Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Master Thesis Proposal

Multimodal Deep Anomaly Detection for Robot Pouring Task

Adithya Narasimhaiah Sathish

Supervised by

Prof. Dr. Paul G. Plöger
Prof. Dr. Nico Hochgeschwender
M.Sc. Santosh Thoduka

June 26, 2023

1 Introduction

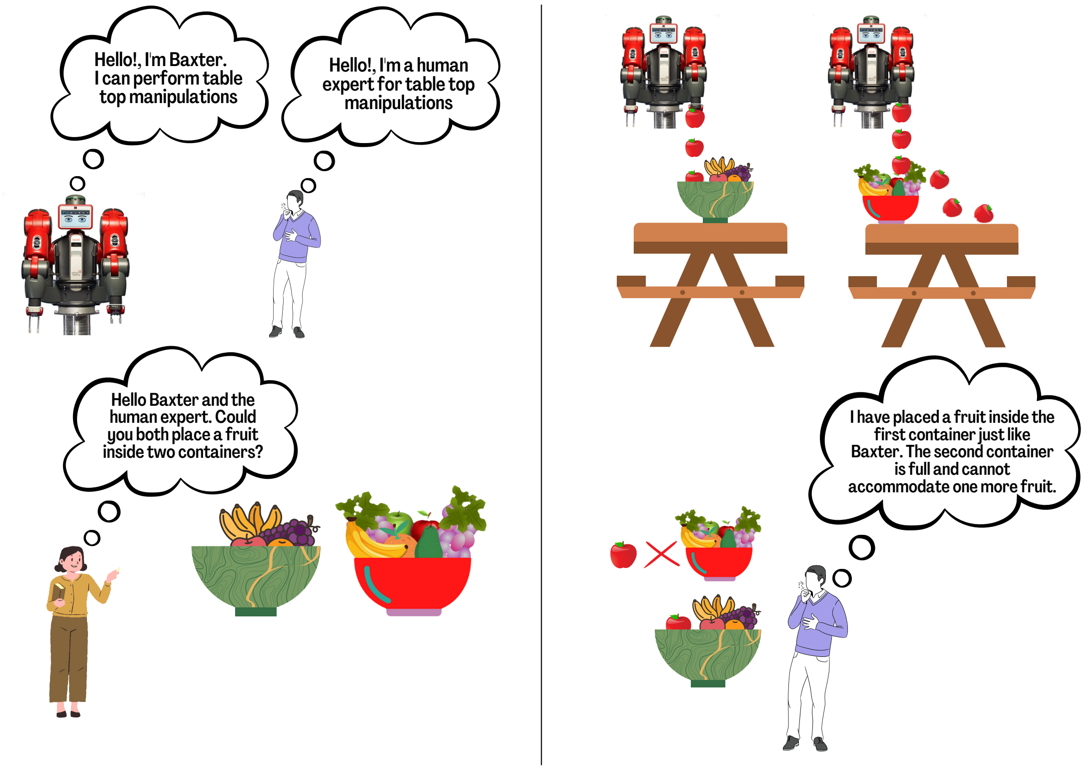


Figure 1: *A robot and a human expert are instructed to complete two tasks. The robot succeeds in the first task, but fails to complete the second task. Whereas the human expert succeeds in completing both tasks. Image copyright of Baxter robot belongs to Rethink Robotics.*

The robot illustrated in Figure 1 is responsible for placing a fruit in two containers filled with various kinds of fruit. The first container is partially filled and has enough space to accommodate more fruits whereas the second container is almost full. The robot picks up a fruit and places it inside the first container. The robot exhibits a behavior to successfully complete the task and the behavior is considered a *normal behavior*. The robot drops the fruit outside of the second container and the behavior of the robot fails to fulfill the objective of the task. The behaviour of the robot is significantly different from normal behaviour and is considered *anomalous behaviour*.

In general, any behaviour of the robot that results in operator overload and cognitive burden is an anomalous behaviour [9]. The robot failing to pick a fruit or dropping it in the middle of execution are some examples of anomalous behavior. The anomalous behavior of a robot is reflected as an unusual pattern in the information provided by its onboard sensors and these unusual patterns in sensor data are known as *Anomalies*.

