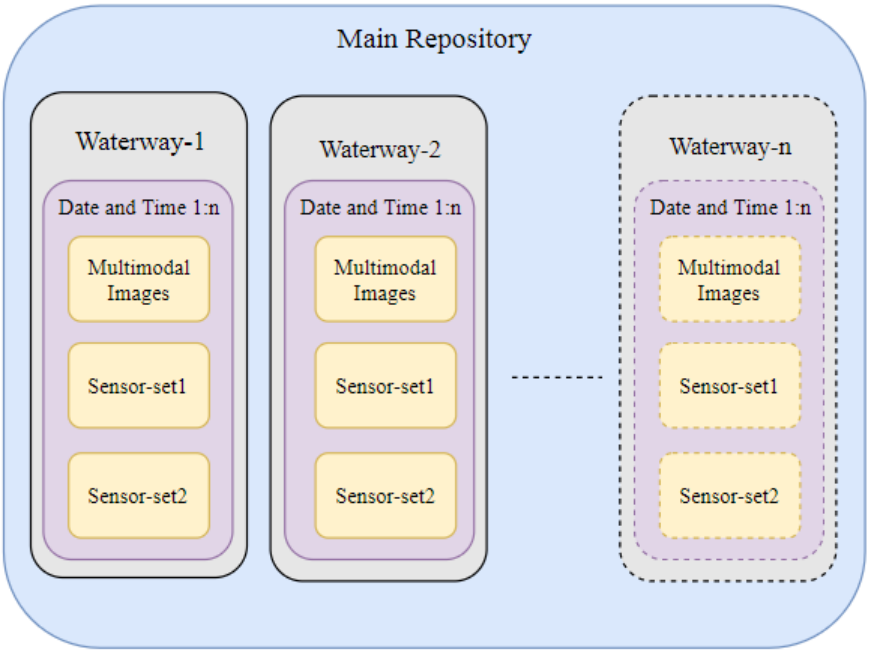


Figure 32: Repository structure

6.1.5 Organization of Repository for Storing Multimodal Data

The data repository was designed with an application specific requirement. The primary considerations were to include physical addresses with time-stamps. It was architected in a hierarchical structure where the central repository has a provision to store data in multiple folders dedicated to various waterways. Within every waterway storage, sensor data is saved as per their acquisition time. Thus, a systematic approach was applied to avoid mix up of the acquired data and preserve it for any future events.

6.1.6 Analytics and Reporting

The application layer operates is a cloud environment and uses multimodal features to convert them into observations. Various business specific reports generated and presented. A typical example of user-level presentation of information extracted is shown in Figures 33 and 34. Pollution density or intensity can be presented either in a mapped picture or by using plots. The purpose here is to develop a reporting scheme readily understandable to users.

Thus, iMuDS provides a complete end-to-end solution using evolutionary intelligence for detecting and classifying multimodal datasets. The further logic is applied and converts the pertinent results into user readable reports. Trials were conducted on a river in Sydney metropolitan and did show very encouraging results.

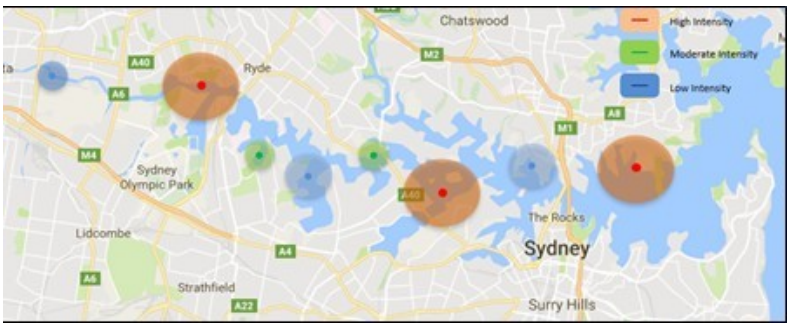


Figure 33: Example of Pollution Monitoring Map

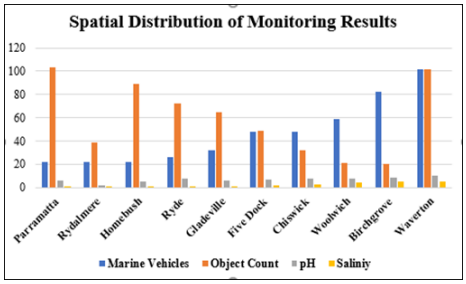


Figure 34: Example of Suburb-wise Pollution Density Plotting

6.2 Evolutionary Multimodal Network for Patient Monitoring

Remote patient monitoring is a significant research study due to limited access to medical services. One of the most common scenarios is an elderly person living independently. In many countries, there is a limited number of trained nursing staff to supervise every individual patient at their residence. In such case routine monitoring of events and activities requires an automated approach. A combination of recent technologies including smartphones and IoT can be used to create an expensive and accessible solution to the problem. Typically, a sensor attached to a human body or an environment he is operating in can be connected to the user's smartphone. The evolutionary algorithms can be employed to conduct an efficient usage of sensory and transmission resources. Here, an ongoing investigative study is in progress to create a system to address the discussed problem. A combination of evolutionary algorithms is employed to select and optimise the patient data.

Combining three algorithms and then applying them consisted of the proposed home-based monitoring domain. The experiment to validate proposed concept includes generation of sensory data and then saving that to several repositories. After that, evolutionary algorithms are used to learn the datasets for their redundancy, environmental dependencies and selection purposes. Initial experiments for validating proposed concepts show very promising signs and merits of it. More enhancements can be done in future to address specific scenarios like sudden environmental changes imposing total outage or severe inaccuracies in the data classification process. Thus, the proposed concepts can be beneficial to an IoT application for health monitoring of elderly population due to its ability to adapt variability in data sets and to classify received data efficiently.

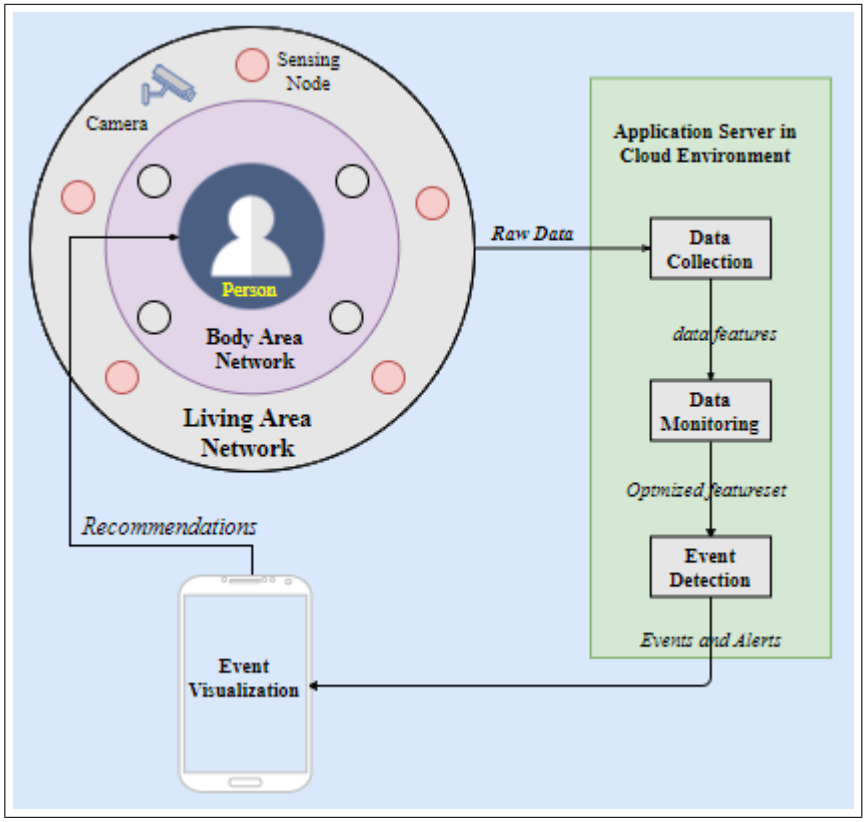


Figure 35: Proposed Architecture for Patient Monitoring System

6.3 Machine Vision Applications

Feature level and modality level detection was achieved for machine vision applications. This work started with an implementation of a GA for detecting significant features of the objects to be monitored. Later a further expansion was done to accommodate infrared and ultra-violet modalities by applying a combination of GA and phenotypic modelling approach. Here the phenotypic method focuses on developing a cluster of features on a periodic basis which is called phenotypes. The phenotype selection allows creating an adaptive mechanism to switch the detection algorithm between changed environmental conditions. A vision system for a Delta parallel robot was developed to detect industrial parts. In another case study, a

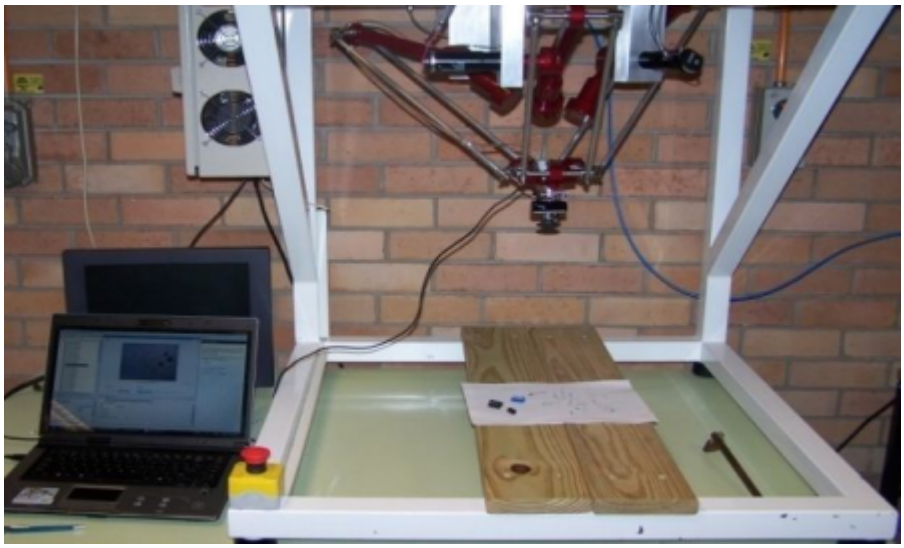


Figure 36: Parallix Delta Robot Setup

machine vision solution for medicine tablet was developed. Both case studies show the advantages of using evolutionary algorithms for detecting targeted objects.

