CHAPTER II LITERATURE REVIEW

2.1 Theoritical Basis

2.1.1 Understanding and History of Deep Learning Technology

Machine learning techniques and methods have evolved significantly since their inception in the 1980s. Some of the early techniques include: Decision Tree Learning (DTL): A method that uses a tree-like model to make decisions based on input data. Bayesian Learning (BL): A probabilistic approach that uses Bayes' theorem to make predictions. Support Vector Machine (SVM): A supervised learning algorithm that finds the best boundary to separate data points. Neural-Based Learning (NBL): A family of techniques inspired by the human brain's neural network, including Artificial Neural Network (ANN), Self Organizing Map (SOM), and others. Unsupervised Learning (UL): A method that finds patterns in data without labeled examples, such as k-Means Clustering (kMC) and Hierarchical Clustering (HC). Ensemble Learning (EL): A technique that combines multiple models to improve performance, including bagging, boosting, stacking, and classifier selection. These techniques are often referred to as Shallow Learning (SL) or Traditional Machine Learning (TML). In addition to SL, machine learning has also developed new techniques and methods that are more popular and reliable, such as: Deep Learning (DL): A subfield of machine learning that uses artificial neural networks to perform sophisticated computations on large amounts of data. Nature-inspired Learning (NIL): A family of techniques inspired by natural processes, including Evolutionary Algorithm (EA) and Swarm Intelligence (SI). Reinforcement Learning (RL): A method that learns by interacting with an environment and receiving feedback in the form of rewards or penalties. These techniques have revolutionized the field of machine learning and have been applied to various real-world applications, such as customer service, voice virtual assistants, eCommerce, text generation, fraud detection, computer vision, and self-driving vehicles . (Naaz Mir et al., n.d.)

DL's popularity has skyrocketed since becoming the first winner in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) image recognition competition in 2012. DL provides accuracy far beyond SL. Until that day, DL became a state-of-the-art AI technique, which in some cases, provided performance beyond human capabilities. DL is capable of handling a variety of data sets, especially very

large data sets. In the new DL technique, many methods, architectures, schemes and learning algorithms have been developed. The first generation of DL was a convolutional Neural Network (CNN), which was specifically used for image processing.(Kazemi et al., 2013)

Since 2012, the development has been very rapid. Many types of data: text, images, sequences, sound, video, and so on. Several DL methods that have been developed are: Capsule Network (CapsNet), Recurrent Neural Network (RNN), Long Short-term Memory (LSTM), Generative Deep Leaning (EDL), Deep Reinforcement Network (LTSM), Generative Adversarial Network (GAN) , Transformer Network (TN), Evolutionary Deep Learning (EDL), Deep Reinforcement Learning (DRL), Lifelong Learning (LL), and so on. (Suyanto, 2021)

The DL has a lot of strengths. DL does not require data engineers to manually engineer features. DL is capable of performing feature engineering automatically, for example the stacked auto encoder model which is explained in full and detail in the book Deep Learning: Machine Learning Modernization for Big Data. In general, DL is able to provide high accuracy when trained with large training data. The addition of training data usually provides a significant increase in model accuracy, which is far more valuable than the effort to add the data, DL is still able to provide increased accuracy. (Suyanto, 2019)(Universitas Atma Jaya Yogyakarta et al., n.d.)

However, DL also has a number of drawbacks. First, DL is usually less suitable for small data, with a relatively small number of rows (records or objects) and columns (features or dimensions). Second, DL requires data scientists to design and build model architectures manually. The large number of hyperparameters in the DL makes the design and implementation process of the DL architecture very difficult to do manually. Third, DL techniques and methods generally have low explainability, namely a relatively low level of ease of explaining. This means that the DL is less able to provide an explanation to the user why it outputs that way. Therefore, DL is often referred to as a black box which is only able to provide accurate output but is unable (or only slightly) to provide an explanation as to why it produces that output. (Suyanto, 2021)

2.1.2 Understanding Deep Learning for Anomaly Detection

Detection of anomalies is a field of research that active with wide applications in various industries, such as maintenance, fraud detection, healthcare, cybersecurity and industrial predictive. It involves identifying outliers, expected characteristics, unusual observations, and abnormal patterns that deviate from established normal condition and in a dataset/environment and system. In many app for anomaly detection applications, domain-specific knowledge is required to draw out timely actionable insights for decision making and risk mitigation. Anomaly detection examines specific data points and detects rare events that appear suspicious because they differ from existing behavioral patterns. (Yahya & Ullah, n.d.)