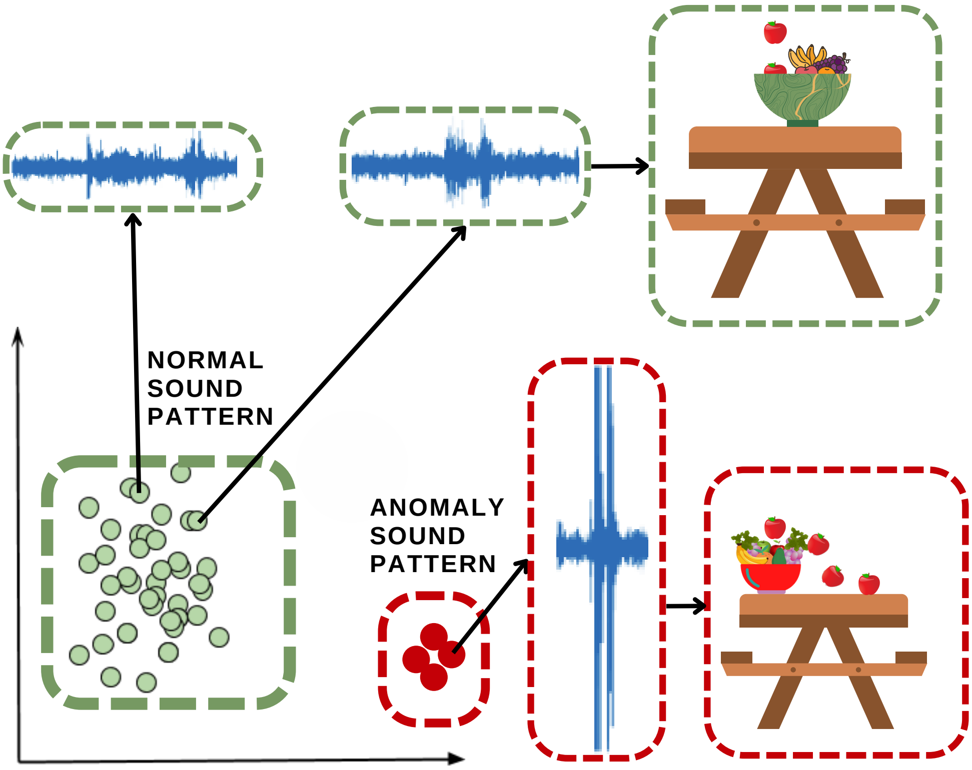


Figure 2: *The plot is a 2D representation of microphone-recorded sound patterns from a robot in a quiet environment. Green data instances are the sound generated when the robot successfully places a fruit within a container. The red dot represents anomalous sound patterns generated when a fruit falls from the container[2]*

Consider, for example, the fruit placement activity described in Figure 1. The microphone on the robot can record audio in the surrounding environment. The sound of a fruit falling from the container is very different from the sound of a robot putting a fruit in a container. Anomalous sound patterns is an indication of anomalous robot behaviour.

Placing a fruit in a completely filled container by the robot remains incomplete and for this reason, the robot must identify anomalies [2]. A mathematical formula or model is needed to identify anomalies in sensor data and it is known as *Anomaly Detection* [19].

This research focuses on detecting anomalies in a task performed by a service robot. In particular, this research focuses on anomalies related to the indoor robotic pouring task for granular substance.

1.1 Motivation

Pouring is a physical activity in which materials (solids/liquids) are transferred from one container to another [15]. A bartender transferring your favorite drink from a bottle to a mocktail glass or a foundry setup transferring molten metal from bucket to moulds are a few examples of pouring task as illustrated in Figure 3.

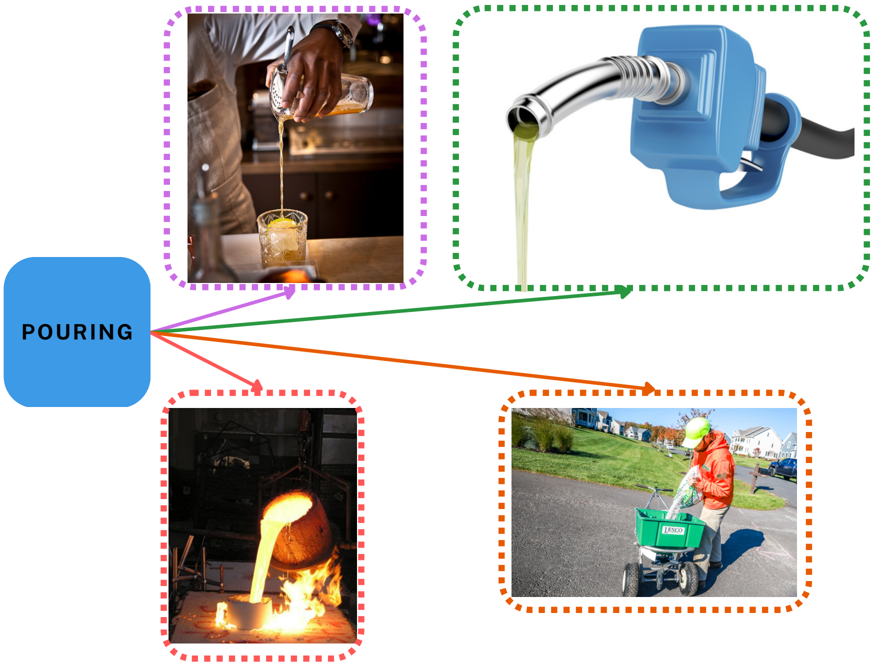


Figure 3: *Image of molten metal pouring is taken from. Image of petrol pouring is taken from Gastro*

Pouring activities, such as serving coffee at a social gathering, are tedious, repetitive and sometimes laborious. Pouring molten metal from a large vessel to

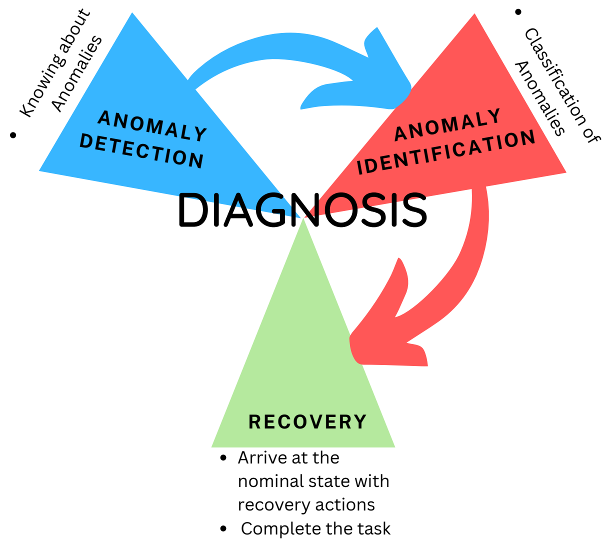


Figure 4: *Diagnostic is a three-step process. Depending upon the anomalies, the robot performs recovery actions to accomplish the task [2]*

moulds in a foundry is a hazardous task for a human expert. Traditionally, all repetitive, tedious and hazardous tasks for human experts are automated by robots. The *Adam* robot developed by *Rich Tech Robotics* is employed as a bartender and the *Kawasaki* robot is employed for pouring hot molten aluminum into foundry moulds.