The delta parallel robot was equipped with two conventional vision methods, eye-at-hand or eye-in-hand. More than 40 types of different objects were observed and classified. The vision systems were tested with relevant constraints including partial occlusion, distance variation and illumination fluctuations. Final results of the trials show significant advantages of using genetic algorithms for learning and adapting to environmental changes.

6.4 Summary

Thus, action research with case studies shows the importance of using evolutionary algorithms when a vision system works in unpredictable conditions. Systems like the iMuDS ecosystem consist of a natural environment with very regular human intervention. Here, the environmental dynamism consists of illumination variation,

occlusion and high-reflection. Detecting presence of pollutants within this situation was a huge challenge. Thanks to the evolutionary methods, every changed scenario was learned and adapted to the detection systems. In case of the machine vision solutions, variability consisted within an enclosed laboratory environment. The illumination was varied artificially to verify accuracy and robustness of the vision system. Equipping a GA for tracking minor changes did help to improve detection accuracy. Applicability in the healthcare domain was proved in initial trials; this work is still in progress and will be extended to a case study relating to a patient monitoring.

7 Conclusion

This chapter discusses the conclusion and future directions for the proposed approach. During this discussion, a brief overview of the scholarly contributions related to this research is provided. This chapter is divided into six parts: the first part revisits the proposed problem and solutions, the second part discusses the research findings, the third part presents the scholarly research contributions, the fourth part summarizes limitations of this research work, and the fifth part gives future directions and finally, the concluding remarks are discussed.

7.1 Revisiting the Proposed Method

A multimodal object detection method for dynamic environmental conditions was proposed and validated. This report started with a detailed description of the problem statement where a real-life scenario for object detection problem was considered. An in-depth background study related to the solutions, methods and algorithms for this body of knowledge was conducted. Based on this review, a research opportunity was found to improve the current multimodal detection practices by using a novel adaptive computational approach.

The investigated innovative evolutionary optimisation methodology was seen as a perfect match to achieve this adaptivity. A problem specific set of tools and framework was developed to create an experimental testbed. Once the testbed and dataset models were developed the proposed algorithms were programmed and tested from quantitative and qualitative angles. The results were found promising with very intriguing details. A set of case studies and applications was developed to implement the methodology in real-life scenarios. This subsection revisits the problem and solution considered at the beginning of this study.

7.1.1 Problem

Multimodal computer vision applications have become an essential part of human lives. The camera-based detection techniques are used in day-to-day life for security, health, manufacturing, environmental monitoring and many more possible areas of applications.

These applications many times operate in unpredictable scenarios where surroundings and light sources make their effect on the objects to be detected. The environmental variability can impose limitation in object detection operations due to scarcity in illumination and other obstructive parameters like dust, rain and fog. In such situations, as an adaptive algorithmic solution is required to overcome the environmental limitations.

7.1.2 Hypothesis and Validation

The hypothesis inspired by evolutionary biological concepts was proposed. The hypothesis suggested on the usage of evolutionary computing models for improving accuracy, performance and robustness. It was proved that the evolutionary processes working on the multimodal metadata should vary in data-structure as per changes in the environmental scenario.

An experimental platform capable of creating scenarios to emulate the environmental conditions was developed. Formal and problem specific validation procedures were applied to run the test cycles. The hypothesis was directed by from various perspectives to record its computational response to changed environmental states. During the validation process, traditional fusion and classification techniques were compared with the proposed methodology. The research findings were recorded and published periodically.

Various applications and case studies were developed by embedding the proposed algorithms. These case studies prove the practical usage of the multimodal evolutionary algorithms. Furthermore, the studies demonstrate the capability of applying the proposed algorithms in a practical context and its applicability for real-world solutions.

7.2 Post Validation Research Finding

7.2.1 Computational Perspective

The proposed set of algorithms was validated for the time consumed for classification and detection processes. The proposed feature based classification significantly improves the accuracy of all the fusion and classification methods. Whereas, the performance and robustness as well are marginally enhanced.

In case of the Phenotype selection algorithm, it improves the performance significantly since it selectively uses input modalities. The Gene-Silencing algorithm is found very useful in increasing the robustness of a fusion or classification algorithm. Table 14 summarises the improvements achieved during this experimental validation.

Table 14: High-Level Comparison between Improvements for Proposed Algorithms

Algorithm	**Accuracy**	**Performance**	**Robustness**
Enhanced GA	High	Medium	Marginal
Phenotype Selection	Marginal	High	Marginal
Gene Silencing	Marginal	Marginal	Medium
Overall Effect	High	High	Medium

7.2.2 Application Perspective

Machine Vision

Two case studies were developed for machine vision detection problems for robotic and industrial applications. Camera-based detection methods were added with evolutionary computing intelligence to enhance the adaptivity in the environmental variable workspace. The industrial monitoring applications many times tend to operate in relatively controlled lighting conditions. But in case of degradation in visibility due to smoke and dust, an added modality provides a more prominent bandwidth of information. The evolutionary algorithms allow reducing redundancy from the numerous bandwidth datasets.

Camera based vision techniques are prevalent in biomedical and healthcare applications for diagnosis and real-time monitoring purposes. During this research case studies for patient monitoring were developed using a multimodal approach. A computer vision along with wireless sensor technologies was deployed in the patient environment.

Here, a partially hybrid method was employed to combine macro level information from a camera modality with the micro level information with sensors. A combination of a neural network method with a genetic-algorithm allows combining these two sensory channels and detect human activities. Simulation results show the applicability and benefits of the proposed case study.

IoT & Environmental Monitoring

An in-depth investigation was conducted on an urban waterway monitoring application. This application has a data acquisition system to acquire data from multiple cameras and sensing modalities. A set of evolutionary algorithms to optimise data at modality and feature level was applied to cut down redundancies and improve detection results.

A real-life case study was performed with a system specially developed for this monitoring activity. This concept was further enhanced by deploying it in an 'Internet of Things' or an 'IoT' architecture. Trials conducted on a river in Sydney metropolitan area does show significant advantages of this project. Thus, the applicability of the proposed methodology was proved using a vital application which was deployed in a distributed environment like the IoT.

7.3 Research Contributions

A set of new algorithms and methods have been presented and published in internationally renowned and recognised conferences and journals. The following list provides most up to date publication work related to the methodology:

7.3.1 Intelligent Healthcare - Motion Analysis System for Health Practitioners

- Published*:*

  'Proceedings of the 2010 Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Brisbane, Australia, 7–10 December 2010; pp. 303–308'.

The core objective of this work was to combine sensing paradigms cameras and wireless sensors. The combination allows fusion of macro and feature using an evolutionary algorithm and an extended Kohonen maps to gain an exploratory understanding of the theoretical concepts underpinning the current state of the art

research in the given domain. This work was explicitly intended to be implemented in a healthcare domain.

7.3.2 Automated Tablet Quality Assurance and Identification for Hospital Pharmacies

- Published*:*

  'International Journal of Electronics and Telecommunications, 57(2), pp.153-158'.

A machine vision solution using a genetic algorithm is proposed. A complete system was developed for a miniature sized conveyor belt to detect defects in medicine tablets. This solution provides an in-house solution for hospital pharmacies to identify the type and quality of the medicine.

7.3.3 Parallel Robot Vision Using Genetic Algorithm and Object Centroid

- Published*:*

  'Proceedings of EUROCAST'11 Proceedings of the 13th international conference on Computer Aided Systems Theory - Volume Part II, Springer-Verlag Berlin, Heidelberg, pp. 170-178, 2012'.

An object detection methodology which a genetic algorithm (GA) to detect objects. In this work, randomly extracted sets of coordinates are used to model into chromosomes of the GA. These chromosomes are a possible representation of the region of interests or (RoIs) in the problem search space.

The objective function of the GA periodically assesses the presence of the centroid within the subspace by creating cropped images from the chromosomes. This work provided a foundation for many future investigations. Experimental results did show superiority against the contemporary methods.

7.3.4 Genetically Inspired modelling of the Multimodal Feature Space

- Published:
  '1st Australian Conference on the Applications of Systems Engineering ACASE'12 (p. 26)'.

Evolutionary biological concepts are suggested for modelling the metadata of a multimodal feature set. During this process, concepts including genotype, phenotype and reaction norms are replicated in a computational model. A genotype or gene structure used as an inspiration to model the structure of the metadata.

In this article, the phenotypes are the functional representation of genes to create physical properties of objects like colour, geometry and texture. The phenotypic behaviour is then tracked to search cues to representing environmental states. Finally, these environmental states are used as triggers to select phenotypes. Pilot tests were conducted to verify the functionality of this method.

7.3.5 Evolutionary Switching for Modality Selection in Water Contamination Monitoring

- Published:

  'Proceedings of the 7th International Conference on Broadband and Biomedical Communications (IB2COM), November 5-8, 2012, Sydney, Australia'.

A phenotypic modelling approach developed previously is enhanced further to use it as a modality selection switch. In this work, multimodal image features as a phenotypic representation are extracted and their response is measured. The most sensitive or variable phenotypes are used as environment trackers.

Based on the slope of such phenotypes a modality is selected at a given time. This method needs periodic and continuous learning since it needs to keep itself adapting to varied environmental scenarios. A case study to detect objects on a water surface is validated to prove a practical use of this algorithm. The modality selection algorithm improves the accuracy and performance of the object detection process.

7.3.6 Managing Dynamism of Multimodal Detection in Machine Vision Using Selection of Phenotypes

- Published:

  'International Conference on Computer Aided Systems Theory (pp. 483-490). Springer, Berlin, Heidelberg'.

A combination of phenotypic modelling and a genetic algorithm is used to optimise data at modality and feature level. The genotypes are optimised and selected using the GA to improve detection accuracy and performance. This method was embedded in a machine vision problem to detect objects. Experiments show very encouraging results and superiority as compared to the contemporary practices.