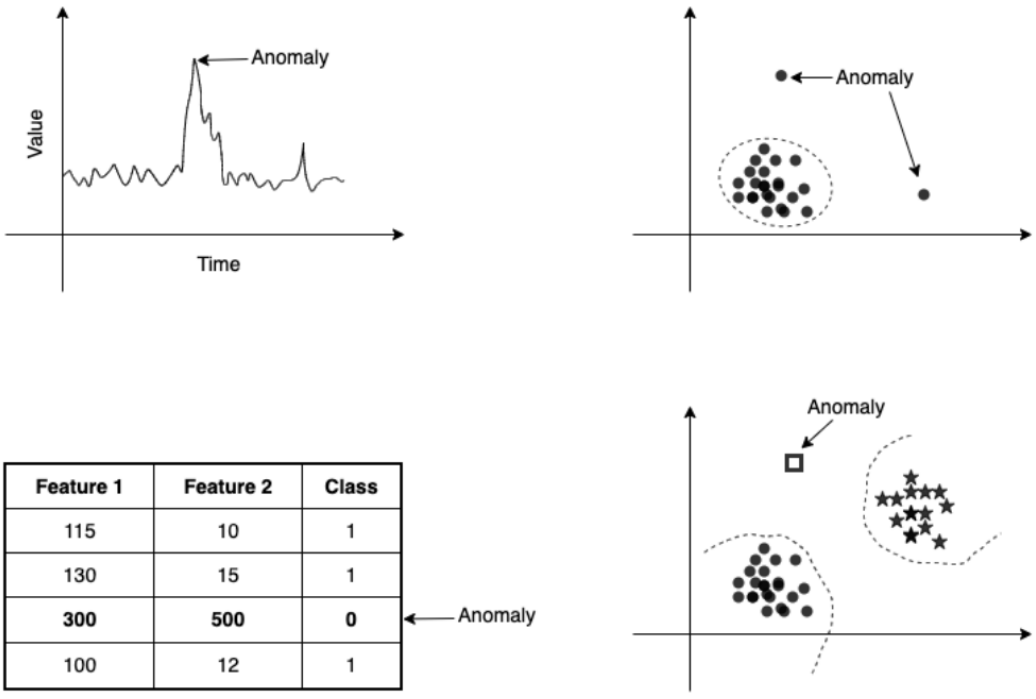
Anomaly detection provides a way to automatically detect dangerous outliers and protect data. Some industries that can benefit from anomaly detection are banking, retail, and cybersecurity, but every business should consider an anomaly detection solution. In some cases, hospitals that don't know an attack is coming can leverage predictive capabilities to write rules to prevent attacks, protect sensitive data, and their environment. For instance, the early detection of equipment performance degradation can prevent unplanned downtime, while the early discovery of disease threats can prevent a pandemic outbreak.(Yahya & Ullah, n.d.)

Promising results have been demonstrated by deep learning-based anomaly detection in address challenges related to the rare nature of anomalies, we need to model complex, high-dimensional data and identify new classes of anomalies.. The primary focus of anomaly detection is often on isolating undesirable data instances, such as product defects and safety risks, from the targeted domain. Other interests include improving model performance by removing problematic data or irrelevant outliers and hopefully identifying anomalies emerging from the data set for superior results that can be used in implementation. (Kecerdasan Buatan Kajian Konsep Dan Penerapan (Jarot Dian Susatyono, M.Kom.) (Z-Library), n.d.)

Khair says that they can exist within a static entity or a temporal context. Below we show a comparison of the various types of anomalies that exist (Khair et al., 2019).:

Anomalies often occur in data sets or those whose top points are outside the normal distribution limits. For example, when we make a purchase using a credit card that is expensive and out of habit, it is an anomaly. Collective anomalies occur when several related data records or sets of observations appear collectively and differ significantly from the rest of the data set. Contextual anomalies occur when viewed based on contextual attributes such as day and time. An example of a temporal contextual anomaly is a sudden increase in online orders outside of expected peak hours of shopping activity.(Saltz & Gupta, 2020)

Anomalies are characterized by at least one (univariate) or several (multivariate) attributes in the continuous, numeric, binary, categorical or binary data types. These attributes determine the dimensions of an anomaly, features, and characteristics. Common anomaly types are illustrated in Figure 1.1.



Feature 1	Feature 2	Class
115	10	1
130	15	1
300	**500**	**0**
100	12	1

Figure 1 Types of anomaly (Tereshchenko et al., 2015)

Defining Detecting anomalies is not an easy task because the boundaries between normal and abnormal behavior can be domain specific and subject to risk tolerance levels determined by the business or organization and industry. For example, if a patient's irregular heartbeat from data collected in an electrocardiogram (ECG) time series can pose a risk of cardiovascular disease, while stock price fluctuations can be considered normal based on market demand. On that basis,

there is basically no definition of an anomaly, and no general solution for which we can detect an anomaly. (Jung et al., 2021)

Unsupervised deep learning techniques using unlabeled datasets are widely used in security-related anomaly detection applications such as intrusion detection systems (IDS), web attack detection, video surveillance, and advanced persistent threats (APT) to detect zero-day attacks or unknown threats unpredictable on web security. There are two categories of IDS: host-based and network-based. Host-based IDS detects collective anomalies such as malicious applications, policy violations, and unauthorized access by analyzing sequential call traces at the operating system level. Network-based IDS analyzes high-dimensional network data to identify potential external attacks for unauthorized network access. Anomaly detection is critical in industries, such as finance, retail, and cybersecurity, but every business should consider an anomaly detection solution. Anomaly detection provides a way to automatically detect dangerous outliers and protect data.(Lee & Shin, 2019)

The targeting of web applications by cybercriminals has increased due to the ubiquity of data. Traditional signature-based techniques using static rules are no longer adequate for web attack protection, as the quality of rulesets depends on known attacks in the signature dataset. Anomaly-based attack detection methods differentiate anomalous web requests by assessing the probability threshold of attributes in the request against established normal request profiles.