As with human experts and machinery, robots also suffer from anomalies due to various sources of uncertainty [2]. Uncertainties may be due to human or power intervention during the execution of a task or due to physical damage to the on-board sensory system [1]. As a consequence, robots exhibit anomalous behaviour, produces undesired results (failure) and innocent robots are not even aware of anomalies.

A *Diagnostic* procedure is required to detect and identify the anomalies [2] and the procedure is made up of three steps illustrated in Figure 4. Let us understand the concept of diagnosis with robotic pouring as an example. A robot behavior that pours liquid into a glass (empty or partially filled) until the brim

[1]The present research assumes that robotic sensors are free from physical damage and plans to replace a damaged sensor for data collection and experimentation. Thus, we eliminate the occurrence of anomalies from damaged sensors.

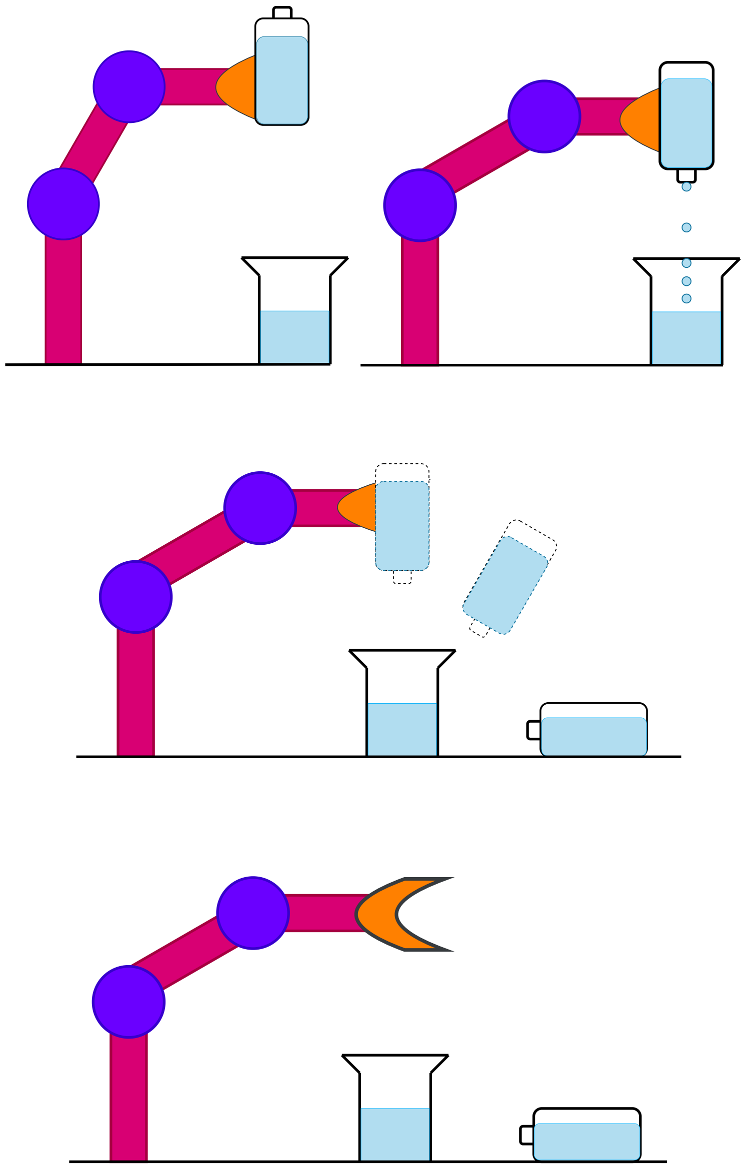


Figure 5: *The robot begins to pour liquid from bottle to beaker. Due to improper calculation of gripping force, robot drops the pouring container in the middle of execution.*

level is considered normal behavior. Any robot behavior that produces different results from normal behavior is considered anomalous behavior. The robot exhibits anomalous behavior by dropping the pouring container as shown in Figure 5. The first step of the *Diagnostic* procedure, *Anomaly Detection*, detects an anomaly in the reading of the torque sensor. Anomalies are classified in the middle step of *Anomaly Identification* [2]. The anomaly is identified as a slip in the robot gripper. And, in the final stage of *Recovery*, the robot uses recovery actions to arrive at the nominal state to finish the task [2]. In this case, the robot is instructed to grab the pouring container and to resume pouring till the liquid reaches the brim level.

The scope of this research is the first step in the diagnostic procedure, ie., anomaly detection for the robot pouring task in an indoor environment. We formulate anomaly detection as a machine learning problem in which a model is trained in semi-supervised fashion using multimodal data obtained from on-board sensors of a service robot. We take advantage of the recent success of deep learning models for anomaly detection and present a comparative study of the various architectures. We also highlight the limitations of an existing collection of robot pouring dataset and present a new dataset.