7.3.7 Evolutionary Feature optimisation and Classification for Monitoring Floating Objects

- Published*:*
  'Computational Intelligence and Efficiency in Engineering Systems (pp. 3-16). Springer International Publishing'.

This method uses a bet-hedging principle for modelling a feature subset. In this work, a combination of features with most and least probability of occurrence is used to mitigate the environmental uncertainty.

The learning process is performed by searching the accuracy level of the composition at different proportion levels for both types of features. An SVM classifier is used to classify the feature subset. This method was validated in a case study to monitor water surfaces for floating pollutants. This technique allows improving accuracy and stability of the detection process.

7.3.8 HyMuDS: A Hybrid Multimodal Data Acquisition System

- Published*:*

  'Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on (pp. 107-112). IEEE'.

A multi-modal data acquisition system 'HyMuDS' was developed for acquiring datasets from a camera and wireless-sensing modalities. A modular and scalable system architecture was designed to accommodate requirements of an environmental monitoring system.

A combination of modalities is packaged together using a controller computer. A multithreaded software running on a server computer is interfaced with the controller to acquire datasets continuously. The server also has a graphical user interface to visualise acquired datasets and control the system operations.

The HyMuDS provided an implementation framework for a case study for conducting experiments and rapid prototyping. A trial study conducted using this platform proves it the flexibility to develop application case studies in a quick manner.

7.3.9 Deep Learning and Learning Classifiers for Multimodal Image Processing

- Published*:*

  'Proceedings of Computer Aided System Engineering (APCASE), 2017 Asia-

Pacific Conference on (pp. 21-22) and submitted to Springer International Publishing Berlin, Heidelberg as Book Chapter'.

A survey study to investigate the applicability of the machine learning classifiers for multimodal detection and labelling. In this work set of most commonly practised machine-learning algorithms were applied for multimodal feature classification.

The classification operation was performed on the datasets which were modelled to emulate environmental variability. Effect of variations in training dataset size, input vector size and number of modalities on classification accuracy and performance was measured. An enhanced version of this work was submitted to Springer as a book chapter.

7.3.10 iMuDS: An Internet of Multimodal Data Acquisition and Analysis Systems for Monitoring Urban Waterways

- Published*:*
  '2017 25th International Conference on Systems Engineering (ICSEng) (pp. 431-437), IEEE'.

An IoT case study iMuDS was developed by extending the previously developed system HyMuDS. Here an IoT architecture with the internet of HyMuDS systems was deployed over a geographical span.

The mapped set of nodes were used for monitoring an urban waterway for detecting the presence of pollutants floating on its surface. A business logic server

environment was proposed to deploy in a cloud environment. This server receives raw datasets in a feature form and then analyses them. Graphical user interfaces for reports, data-visualisation and dashboards were designed.

The multimodal evolutionary algorithms were an essential core component of the business logic to acquire, select and optimise the received data. This case study provides a low cost and practical solution for pollution monitoring problem in variable environmental conditions.

7.3.11 IMMuNE: An IoT Methodology for Multimodal Networks using Evolutionary Algorithms in Healthcare Context:

- Published*:*

  'Proceedings of Computer Aided System Engineering (APCASE), 2017 Asia-Pacific Conference on (pp. 21-22) and subsequently submitted to Springer International Publishing Berlin, Heidelberg'.

This work is a short investigative study conducted to find applicability of the evolutionary algorithms in the healthcare domain using IoT technologies. The focus here is on elderly-patient activity monitoring applications.

The idea here is to develop a multimodal detection system to monitor activities, human-body parameters and uncertain events. A non-intrusive approach using silhouettes is used to overcome privacy concerns. Use of evolutionary algorithms for optimising data at modality, feature and decision level is proposed. This work is still in progress and will be completed in the near future.

7.4 Limitations of Findings

The current methodology was tested up to a set of four modalities. There was no attempt made to increase the number of modalities due to inadequate availability of such data resources. Few other scalability parameters missed during this validation work include increasing the number of object labels and features. Insufficient effort was applied to combine spatial and spectral datasets.

The spectral aspect may allow discovering the problem space beyond the visible features. Also, there was no attempt made to include a few other modalities like audio and text. As per the literature review, most of the parallel processing architectures heavily rely on underlying computer hardware. There was no attempt made to exploit the proposed evolutionary methods to apply them on hardware-dependent architectures.

7.5 Future Directions

7.5.1 Computational Perspective

Methodology and validation work completed so far is mostly for processing data in the spatial domain. An insufficient attempt was given to the combining of the hybrid data fusion model using the evolutionary approach.

One of the future directions is to investigate further for developing a composite multimodal solution to connect in spatial with spectral domains. Such an attempt will stimulate the current evolutionary model to build many more real-life applications. One of the best examples can be monitoring of human gait using a camera and analogue sensors.

Most of the research work in this report is related to adding value to conventional image fusion and machine learning classification techniques. A further enhancement

to increase portability of these algorithms is to select and optimise processor cores in a parallel processing environment as technically required.

The recent technological trends are very supportive of the concurrent processing applications using multicore microprocessors. Use of an evolutionary selection algorithm for selecting a situation specific microprocessor core will increase operational efficiency.

7.5.2 Application Perspective

There was no attempt made to investigate the use of a low-resolution wireless camera network or Visual-Sensor-Network or VSN. The VSN technologies can allow discovering information from a large set of cameras in a distributed processing environment. An attempt needs to be given to enhancing the current evolutionary method, optimising information received from massively scalable VSN environments.

The HyMuDS system can capture a very massive amount of sensory data from the targeted environment. An inadequate focus was provided on the scalability perspective in the experimental work as it is considered an exercise in future. Further validation to test the applicability of big-data scenarios needs to be performed.

7.6 Concluding Remarks

This investigation adopted an evolutionary algorithmic methodology that proved to overcome the issues of uncertainty while detecting objects using multimodal camera-based methods. The designed algorithms were tested thoroughly to validate the hypothesis. The results of experimentation are showing significant improvements in accuracy and performance reaching 11% and 9% respectively (detailed results in

chapter 5.2). The robustness was observed to be improved in the range of 4% to 7%. Thus, proposed evolutionary algorithms are seen to be beneficial, and it is recommended the future endeavour could explore highly scalable 'Big-Data' applications.

Part III

Bibliography and Appendix

Bibliography

[1] M. Abou-Rayan, M.M. Jamali, and V. Schmidt. Image fusion of astronomical images via parallel/cloud processing. *Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National (pp. 100-104). IEEE*, 2016.

[2] T. Achalakul and N. Wattanapongsakorn. Performance and reliability modeling for a parallel image fusion system. *Mathematical and computer modelling, 39(1), pp.61-73*, 2004.

[3] Anurag A. Agrawal. Phenotypic plasticity in the interactions and evolution of species. *Science Journal*, 294(5541):321–326, 2001.

[4] T. Akamatsu, T. Ura, H. Sugimatsu, R. Bahl, S. Behera, S. Panda, M. Khan, S.K. Kar, C.S. Kar, S. Kimura, and Y. Sasaki-Yamamoto. A multimodal detection model of dolphins to estimate abundance validated by field experiments. *The Journal of the Acoustical Society of America, 134(3), pp.2418-2426*, 2013.

[5] A. Akoguz, S.K. Pinar, A. Ozdemir, S. Bagis, M. Yucel, and M. Kartal. Remote sensing data fusion algorithms with parallel computing. *Recent Advances in Space Technologies (RAST), 2013 6th International Conference on (pp. 87-92). IEEE*, 2013.

[6] A. Amanatiadis, D. Chrysostomou, D. Koulouriotis, and A. Gasteratos. A fuzzy multi-sensor architecture for indoor navigation. *Imaging Systems and*

Techniques (IST), 2010 IEEE International Conference on, pp.452-457, 1-2 July 2010, 2010.

[7] C.A. Ankenbrandt, B.P. Buckles, and F.E. Petry. Scene recognition using genetic algorithms with semantic nets. *Pattern Recognition Letters 11 (4) (1990) 231-304*, 1990.

[8] S. Balakrishnama and A. Ganapathiraju. Linear discriminant analysis - a brief tutorial. *Institute for Signal and information Processing, 18th edition*, 1998.

[9] Leah Bar, Nir A Sochen, and Nahum Kiryati. Image deblurring in the presence of salt-and-pepper noise. In *Scale Space and PDE Methods in Computer Vision, Springer Media*, volume 5, pages 107–118. Springer, 2005.

[10] B. Barney. Introduction to parallel computing. *Lawrence Livermore National Laboratory, 6(13), p.10*, 2010.

[11] Adrian Bird. Perceptions of epigenetics. *Nature*, 447(7143):396–398, 2007.

[12] G. Botella, H. Martin, M. Santos, and U. Meyer-Baese. Fpga-based multimodal embedded sensor system integrating low-and mid-level vision. *Sensors Journal, 11(8), pp.8164-8179*, 2011.

[13] B.P. Buckles and F.E. Petry. Cloud identification using genetic algorithms and massively parallel computation. *Final Report, Grant No. NAG 5-2216, Center for Intelligent and Knowledge based Systems, Department of Computer Science, Tulane University, New Orleans, LA, June 18, 1996*, 1996.

[14] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 205–211. ACM, 2004.

[15] Gemma A Calvert, Peter C Hansen, Susan D Iversen, and Michael J Brammer. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *Neuroimage Journal*, 14(2):427–438, 2001.

[16] Zenon Chaczko and Anup Kale. Automated tablet quality assurance and identification for hospital pharmacies. *International Journal of Electronics and Telecommunications*, 57(2):153–158, 2011.

[17] P. Chalermwat, T. El-Ghazawi, and J. LeMoigne. Ga-based parallel image registration on parallel clusters. *IPPS/SPDP Workshops, 1999, pp. 257-265*, 1999.

[18] D. Chen, H. Bourlard, and J.P. Thiran. Text identification in complex background using svm. *Computer Vision and Pattern Recognition, 2001. CVPR 2001 Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 2, pp. II-II). IEEE*, 2001.

[19] L. Chen, Y. Ma, P. Liu, J. Wei, W. Jie, and J. He. A review of parallel computing for large-scale remote sensing image mosaicking. *Cluster Computing, 18(2), pp.517-529*, 2015.

[20] Min Chen, Sergio Gonzalez, Athanasios Vasilakos, Huasong Cao, and Victor C.M. Leung. Body area networks: A survey. *Mobile networks and applications*, 16(2):171–193, 2011.

[21] S.C. Chen, M.L. Shyu, C. Zhang, and M. Chen. A multimodal data mining framework for soccer goal detection based on decision tree logic. *International Journal of Computer Applications in Technology, 27(4), pp.312-323*, 2006.

[22] Y. Cheng, Y. Li, and R. Zhao. A parallel image fusion algorithm based on wavelet packet. *Signal Processing, 2006 8th International Conference on (Vol. 2). IEEE*, 2006.

[23] C.K. Chow, H.T. Tsui, and T. Lee. Surface registration using dynamic genetic algorithm, technical report 001. *CVIP Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, 2001*, 2001.

[24] Raquel C. Conceicao, Martin O'Halloran, Ricardo M. Capote, Claudia S. Ferreira, Nuno Matela, Hugo A. Ferreira, Martin Glavin, Edward Jones, and Pedro Almeida. Development of breast and tumour models for simulation of novel multimodal pem-uwb technique for detection and classification of breast tumours. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, pages 2769–2772. IEEE, 2012.

[25] R.M. Couture, K. Tominaga, J. Starrfelt, S.J. Moe, O. Kaste, and R.F. Wright. Modelling phosphorus loading and algal blooms in a nordic agricultural catchment-lake system under changing land-use and climate. *Environmental Science: Processes and Impacts, 16(7), pp.1588-1599*, 2014.

[26] Dipankar Dasgupta and Zbigniew Michalewicz. *Evolutionary algorithms in engineering applications.* Springer Science and Business Media, 2013.

[27] E.R. Davies. *Machine vision: theory, algorithms, practicalities.* Elsevier, 2005.

[28] Thomas J. DeWitt, Andrew Sih, and David Sloan Wilson. Costs and limits of phenotypic plasticity. *Trends in Ecology and Evolution*, 13(2):77–81, 1998.

[29] S. Ding, H. Li, C. Su, J. Yu, and F. Jin. Evolutionary artificial neural networks: a review. *Artificial Intelligence Review, pp.1-10.*, 2013.

[30] G. Dong and G. Kuang. Target recognition via information aggregation through dempster-shafer's evidence theory. *IEEE Geoscience and Remote Sensing Letters, 12(6), pp.1247-1251*, 2015.

[31] A. Eklund, P. Dufort, D. Forsberg, and S.M. LaConte. Medical image processing on the gpu-past, present and future. *Medical image analysis, 17(8), pp.1073-1094*, 2013.

[32] M.P. Feret and J.I. Glasgow. Hybrid case-based reasoning for the diagnosis of complex devices. *Proceedings of Eleventh National Conference on Artificial Intelligence, pp 168-175, Washington D.C., July 11-15 1993, AAAI Press/MIT Press*, 1993.

[33] A. Fijany and F. Hosseini. Image processing applications on a low power highly parallel simd architecture. *Aerospace Conference, 2011 IEEE (pp. 1-12). IEEE*, 2011.

[34] A. Fleury, M. Vacher, and N. Noury. Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE Transactions on Information Technology in Biomedicine, 14(2), p.274-283*, 2010.

[35] Frank Willi Floeth, Dirk Pauleit, Hans-Jorg Wittsack, Karl Josef Langen, Guido Reifenberger, Kurt Hamacher, Martina Messing-Junger, Karl Zilles, Friedrich Weber, Walter Stummer, and et al. Multimodal metabolic imaging of cerebral gliomas: positron emission tomography with [18f] fluoroethyl-l-tyrosine and magnetic resonance spectroscopy. *Journal of Neurosurgery*, 102(2):318–327, 2005.

[36] J.S. Franco and E. Boyer. Fusion of multiview silhouette cues using a space occupancy grid. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on (Vol. 2, pp. 1747-1753). IEEE*, 2005.

[37] Andrew C Gallagher. Using vanishing points to correct camera rotation in images. In *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*, pages 460–467. IEEE, 2005.

[38] K. Garg and S.K. Nayar. Detection and removal of rain from videos. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE*, 2004.

[39] A.K. Garga, K.T. McClintic, R.L. Campbell, Yang Chih-Chung, M.S. Lebold, and T.A. Hay. Hybrid reasoning for prognostic learning in cbm systems. *Aerospace Conference, 2001, IEEE Proceedings, vol.6, no., pp.2957-2969 vol.6, 2001*, 2001.

[40] T. Giannakopoulos, A. Makris, D.I. Kosmopoulos, S.J. Perantonis, and S. Theodoridis. Audio-visual fusion for detecting violent scenes in videos. *SETN (pp. 91-100)*, 2010.

[41] X.E. Gros, J. Bousigue, and K. Takahashi. Ndt data fusion at pixel level. *NDT and E International, 32(5), pp.283-292*, 1999.

[42] Svensson L. Guerriero, M. and P. Willett. Bayesian data fusion for distributed target detection in sensor networks. *IEEE Transactions on Signal Processing*, 2010.

[43] M. Guillaumin, J. Verbeek, and C. Schmid. Multimodal semi-supervised learning for image classification. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 902-909). IEEE*, 2010.

[44] P. Gupta, T. Mehrotra, A. Bansal, and B. Kumari. Multimodal sentiment analysis and context determination: Using perplexed bayes classification. in. *Automation and Computing (ICAC), 2017 23rd International Conference on (pp. 1-6). IEEE*, 2017.

[45] L. Gustafsson and A.P. Paplinski. Bimodal integration of phonemes and letters: an application of multimodal self-organizing networks. in neural networks, 2006. *IJCNN'06. International Joint Conference on (pp. 312-318). IEEE*, 2006.

[46] A. Ben Hamza, Y. He, H. Krim, and A. Willsky. A multiscale approach to pixel-level image fusion. *Integrated Computer-Aided Engineering, 12(2), pp.135-146*, 2005.

[47] N. Hautiere, E. Bigorgne, J. Bossu, and D. Aubert. Meteorological conditions processing for vision-based traffic monitoring. *The Eighth International Workshop on Visual Surveillance-VS2008*, 2008.

[48] S. Le Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar. Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE transactions on geoscience and remote sensing, 35(4), pp.1018-1031*, 1997.

[49] A. Hill and C.J. Taylor. Model-based image interpretation using genetic algorithms. *Image and Vision Computing 10 (5) (1992) 295-300*, 1992.

[50] Keith R. Hopper. Risk-spreading and bet-hedging in insect population biology. *Annual Review of Entomology Journal*, 44(1):535–560, 1999.

[51] M.K. Hossain, M.A. Ashique, M.A. Ibtehaz, and J. Uddin. January. parallel edge detection using sobel algorithm with contract-time anytime algorithm in cuda. *In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA) (p. 130). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp)*, 2016.

[52] M.S. Hosseini and A.M. Eftekhari-Moghadam. Fuzzy rule-based reasoning approach for event detection and annotation of broadcast soccer video. *Applied Soft Computing, 13(2), pp.846-866*, 2013.

[53] John Immerkaer. Fast noise variance estimation. *Computer vision and image understanding*, 64(2):300–302, 1996.

[54] F. Ino, K. Ooyama, and K. Hagihara. . a data distributed parallel algorithm for nonrigid image registration. *Parallel Computing, 31(1), pp.19-43*, 2005.

[55] Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck. *Machine Vision.* MacGraw Hill Series, 1995.

[56] MingHua Jiang, Ming Hu, Tao Peng, and Yi-Xiang Ding. Fuzzy reliability analysis of a multi-sensor fusion system. *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on, vol.3, pp.81-84, 18-20 Oct. 2008*, 2008.

[57] Anup Kale and Zenon Chaczko. Genetically inspired modeling of the multimodal feature space. In *1st Australian Conference on the Applications of Systems Engineering ACASE'12*, page 26, 2012.

[58] Anup Kale and Zenon Chaczko. Evolutionary feature optimization and classification for monitoring floating objects. In *Computational Intelligence and Efficiency in Engineering Systems*, pages 3–16. Springer, 2015.

[59] Anup Kale, Zenon Chaczko, and Imre Rudas. Parallel robot vision using genetic algorithm and object centroid. In *International Conference on Computer Aided Systems Theory*, pages 170–178. Springer, 2011.

[60] Anup Kale, Zenon Chaczko, and Imre Rudas. Managing dynamism of multimodal detection in machine vision using selection of phenotypes. In *International Conference on Computer Aided Systems Theory*, pages 483–490. Springer, 2013.

[61] Anup Kale, Zenon Chaczko, and Shaher Slehat. Hymuds: A hybrid multimodal data acquisition system. In *Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on*, pages 107–112. IEEE, 2015.

[62] Anup Kale, Zenon Chaczko, and Raniyah Wazirali. Evolutionary switching for modality selection in water contamination monitoring. *Proceedings of the 7th IB2COM, November 5-8, 2012, Sydney, Australia*, 2012.

[63] R. Kastner, A. Lin, C. Schurgers, J. Jaffe, P. Franks, and B.S. Stewart. Sensor platforms for multimodal underwater monitoring. *Green Computing Conference (IGCC), 2012 International (pp. 1-7). IEEE*, 2012.

[64] S.W. Keckler, W.J. Dally, B. Khailany, M. Garland, and D. Glasco. Gpus and the future of parallel computing. *IEEE Micro, 31(5), pp.7-17*, 2011.

[65] Er Kenneth, AU Rajendra, N Kannathal, and Choo Min Lim. Data fusion of multimodal cardiovascular signals. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 4689–4692. IEEE, 2005.

[66] W.T. Kerr, E.S. Hwang, K.R. Raman, S.E. Barritt, A.B. Patel, J.M. Le, J.M. Hori, E.C. Davis, C.T. Braesch, E.A Janio, and E.P. Lau. Multimodal diagnosis of epilepsy using conditional dependence and multiple imputation. *Pattern Recognition in Neuroimaging, 2014 International Workshop on (pp. 1-4). IEEE*, 2014.

[67] T.K. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE transactions on pattern analysis and machine intelligence, 27(3), pp.318-327*, 2005.