1.2 Problem Statement

The current research proposes to create a dataset for the robot pouring task for granular media. The dataset will be used as a testbed for comparative analysis of multimodal deep anomaly detection methods. We present a literature review of the robotic pouring dataset, multimodal sensor fusion schemes and anomaly detection methods. We formulate multimodal deep anomaly detection as machine learning problem with deep learning models trained on proposed dataset in semi-supervised fashion. As part of the research, the following research questions are answered.

- **RQ1** What are the deficits in the existing collection of robotic pouring datasets and why do we need a new dataset?
- **RQ2** Which sensor fusion scheme works best for multimodal deep anomaly detection?
  - RQ2.1 Which sensor modalities should be fused for deep anomaly detection?
  - RQ2.2 When to fuse different types of sensor modalities for deep anomaly detection?
  - RQ2.3 How to fuse different types of sensor modalities for deep anomaly detection?
- **RQ3** What influences do each sensor modality have on multimodal deep anomaly detection?
- **RQ4** What is the performance of the selected deep learning architectures on the proposed dataset [2]?
- **RQ5** What are the complex anomaly detection issues that selected deep learning architectures address?

[2] Selected architectures/methods are listed in Section 2.3.2

1.3 Challenges and Difficulty

Problem complexities and challenges in anomaly detection are as follows:

1.3.1 Problem Complexities

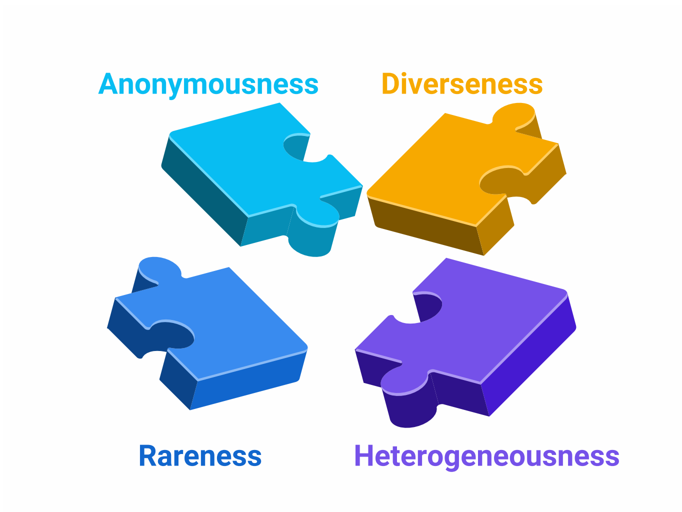


Figure 6: *Problem complexities in Anomaly Detection [16]*

- **Anonymousness:** Anomalies remain anonymous until they occur. All space anomalies or the nuclear explosion at Hiroshima are best known for anonymous nature.
- **Heterogeneousness:** Anomalies are heterogeneous by nature. The anomalies are quite different from one another. Mass outbreaks or terrorist attacks on public surveillance are anomalies which are very different from each other.
- **Rareness:** Anomalies are rare in nature. Physical damage to the robot or the environment is rare and just as hard to replicate.

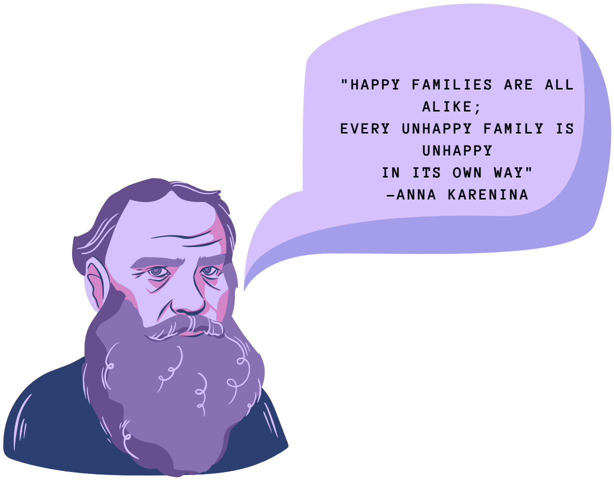


Figure 7: *Normal data instances are like happy families as they are alike. Anomaly data instances from like unhappy families, they are unique, rare and are difficult to recognize. The famous quote is taken from fictional novel* ***Anna Karenina*** *written by a famous Russian author* ***Leo Tolstoy****.*