[68] Ron Kohavi and et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, volume 14, pages 1137–1145. Stanford, CA, 1995.

[69] H.R. Kolar, J. Cronin, P. Hartswick, A.C. Sanderson, J.S. Bonner, L. Hotaling, R.F. Ambrosio, Z. Liu, M.L. Passow, and M.L. Reath. Complex real-time en-

vironmental monitoring of the hudson river and estuary system. *IBM Journal of Research and Development, 53(3), pp.4-1*, 2009.

[70] Abdullah Konak, David W. Coit, and Alice E. Smith. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety, Volume 91, Issue 9, Special Issue - Genetic Algorithms*, 1991.

[71] S.B. Kotsiantis, I. Zaharakis, and P. Pintelas. Supervised machine learning: A review of classification techniques. *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies Pages 3-24, IOS Press Amsterdam, The Netherlands*, 2007.

[72] J.R. Koza. Survey of genetic algorithms and genetic programming. *WESCON'95. Conference record. 'Microelectronics Communications Technology Producing Quality Products Mobile and Portable Power Emerging Technologies*, 1995.

[73] Anders Wang Kristensen, Tomas Akenine-Moller, and Henrik Wann Jensen. Precomputed local radiance transfer for real-time lighting design. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 1208–1215. ACM, 2005.

[74] Vladimir Kubelka, Lorenz Oswald, Francois Pomerleau, Francis Colas, Tomas Svoboda, and Michal Reinstein. Robust data fusion of multimodal sensory information for mobile robots. *Journal of Field Robotics*, 32(4):447–473, 2015.

[75] M. Kumar and S. Dass. A total variation-based algorithm for pixel-level image fusion. *IEEE Transactions on Image Processing, 18(9), pp.2137-2143*, 2009.

[76] Praveen Kumar, Ankush Mittal, and Padam Kumar. Addressing uncertainty in multi-modal fusion for improve object detection in yn mic environment. *Information Fusion*, Volume 11, Issue 4:Pages 311–324, 2010.

[77] F. De la Torre and T. Kanade. Multimodal oriented discriminant analysis. *Proceedings of the 22nd international conference on Machine learning (pp. 177-184). ACM*, 2005.

[78] En Li, Caroline Beard, and Rudolf Jaenisch. Role for dna methylation in genomic imprinting. *Nature Journal*, 366(6453):362, 1993.

[79] W. Li, S. Prasad, J.E. Fowler, and L.M. Bruce. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing, 50(4), pp.1185-1198*, 2012.

[80] A. Liaw and M. Wiener. Classification and regression by randomforest. *R news, 2(3), pp.18-22*, 2002.

[81] Y. Liu, X. Feng, and Z. Zhou. Multimodal video classification with stacked contractive autoencoders. *Signal Processing, 120, pp.761-766*, 2016.

[82] Oscar J. Romero Lupez and Angelica de Antonio. Analysis of emergent properties in a hybrid bioinspired architecture for cognitive agents. *Innovations in Hybrid Intelligent Systems Advances in Soft Computing, 2007, Volume 44/2007,p.p. 1-8*, 2007.

[83] Igor V. Maslov and Izidor Gertner. Multi-sensor fusion: an evolutionary algorithm approach. *Information Fusion, Volume 7, Issue 3, September 2006, Pages 304-330*, 2006.

[84] Thomas Mitchell-Olds, John H Willis, and David B Goldstein. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, 8(11):845–856, 2007.

[85] K.P. Murphy. *Naive Bayes Classifiers.* University of British Columbia, 2006.

[86] F. Murtagh, A.E. Raftery, and J.L. Starck. Bayesian inference for multiband image segmentation via model-based cluster trees. *Image and Vision Computing, 23(6), pp.587-596*, 2005.

[87] V.P.S. Naidu and J.R. Raol. Pixel-level image fusion using wavelets and principal component analysis. *Defence Science Journal, Vol. 58, No. 3, May 2008, pp. 338-352*, 2008.

[88] E. O'Connor, A.F. Smeaton, and N.E. O'Connor. A multi-modal event detection system for river and coastal marine monitoring applications. *OCEANS, 2011 IEEE-Spain (pp. 1-10). IEEE*, 2011.

[89] M.R. Paulsen and W.F. McClure. Illumination for computer vision systems. *Transactions of the ASAE*, 29(5):1398–1404, 1986.

[90] G. Pavlin and J. Nunnink. Inference meta models: Towards robust information fusion with bayesian networks. *Information Fusion, 2006 9th International Conference on, vol., no., pp.1-8, 10-13 July 2006*, 2006.

[91] Dimitri Perrin, Heather J Ruskin, Martin Crane, and Ray Walshe. Epigenetic modelling. *ERCIM News*, 72:46, 2008.

[92] V.S. Petrovic and C.S. Xydeas. Gradient-based multiresolution image fusion. *IEEE Transactions on Image processing, 13(2), pp.228-237*, 2004.

[93] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[94] A. Plaza, J. Plaza, A. Paz, and S. Sanchez. Parallel hyperspectral image and signal processing. *IEEE Signal Processing Magazine, 28(3), pp.119-126*, 2011.

[95] P. Pong and S. Challa. Empirical analysis of generalised uncertainty measures with dempster-shafer fusion. *Information Fusion, 2007 10th International Conference on, pp.1-9, 9-12 July 2007*, 2007.

[96] C. Prakash, S. Rajkumar, and P.C. Mouli. Medical image fusion based on redundancy dwt and mamdani type min-sum mean-of-max techniques with quantitative analysis. *Recent Advances in Computing and Software Systems (RACSS), 2012 International Conference on (pp. 54-59). IEEE*, 2012.

[97] P. Prodanov and A. Drygajlo. Bayesian networks based multi-modality fusion for error handling in human-robot dialogues under noisy conditions. *Speech Communication, 45(3), pp.231-248*, 2005.

[98] D. Rao, M. De Deuge, N. Nourani-Vatani, S.B. Williams, and O. Pizarro. Multimodal learning and inference from visual and remotely sensed data. *The International Journal of Robotics Research, 36(1), pp.24-43*, 2017.

[99] Aharon Razin and Arthur D. Riggs. Dna methylation and gene function. *Science*, 210(4470):604–610, 1980.

[100] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. In *Encyclopedia of database systems*, pages 532–538. Springer, 2009.

[101] Olivier Romain, Aymeric Histace, Juan Silva, Jade Ayoub, Bertrand Granado, Andrea Pinna, Xavier Dray, and Philippe Marteau. Towards a multimodal wireless video capsule for detection of colonic polyps as prevention of colorectal cancer. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–6. IEEE, 2013.

[102] J.M. Rouet. Robust 3-d elastic multimodality image registration through genetic algorithms. *Internal Report, Thesard Departement ITI, 27 May, 1999*, 1999.

[103] Daniel L Ruderman and William Bialek. Statistics of natural images: Scaling in the woods. In *Advances in neural information processing systems*, pages 551–558, 1994.

[104] S.R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics, 21(3), pp.660-674. 14*, 1991.

[105] D.K. Sahu and M.P. Parsai. Different image fusion techniques-a critical review. *International Journal of Modern Engineering Research (IJMER)*, 2(5):4298–4301, 2012.

[106] T. Saitoh and R. Konishi. Lip reading using video and thermal images. *SICE-ICASE, 2006 International Joint Conference, pp.5011-5015, 18-21 Oct. 2006*, 2006.

[107] A. Saxena, B. Wu, and G. Vachtsevanos. A hybrid reasoning architecture for fleet vehicle maintenance. *Instrumentation and Measurement Magazine, IEEE, vol.9, no.4, pp.29-36, Aug.2006*, 2006.

[108] S.G. Schlosser, J.M. Trenkle, and R.C. Vogt. Feature set optimization for the recognition of arabic characters using genetic algorithms. *Proceedings of the 21st Applied Imagery Pattern Recognition Workshop, Proc. SPIE, vol. 1838, Washington, DC, 1992, pp. 64-73*, 1992.

[109] B. Scholkopf and A.J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. *MIT press*, 2001.

[110] S.T. Shivappa, M.M. Trivedi, and B.D. Rao. Person tracking with audio-visual cues using the iterative decoding framework. *Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on, pp.260-267, 1-3 Sept. 2008*, 2008.

[111] R. Singh, M. Vatsa, and A. Noore. Multimodal medical image fusion using redundant discrete wavelet transform. *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on (pp. 232-235). IEEE*, 2009.

[112] A. Singhal, J. Luo, and C. Brown. A multilevel bayesian network approach to image sensor fusion. in information fusion, 2000. fusion 2000. *Proceedings of the Third International Conference on (Vol. 2, pp. WEB3-9). IEEE*, 2000.

[113] A.H.S. Solberg. Remote sensing of ocean oil-spill pollution. *Proceedings of the IEEE, 100(10), pp.2931-2945*, 2012.

[114] M. Srinivas and L.M. Patnaik. Adaptive probabilities of crossover and mutation in genetic algorithms. *Systems, Man and Cybernetics, IEEE Transactions on, vol.24, no.4, pp.656-667, Apr 1994*, 1994.

[115] D. Su, P. Fung, and N. Auguin. Multimodal music emotion classification using adaboost with decision stumps. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 3447-3451). IEEE*, 2013.

[116] Timothy Sullivan, Edel O'Connor, Dian Zhang, Ciprian Briciu-Burghina, Brendan Heery, Noel E. O'Connor, Alan F. Smeaton, and Fiona Regan. Integration of multimodal data sources in designing decision support tools for marine resource management. *ICES*, 2010.

[117] Xiaoyun Sun, Yongchun Liang, Dandan Lv, Donghui Liu, and Jianpeng Bian. Application of kernel c-means clustering and dempster-shafer theory of evidence in power transformers fault diagnosis. *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on, vol.3, pp.1060-1063, 10-12 Aug. 2010*, 2010.

[118] R. Tinos and S. Yang. Genetic algorithms with self-organizing behaviour in dynamic environments. *Genetic algorithms with self-organizing behaviour in dynamic environments, Studies in Computational Intelligence*, 2007.

[119] Alexander Toet. The tno multiband image data collection. *Data in brief*, 15:249, 2017.