1.3.2 Challenges

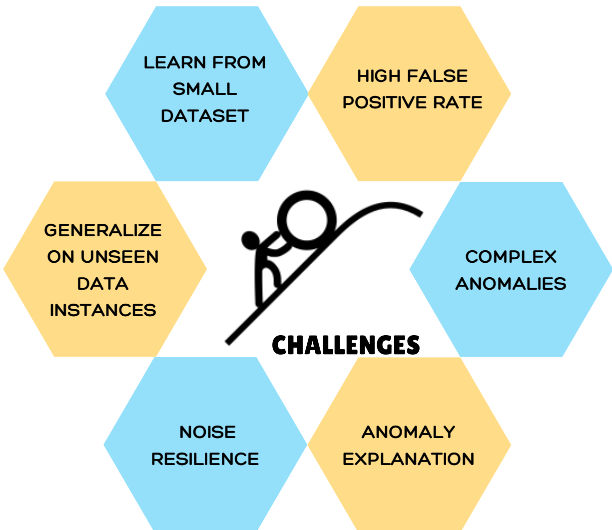
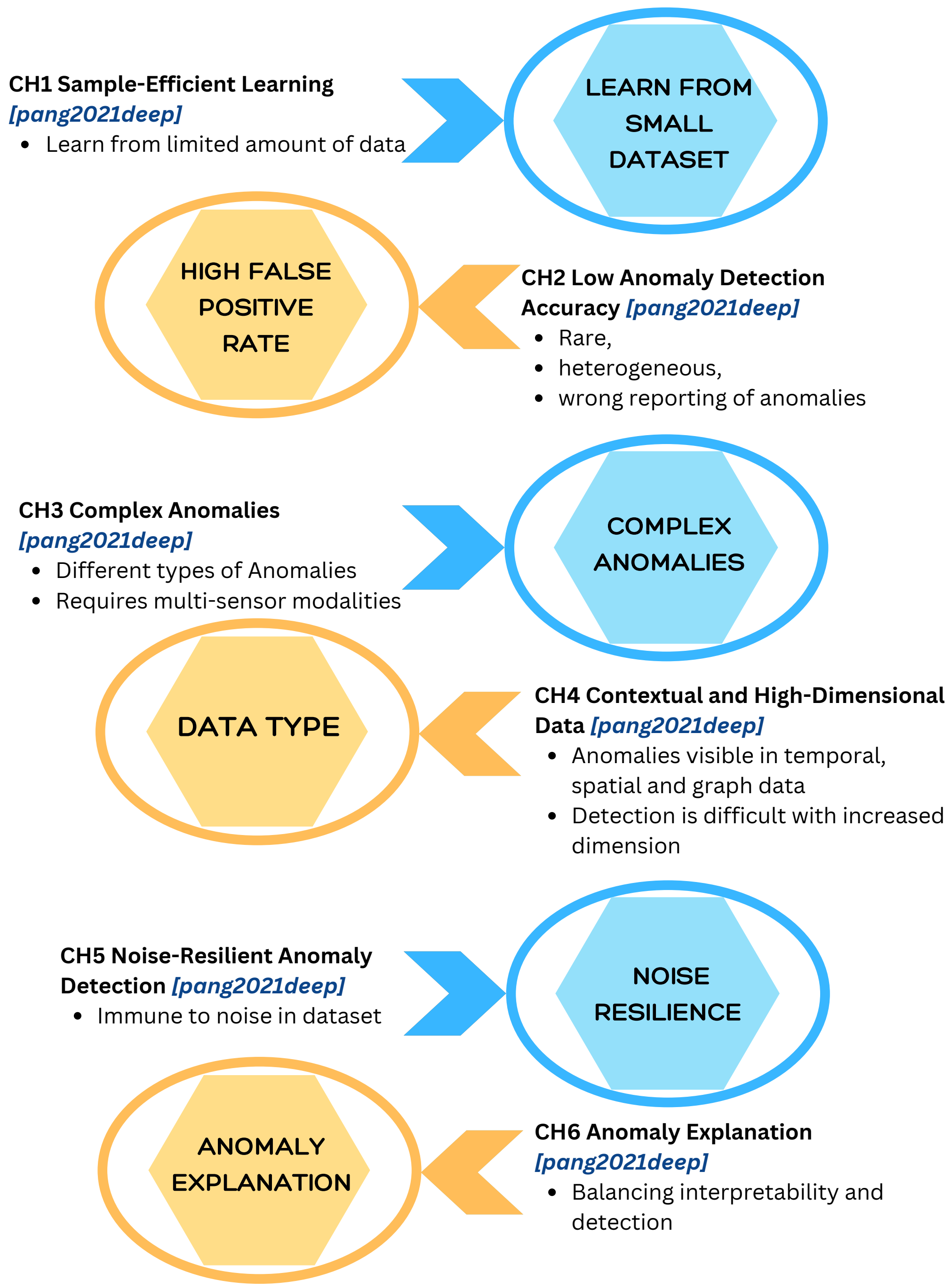


Figure 8: *Challenges address by deep anomaly detection methods [16]*


CH1 Sample-Efficient Learning
[pang2021deep]
• Learn from limited amount of data
LEARN FROM SMALL DATASET
HIGH FALSE POSITIVE RATE
CH2 Low Anomaly Detection Accuracy [pang2021deep]
• Rare,
• heterogeneous,
• wrong reporting of anomalies
CH3 Complex Anomalies
[pang2021deep]
• Different types of Anomalies
• Requires multi-sensor modalities
COMPLEX ANOMALIES
DATA TYPE
CH4 Contextual and High-Dimensional Data [pang2021deep]
• Anomalies visible in temporal, spatial and graph data
• Detection is difficult with increased dimension
CH5 Noise-Resilient Anomaly Detection [pang2021deep]
• Immune to noise in dataset
NOISE RESILIENCE
ANOMALY EXPLANATION
CH6 Anomaly Explanation
[pang2021deep]
• Balancing interpretability and detection

2 Related Works

In this section, we present a brief literature review on robotic pouring datasets, multimodal sensor fusion schemes and various aspects of anomaly detection methods.

2.1 Robotic Pouring Datasets

Most of the pouring data sets are related to the liquid material as opposed to granular pouring. A portion of the significant pouring dataset for the granular substance are. We are proposing a new pouring dataset due to limitations in existing robotic granular pouring datasets.

- **Highly Instrumented Environment:** All the datasets are collected in a highly instrumented environment [3]. For instance, the microphone for recording the pouring process is located right next to the receiving container. The sound patterns differ considerably from the microphone positioned next to the receiver to the microphone on the robot platform. Model trained on former audio modality may present a poor generalisation service robot or robot with microphone on its platform.
- **Performed in a Quiet Environment:** Datasets are collected in a calm setting [13]. Again, the model may not generalize with service robots deployed in noisy environments.
- **No Human-Robot Interaction:** All the datasets are collected using statically positioned robot manipulators [22]. The generalization aspect of models trained on such datasets is questionable when they are deployed on service robots.
- **Human Demonstration:** In some literature, the pouring activity is performed by human expert [3]. Thus, haptic data, the most significant modality, will not be included in the dataset. What's more, demonstrations by human experts are different from robotic pouring. Therefore, we rule out the dependence of the human expert for the pouring task in the proposed dataset.

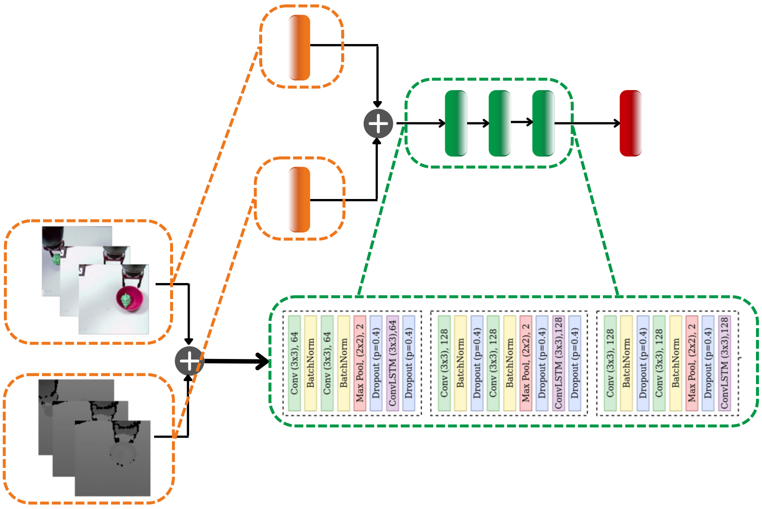


Figure 9: *Early fusion of RGB and depth images [11]*

- **Lack of Real World Datasets**: Only a handful of real world anomaly detection datasets are available [16]. And a very few of them are multimodal in nature with focus on robotics [17]. We offer a new data set that can be used for model training and other data sets in the existing collection can be used for testing or vice versa.

2.2 Multimodal Sensor Fusion Schemes

Based on the number of sensor modalities, anomaly detection methods can be classified into two categories as *uni-modal* and *multi-modal* methods. The main limitation of uni-modal methods lies in their reliance on single sensor modality [10]. Even human experts rely on more than one sensing organ to detect failures.

Given the importance of utilizing multi sensor modalities, various sensor fusion schemes are available [6]. Early fusion, intermediate fusion, late fusion answers when to fuse sensor modalities. FINO-Net architecture [11] shown in Figure 3.4 uses early fusion for combining RGB and depth images. Mixture of Experts and Ensemble are different ways to fuse modalities [6].

2.3 Anomaly Detection Methods: Aspects

Anomaly Detection methods can generally be classified into two categories, model-based methods and model-free methods. The initial set of work comes under the first category of model-based methods. For example, [7] proposed to use *Kalman Filter* to four wheel and three sensor errors. The major drawback of model based methods is the knowledge of process and measurement models. The assumption of a prior distribution as *Gaussian white noise* is an additional limitation. Model free methods have become very popular as they mitigate the limitations of model based methods.

Traditional algorithms offer sub-optimal performance on image data [4] and the challenges listed in the section above are addressed through deep learning. Below, we discuss about potential deep learning architectures for the comparative study along with the model training method.

2.3.1 Training Methods

We have four training methods for anomaly detection as shown in Figure 10.

- **Supervised Method:** The success of supervised training methods depends on the availability of balanced dataset. As anomalies are rare, it is not possible to collect enough anomaly data instances relative to normal data instances. Therefore, the resulting dataset is highly imbalanced and the models tend to bias the majority class.

- **Semi-Supervised Method:** Models are trained on normal data instances with low reconstruction error. During testing, data instances with high reconstruction error are identified as anomalies. Preparation of train dataset is relatively easier as it only contains normal data instances. Additional effort is required for preparing test dataset with human induced anomaly data instances. Semi-Supervised is the most preferred training method [16].

- **Weakly-Supervised Method:** Models are trained on dataset with normal data instances and few local anomaly data instances. Local anomaly data instances belong to anomaly class but are very close to normal data instances. Hence, this training method is preferred in tasks with local anomalies [21].

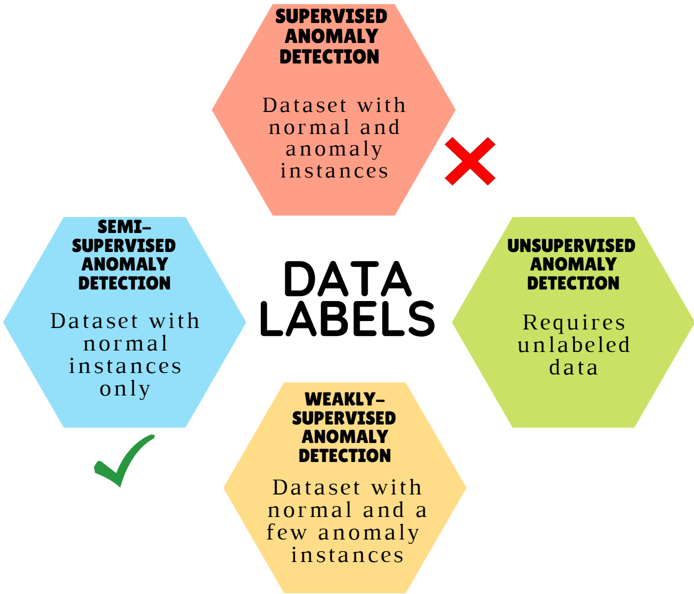


Figure 10: *Training paradigms for anomaly detection methods [5]*

- **Un-supervised Method:** There is no need for training data, but the main limitation of the method is that it is based on a prior assumption about data distribution.