[120] J.M. Trenkle, S.G. Schlosser, and R.C. Vogt. Morphological feature set optimization using the genetic algorithm. *Proceedings of Image Algebra and Morphological Processing II SPIE, vol. 1568, San Diego, CA, 1991, pp. 212-223*, 1991.

[121] B. Turton, T. Arslan, and D. Horrocks. A hardware architecture for a parallel genetic algorithm for image registration. *Proceedings of the IEE Colloquium on Genetic Algorithms in Image Processing and Vision, October 1994, pp. 11/1 -11/6*, 1994.

[122] G.K. Verma and U.S. Tiwary. Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage, 102, pp.162-172*, 2014.

[123] J. Wan and Y. Cheng. Remote sensing monitoring of gulf of mexico oil spill using envisat asar images. In *2013 21st International Conference on Geoinformatics*, pages 1–5, June 2013.

[124] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li. A comparative analysis of image fusion methods. *IEEE transactions on geoscience and remote sensing, 43(6), pp.1391-1402*, 2005.

[125] Allen M Waxman, Alan N Gove, Michael C Siebert, David A Fay, James E Carrick, Joseph P Racamato, Eugene D Savoye, Barry E Burke, Robert K Reich, William H McGonagle, and et al. Progress on color night vision: visible/ir fusion, perception and search, and low-light ccd imaging. In *Enhanced and Synthetic Vision 1996*, volume 2736, pages 96–108. International Society for Optics and Photonics, 1996.

[126] A.M. Waxman, A.N. Gove, D.A. Fay, J.P. Racamato, J.E. Carrick, M.C. Seibert, and E.D. Savoye. Color night vision: opponent processing in the fusion of visible and ir imagery. *Neural Networks, 10(1), pp.1-6*, 1997.

[127] H. Wei, M. Viallon, B.M. Delattre, K. Moulin, F. Yang, P. Croisille, and Y. Zhu. Free-breathing diffusion tensor imaging and tractography of the human heart in healthy volunteers using wavelet-based image fusion. *IEEE transactions on medical imaging, 34(1), pp.306-316*, 2015.

[128] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.Y. Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing, 53(7), pp.3658-3668*, 2015.

[129] L. Winkler. Bayesian infeence and decision. *Second Edition, Probabilistic Publishing, pp. 49-99, 2003*, 2003.

[130] Huadong Wu, M. Siegel, R. Stiefelhagen, and Jie Yang. Sensor fusion using dempster-shafer theory [for context-aware hci]. *Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE, vol.1, pp. 7- 12 vol.1, 2002*, 2002.

[131] J.K. Wu and Y. Wong. Bayesian approach for data fusion in sensor networks. *In Information Fusion, 2006 9th International Conference on (pp. 1-5). IEEE*, 2006.

[132] X. Xue, G. Lei, W. Hongfu, and X. Fang. A parallel fusion method of remote sensing image based on ihs transformation. *Image and Signal Processing (CISP), 2011 4th International Congress on (Vol. 3, pp. 1600-1603). IEEE*, 2011.

[133] Z. Xue, S. Antani, L.R. Long, D. Demner-Fushman, and G.R. Thoma. Classification of ct figures in biomedical articles based on body segments. *Healthcare Informatics (ICHI), 2013 IEEE International Conference on (pp. 264-268). IEEE*, 2013.

[134] CS Xydeas, , and V Petrovic. Objective image fusion performance measure. *Electronics letters*, 36(4):308–309, 2000.

[135] B. Yang and S. Li. Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement, 59(4), pp.884-892*, 2010.

[136] Bo Yang, Zhong liang Jing, and Hai tao Zhao. Review of pixel-level image fusion. *Journal of Shanghai Jiaotong University (Science)*, 15(1):6–12, Feb 2010.

[137] Shengxiang Yang and Xin Yao. Experimental study on population-based incremental learning algorithms for dynamic optimization problems. *Soft Computing*, 9(11):815–834, 2005.

[138] X. Yang, P. Wang, Y. Du, and H. Zhou. A data-distributed parallel algorithm for wavelet-based fusion of remote sensing images. *Frontiers of Computer Science in China, 1(2), pp.231-240*, 2007.

[139] Z. Yang, Y. Zhu, and Y. Pu. Parallel image processing based on cuda. In *Computer Science and Software Engineering, 2008 International Conference on (Vol. 3, pp. 198-201). IEEE*, pages 198–201, 2008.

[140] Jae-Hyoung Yu, Hwan-Ik Chung, and Hernsoo Hahn. Walking assistance system for sight impaired people based on a multimodal information transformation technique. In *ICCAS-SICE, 2009*, pages 1639–1643. IEEE, 2009.

[141] Frederic Zana and Jean-Claude Klein. A multimodal registration algorithm of eye fundus images using vessels detection and hough transform. *IEEE transactions on Medical Imaging*, 18(5):419–428, 1999.

[142] B. Zhang, W. Perrie, X. Li, and W.G. Pichel. Mapping sea surface oil slicks using radarsat2 quad-polarization sar image. *Geophysical Research Letters, 38(10)*, 2011.

[143] M.L. Zhang and Z.H. Zhou. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering, 26(8), pp.1819-1837*, 2014.

[144] Xuemei Zhang, D.A. Silverstein, J.E. Farrell, and B.A. Wandell. Color image quality metric s-cielab and its application on halftone texture visibility. *COMPCON, pp.44, Proceedings of COMPCON 97*, 1997.

[145] J. Zhao and H. Zhou. Design and optimization of remote sensing image fusion parallel algorithms based on cpu-gpu heterogeneous platforms. *Image and Signal Processing (CISP), 2011 4th International Congress on (Vol. 3, pp. 1623-1627). IEEE*, 2011.

[146] Yi Zheng and Ping Zheng. Multisensor image fusion using fuzzy logic for surveillance systems. *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on, vol.2, pp.588-592, 10-12 Aug. 2010*, 2010.

8 Appendix

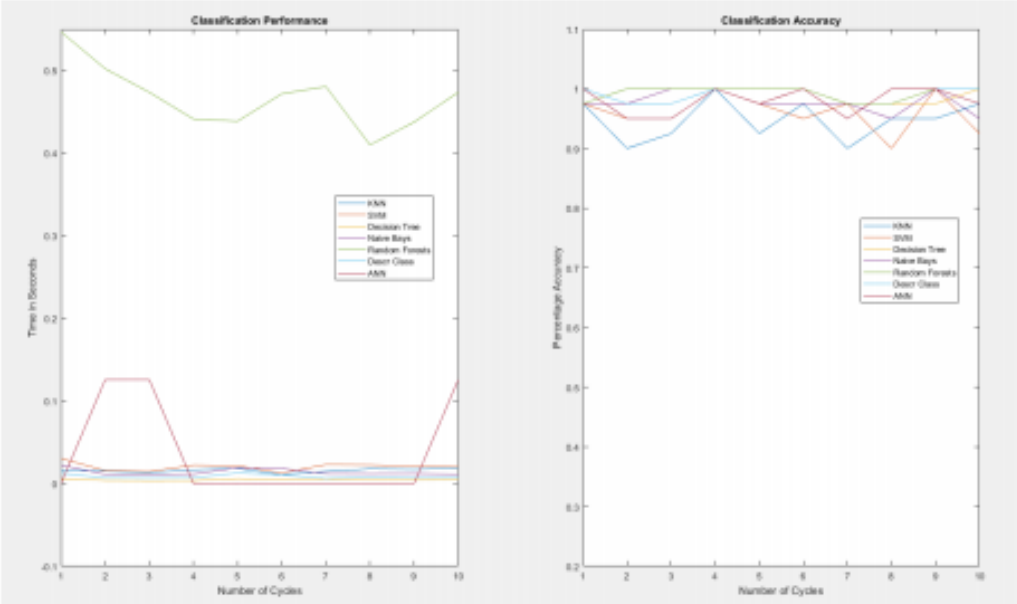
8.1 Unimodal Image Detection Experimetal Results

8.1.1 Experiment Specification

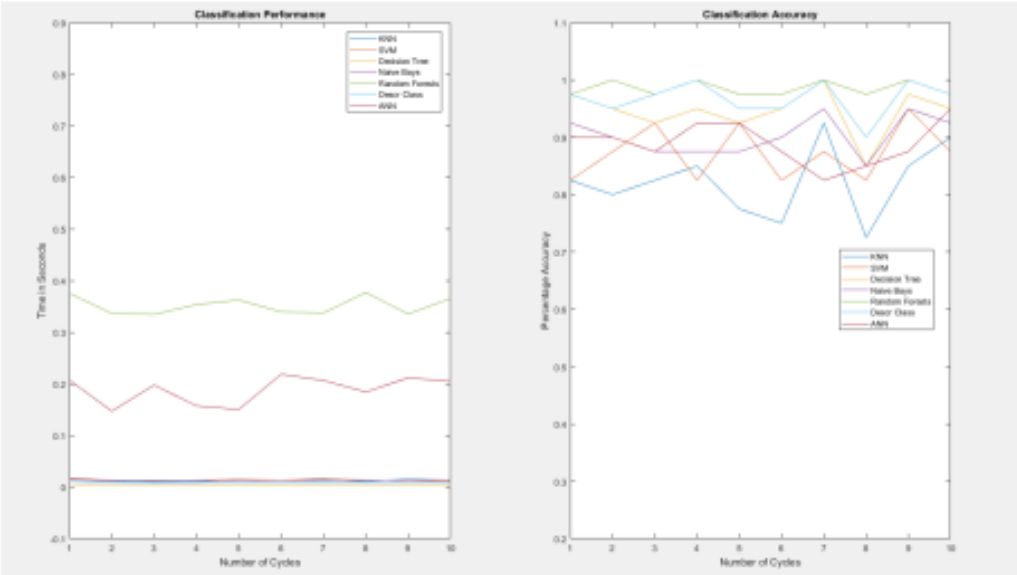
- Accuracy and performance comparison of feature-based object classification with and without optimisation
- Modality: Visual spectrum
- Number of features: 26
- Number of class labels: 2 to 32
- Number of cycles: 10
- Classification Algorithms: SVM, DT, NB, RF, DC, KNN and ANN

Results and Discussion

A quick assessment was applied to verify usage of optimisation Enahnaced-GA over seven commonly used classification algorithms. The Enhanced-GA was applied by varying a count of class labels from 2 to 32. The results show steady improvements in performance and accuracy as the number of class labels were increased. Alternately it can be concluded that, for a unimodal classification application, the optimisation process is more efficient for a substantial search space.