2.3.2 Deep Learning Models: Selection of Methods

Human experts rely on vision and force-torque sensor modalities to accomplish pouring task. As vision is among the primary sensor modalities, video anomaly detection methods (VAD-methods) are selected for the comparative study. Among the plethora of VAD-methods, following two methods are selected for the comparative study.

Probabilistic U-Net: [20] proposed a knowledge-based method for action execution monitoring along with task-specific robotic dataset. The proposed method, a hybrid of model-based and data-driven technique, allows incorporation of domain-knowledge with no dependency on large datasets.

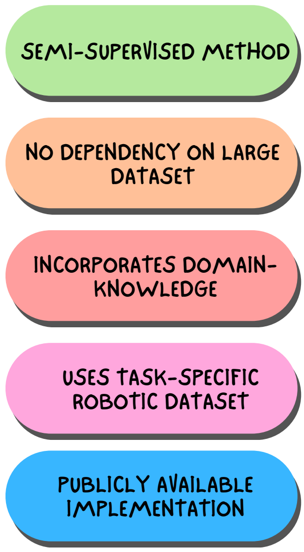


Figure 11: *Keypoints to select the action monitoring method proposed by [20] for the comparative study*

The method uses a combination of *Conditional Variational Autoencoder* and *U-Net* known as *Probabilistic U-Net* to detect motion anomalies. *Probabilistic U-Net* takes multiple ground truth inputs during training and is relatively easier to train with less number of training parameters. The method was compared with two other methods and exhibited superior performance.

HF^2-VAD: [14] proposed a new framework HF^2-*VAD* for video anomaly detection. The framework is a hybrid method based on the reconstruction based

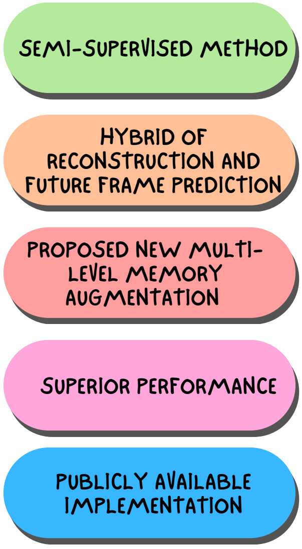


Figure 12: *Keypoints to select VAD method proposed by [14] for the comparative study*

method and the frame prediction method. The framework address the current limitations of memory-based auto-encoders for anomaly detection.

The framework reconstructed normal data instances very well and gave poor results on anomaly data instances during testing. The framework was compared with eight other methods on publicly available video anomaly detection datasets and exhibited superior performance.

3 Project Plan

3.1 Work Packages

Literature Review

- Literature review on robot pouring datasets
- Literature review on multimodal sensor fusion schemes
- Literature review on anomaly detection methods

WORK PACKAGE 1

Dataset

- Collect raw sensor reading from Toyota HSR robot for successful and anomalous granular pouring task
- Data Pre-processing
- Prepare train dataset with successful robot pouring tasks
- Prepare test dataset with human induced anomalous robot pouring data instances

WORK PACKAGE 2

Methodology

- Identify anomaly detection methods for the comparative study
- Design of experiments
- Selection of evaluation metrics
- Formulate a hypothesis based on knowledge of literature review

WORK PACKAGE 3

Implementation

- Implement late and mixture of experts sensor fusion schemes
- Integrate sensor fusion schemes with existing implementation/s
- Implement and integrate sensor fusion schemes with other deep anomaly detection methods

WORK PACKAGE 4

Experimentation

- Analyse results obtained in WP4
- Comparative study of deep learning models on proposed dataset
- Ablation study to know the importance of each sensor modality for anomaly detection

WORK PACKAGE 5

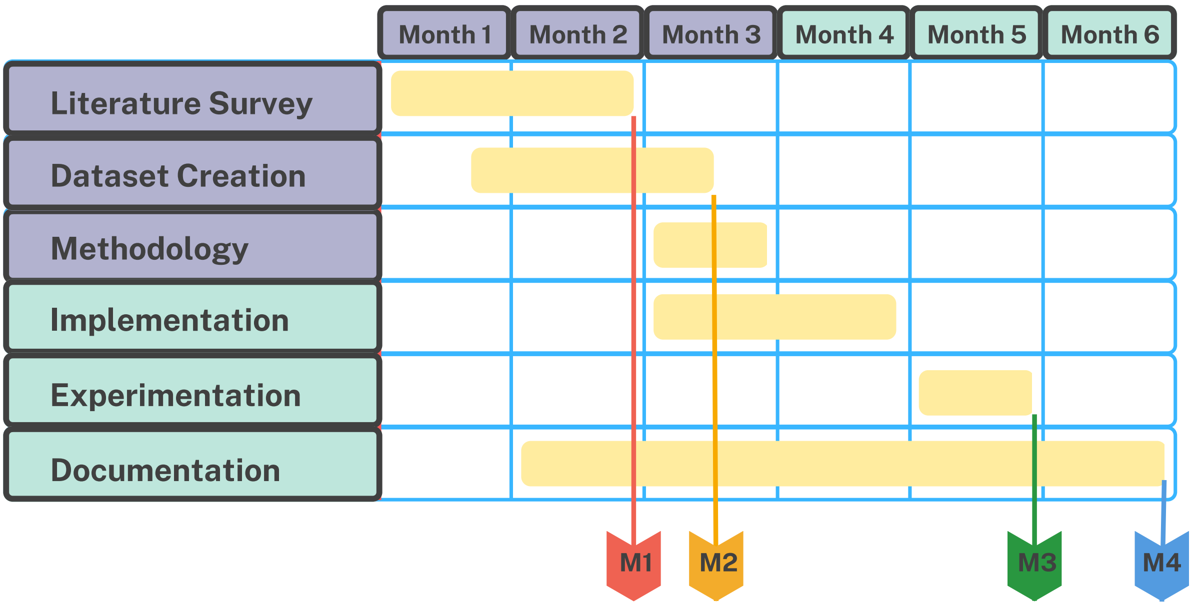
Documentation

- Write a comprehensive report describing
  - the results of the literature review,
  - the dataset developed,
  - the methodology used,
  - the experiments performed and
  - the conclusions drawn from experimentation

WORK PACKAGE 6

3.2 Milestones

3.3 Project Schedule



3.4 Deliverables

MINIMUM VIABLE

- Literature survey on robotic pouring datasets, multimodal sensor fusion schemes and deep anomaly detection methods
- Collection of robotic pouring dataset
- Implementation of (2+) sensor fusion schemes
- Train, validate and test existing implementation on collected dataset

EXPECTED

- Ablation study to investigate the impact/contribution of sensor modalities for anomaly detection in robotic pouring task
- Comparative study of (2+) deep anomaly detection methods

MAXIMUM

- Deployment of the chosen method for identifying anomalies following the execution of tasks on a robot.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian Conference on Computer Vision*, pages 622–637. Springer, 2018.

[2] Dogan Altan and Sanem Sariel. Clue-ai: A convolutional three-stream anomaly identification framework for robot manipulation. *arXiv preprint arXiv:2203.08746*, 2022.

[3] Alexis Burns, Siyuan Xiang, Daewon Lee, Larry Jackel, Shuran Song, and Volkan Isler. Look and listen: A multi-sensory pouring network and dataset for granular media from human demonstrations. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2519–2524. IEEE, 2022.

[4] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[5] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[6] Di Feng, Christian Haase-Schütz, Lars Rosenbaum, Heinz Hertlein, Claudius Glaeser, Fabian Timm, Werner Wiesbeck, and Klaus Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2020.

[7] Puneet Goel, Göksel Dedeoglu, Stergios I Roumeliotis, and Gaurav S Sukhatme. Fault detection and identification in a mobile robot using multiple model estimation and neural network. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 3, pages 2302–2309. IEEE, 2000.

[8] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly

detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1705–1714, 2019.

[9] Ritwik Gupta, Zachary T Kurtz, Sebastian Scherer, and Jonathon M Smereka. Open problems in robotic anomaly detection. *arXiv preprint arXiv:1809.03565*, 2018.

[10] Arda Inceoglu, Gökhan Ince, Yusuf Yaslan, and Sanem Sariel. Failure detection using proprioceptive, auditory and visual modalities. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2491–2496. IEEE, 2018.

[11] Arda Inceoglu, Eren Erdal Aksoy, Abdullah Cihan Ak, and Sanem Sariel. Fino-net: A deep multimodal sensor fusion framework for manipulation failure detection. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6841–6847. IEEE, 2021.

[12] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv:2107.14795*, 2021.

[13] Hongzhuo Liang, Shuang Li, Xiaojian Ma, Norman Hendrich, Timo Gerkmann, Fuchun Sun, and Jianwei Zhang. Making sense of audio vibration for liquid height estimation in robotic pouring. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5333–5339. IEEE, 2019.

[14] Zhian Liu, Yongwei Nie, Chengjiang Long, Qing Zhang, and Guiqing Li. A hybrid video anomaly detection framework via memory-augmented flow reconstruction and flow-guided frame prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13588–13597, October 2021.

[15] Sebastian Neubert, Steffen Junginger, Thomas Roddelkopf, Simon-Johannes Burgdorf, Norbert Stoll, and Kerstin Thurow. Automated system for pouring

and filtration tasks in laboratory applications. *Chemie Ingenieur Technik*, 94 (4):530–541, 2022.

[16] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021.

[17] Jing Ren, Feng Xia, Yemeng Liu, and Ivan Lee. Deep video anomaly detection: Opportunities and challenges. In *2021 International Conference on Data Mining Workshops (ICDMW)*, pages 959–966. IEEE, 2021.

[18] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. *arXiv preprint arXiv:1906.02694*, 2019.

[19] James P Theiler and D Michael Cai. Resampling approach for anomaly detection in multispectral images. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, volume 5093, pages 230–240. SPIE, 2003.

[20] Santosh Thoduka, Juergen Gall, and Paul G Plöger. Using visual anomaly detection for task execution monitoring. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4604–4610. IEEE, 2021.

[21] Ujjwal Verma, Radhika M Pai, et al. Contextual information based anomaly detection for a multi-scene uav aerial videos. *arXiv preprint arXiv:2203.15437*, 2022.

[22] Justin Wilson, Auston Sterling, and Ming C Lin. Analyzing liquid pouring sequences via audio-visual neural networks. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7702–7709. IEEE, 2019.