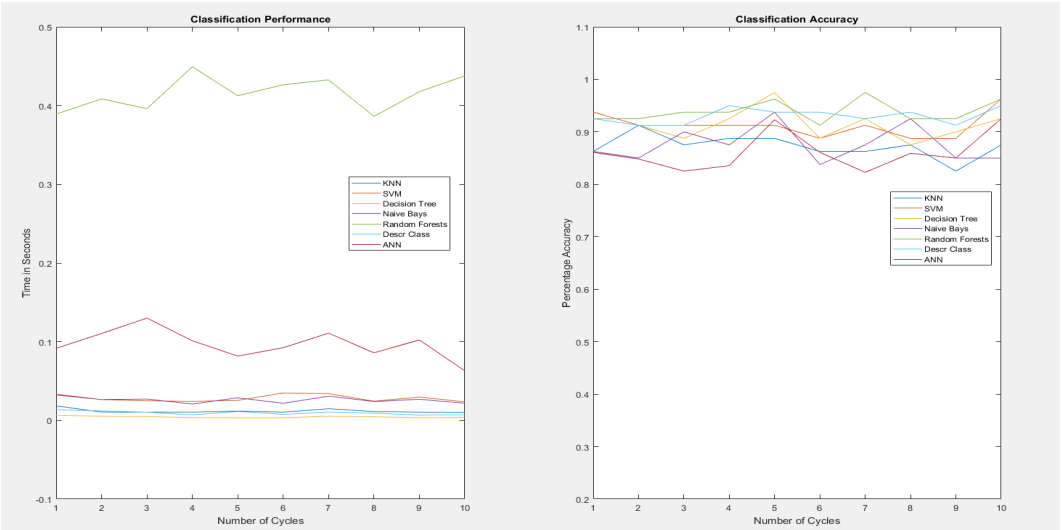


(a) Without Applying Optmization

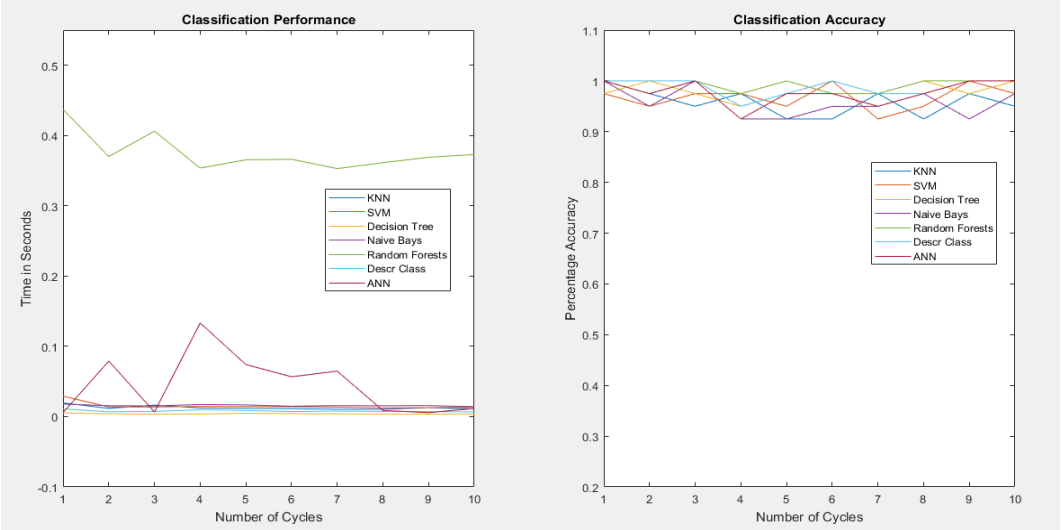


(b) After Applying Optmization

Figure 37: Classification Results with 2 Output Labels

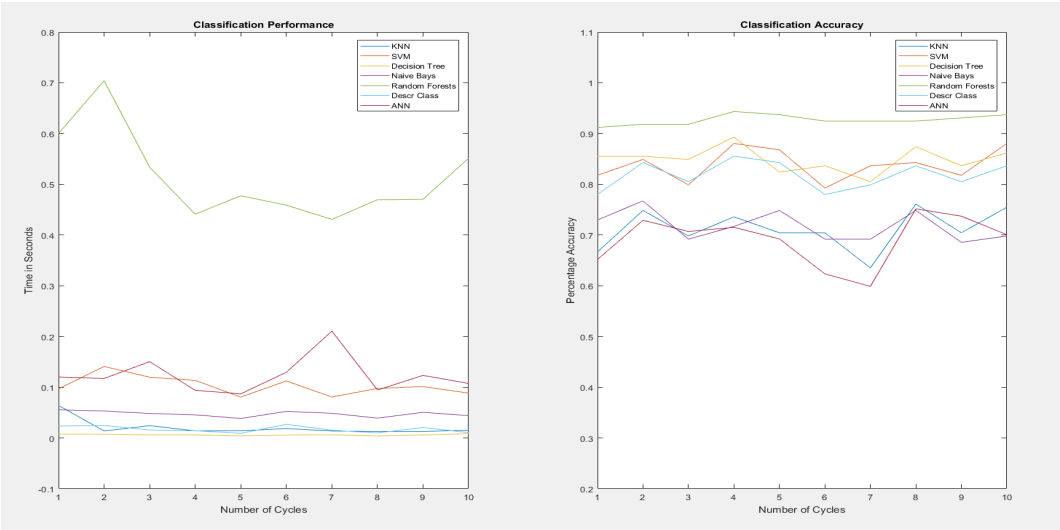


(a) Without Optimisation

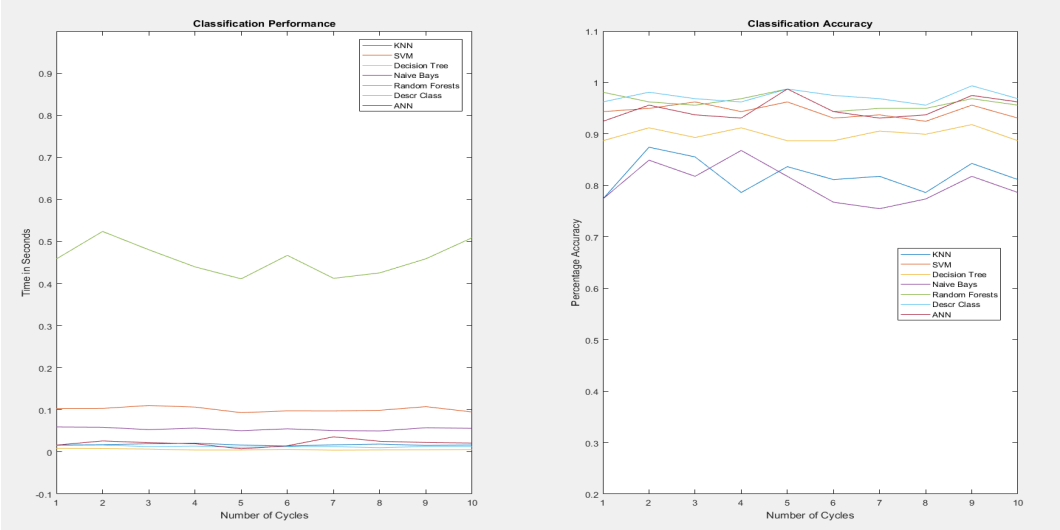


(b) With Optimisation

Figure 38: Classification Results with 4 Output Labels

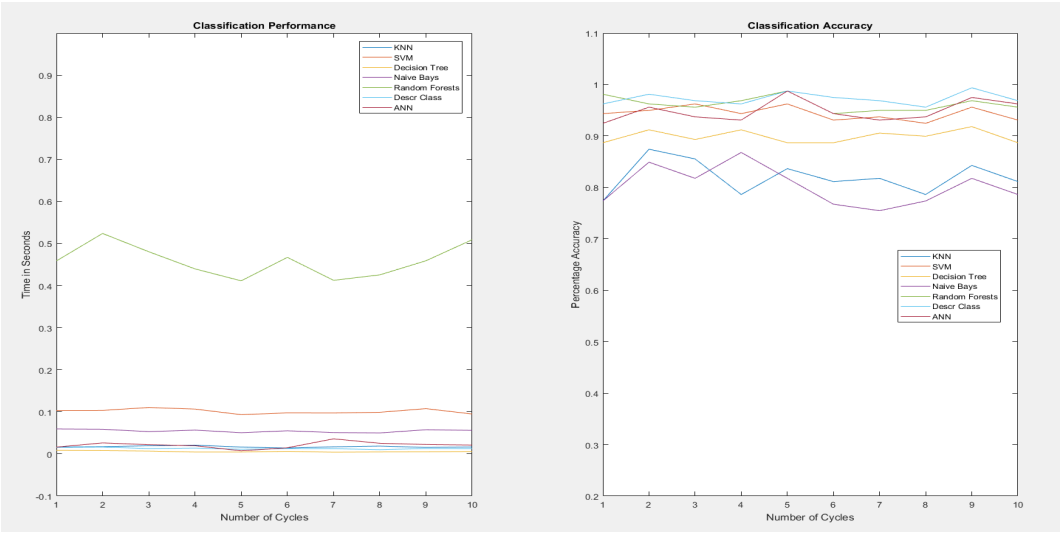


(a) Without Optimisation

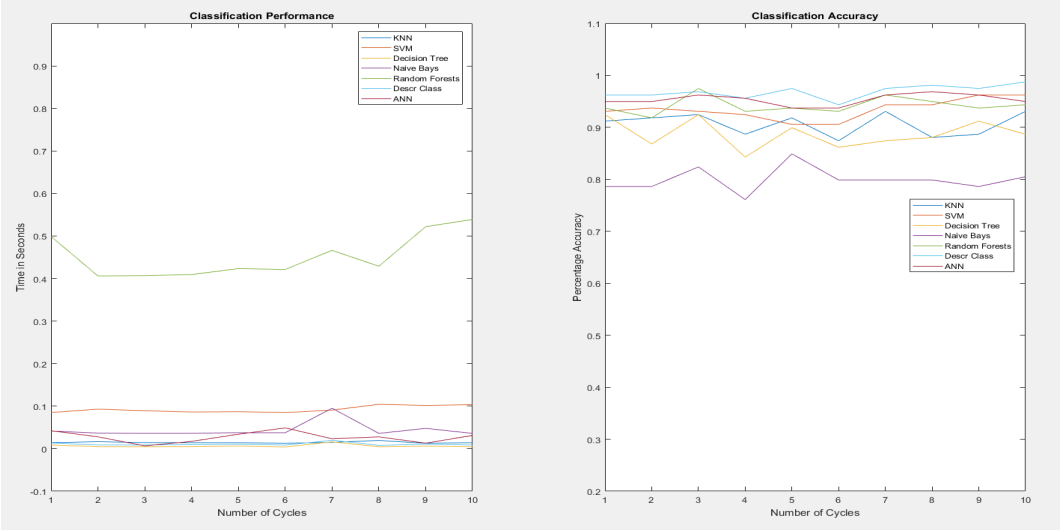


(b) With Optimisation

Figure 39: Classification Results with 8 Output Labels

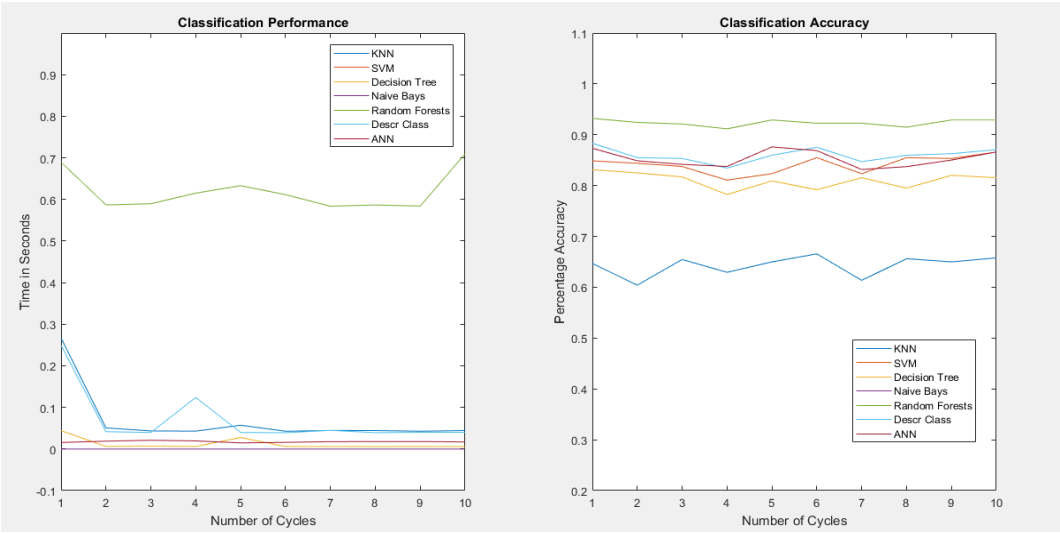


(a) Without Optmization

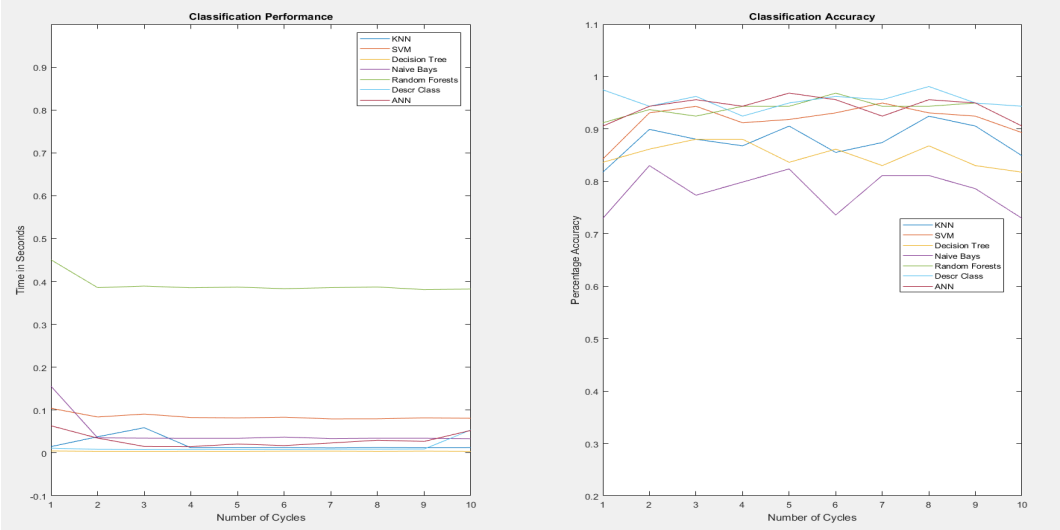


(b) With Optmization

Figure 40: Classification Results with 16 Output Labels



(a) Without Optmization



(b) With Optmization

Figure 41: Classification Results with 32 Output Labels

8.2 HyMuDS Datasamples

Figure 42: Example of Multimodal Data Samples Acquired using HyMuDS

8.3 TNO Dataset Class Library

Table 15: TNO Dataset Class Library

ID	Class	Description	Total mutimodal image Sets	Modeled Images
CL1	Men	Presence of human	3	594
CL2	Bush	Trees and bushes in a scene	3	594
CL3	APVehicle	Military vehicle	10	1980
CL4	Chopper	Helicopter	1	198
CL5	Bench	Sitting bench in a scene	1	132
CL6	Lake	Water body in pictures	1	132
CL7	Path	Walkway in bush	1	198
CL8	Tank	Military tank	1	198
CL9	Wire	Barbed Wire	2	396
CL10	House	Different types of houses	27	5346
CL11	Jeep	Automotive Vehicle	2	396
CL12	Wall	House Wall Structure	1	198
CL13	Sandpath	Walking Path in Sand	1	198
CL14	Balls	Circular Ball like Structure	1	198

8.4 Post-Thesis Research Paper

8.4.1 Towards an IOT Based System for Detection and Monitoring of Microplastics in Aquatic Environments

Conference: 22nd IEEE International Conference on Intelligent Engineering Systems 2018

Authors: Zenon Chaczko, Anup Kale, José Juan Santana-Rodríguez and Carmen Paz Suárez-Araujo

Abstract: The recent event of rescue mission involved in the Thailand soccer team called 'Wild Boar' has attracted our attention due to one of many challenges involved in the process. Apart from complex terrains, slippery paths and narrow size of gaps another essential aspect was the visibility constraint for the rescuers. Due to the kilometres of length without any natural light and presence of the muddy water, the rescue mission faced enormous challenges. At one stage there was a thought by rescue team to use crawling robots for guidance, but till date, it's not very sure whether such attempt was made. This article presents an overview of the possible usage of multimodal sensory apparatus equipped on a robotic vehicle which uses reinforcement learning and other neural networks for observing and learning intricate pathways involved in search and rescue operations. Here, we contemplate on using three on the shelf technologies and then integrate them with an application-specific configuration. Most commonly used mobile robots including crawlers, swimmers, climbers and drones were overviewed for the proposed usage. Appropriate instrumentation was as well overviewed with the proposed robots to capture images, signals and environmental parameters to create a most accurate scenario of the affected site. Finally, computational intelligence is equipped for data acquisition, pre-processing, transmission and classification operations. The proposed computational model involves the use of reinforcement learning strategy

to learn and store the hidden discoveries. Finally, few case studies are discussed on the past usage of similar technologies with recommending future improvements using the most recent techniques.

Keywords: Microplastics, IoT, In-Situ Sensing, Multimodality.

8.5 Future Research Paper

8.5.1 Multimodal Detection using Mobile Robotics for Rescue Missions a Preview

Abstract: The recent event of rescue mission involved in the Thailand soccer team called 'Wild Boar' has attracted our attention due to one of many challenges involved in the process. Apart from complex terrains, slippery paths and narrow size of gaps, another crucial aspect was the visibility constraint for the rescuers. Due to the kilometres of length without any natural light and presence of the muddy water, the rescue mission faced enormous difficulties from a visibility perspective. At one stage there was a thought by rescue team to use crawling robots for guidance, but till date, it's not very sure whether such attempt was made. This article presents an overview of the possible usage of multimodal sensory apparatus equipped on a robotic vehicle which uses reinforcement learning and other neural networks for observing and learning intricate pathways involved in search and rescue operations. Here, we contemplate on using three on the shelf technologies and then integrate them with an application-specific configuration. Most commonly used mobile robots including crawlers, swimmers, climbers and drones were overviewed for the proposed usage. Appropriate instrumentation was as well overviewed with the proposed robots to capture images, signals and environmental parameters to create a most accurate scenario of the affected site. Finally, computational intelligence is equipped for data acquisition, preprocessing, transmission and classification operations. The proposed computational model involves the use of reinforcement learning strategy to learn and store the hidden discoveries. Finally, few two studies are discussed on the past usage of similar technologies with recommending future improvements using the most recent techniques.

Keywords: Mobile Robotics, Multimodality, Reinforcement Learning

8.6 Research Contributions in Non Proceedings

1. Kale, A.V., Chaczko, Z. (2011), Multimodal Approach for Object Detection Dynamic and Uncertain Environments, University of Technology: Faculty of Engineering and IT Research Showcase (UTS FEIT Showcase 2011), 15 June 2011

2. Kale, A.V., Chaczko, Z. (2012), Environment Specific Modality Selection and Feature Subset optimisation, University of Technology: Faculty of Engineering and IT Research Showcase (UTS FEIT Showcase 2012), 13 June 2012

3. Kale, A.V., Chaczko, Z. (2014), 'evoMAP' an Evolutionary Multimodal Approach for Water Surface Pollution Monitoring, University of Technology: Faculty of Engineering and IT Research Showcase (UTS FEIT Showcase 2014), 4 September 2014