2.2 Previous Research

Previous research is very important to strengthen or support the strength of the research to be carried out because there are already scientific references that have relevance so that the research is stronger and more accurate. The following is an explaining previous research :

Nasib Ullah and Partha Pratim Mohanta (2023) conducted a study on video captioning with object detection, Improved object detection has become a focus in the advancement of video captioning tasks. Object detection is a key area of computer vision that has enhanced other computer vision tasks, including classification and detailed captioning. Object detection has been used in a variety of applications, including face detection, vehicle detection, and motion detection. Object detection is also used in security applications like intrusion detection systems (IDS),

video surveillance, web attack detection, and advanced persistent threats (APT). Object detection has helped improve safety and efficiency in a variety of industries, including automotive, manufacturing, and healthcare. The study provided an overview of both types of object detection methodology:

1. Two-stage region proposal-based methods: These methods first generate region proposals and then classify the objects within those regions. They are more accurate but slower compared to single-stage methods.
2. Single-stage regression-based methods: These methods directly predict the object boundaries and class labels in a single step. They are faster but less accurate compared to two-stage methods.

The study aimed to explore the potential of object detection in enhancing video captioning tasks by utilizing the various objects that appear in the video and employing off-the-shelf object detection techniques on video frames

Continuing with Asifuzzaman (2023) conducted a study on a deep learning-based framework for COVID-19 identification using chest X-ray images. The study aimed to develop a deep learning-based system to categorize four distinct kinds of chest X-ray pictures. The researchers compared their results with the state-of-the-art methods and achieved better performance. The proposed framework used a deep convolutional neural network (CNN) architecture to extract features from the chest X-ray images.

Vipul (2023) conducted a study on the detection of surface defects in aluminum tubes using Faster Region-based Convolutional Neural Networks (Faster R-CNN) . The study aimed to develop a deep learning-based system to detect surface defects in aluminum tubes. The proposed system used Faster R-CNN, which is a state-of-the-art object detection algorithm. The Faster R-CNN algorithm consists of two modules: a region proposal network (RPN) and a Fast R-CNN detector. The RPN generates region proposals, and the Fast R-CNN detector classifies the proposals into different categories. The study used a publicly available dataset of aluminum tube images to train and test the proposed system. The results showed that the proposed system achieved high accuracy in detecting surface defects in aluminum tubes. The study concluded that the proposed system could be used as a reliable tool for surface defect detection in aluminum tubes.

Navpreet (2023) conducted a study on a real-time face detection-based automobile safety system using computer vision and supervised machine learning. The study aimed to develop a system that can detect the driver's face in real-time and alert the driver if they are not paying attention to the road. The proposed system used computer vision and supervised machine learning techniques to detect the driver's face and classify it as attentive or inattentive. The study used a publicly available dataset of face images to train and test the proposed system. The results showed that the proposed system achieved high accuracy in detecting a driver's face and classifying it as attentive or inattentive. The study concluded that the proposed system could be used as a reliable tool for automobile safety.

Sanoar (2023) conducted a study on texture feature descriptors to analyze facial patterns in facial expression recognition systems. The study proposed a multipose facial expression recognition system that has three components: image preprocessing, feature extraction, and classification. The feature extraction component used texture feature descriptors to analyze facial patterns. This study used a publicly available dataset of facial expression images to train and test the proposed system. The results showed that the proposed system achieved a high accuracy in recognizing facial expressions.

The study by Chiara Pero (2023) provides a brief overview of recent techniques in crowd counting and density estimation. The study highlights the importance of estimating crowd counting from images and the challenges associated with it. The study also discusses various techniques such as convolutional neural networks (CNNs), local binary patterns (LBPs), and scale-invariant feature transform (SIFT) for crowd counting and density estimation

And all previous research is summarized in the following table:

Table 1 Previous Research

no	Researcher, Year, Title	Research Purposes	Research Summary	
			Equality	Difference
1.	Nasib Ullah, Partha Pratim Mohanta. (2023). Recent Advances in Video Captioning with Object Detection. ISBN 978-1-003-39365-8 (ebook master)	describes the advancements in the **video captioning task** due to improved object detectors.	Using deep learning for fine-grained classification	Different case and purpose detection
2.	Asifuzzaman L., Mridul Ghosh., Chandan Chakrabort.(2023). A Deep Learning-based Framework for COVID-19 Identification using Chest X-Ray Images. ISBN 978-1-003-39365-8 (ebook master)	the **detection** of COVID-19 among disparate sets of lung images captured normal as well as affected using homogeneous chest X-ray images.	Using **deep learning RCNN-based** Defect Recognition	Different study case and purpose detection
3.	Vipul Sharma., Roohie Naaz Mir .(2023). Faster Region-based Convolutional Neural Networks for the Detection of Surface Defects in Aluminium Tubes. ISBN 978-1-003-39365-8 (ebook master)	The proposed faster RCNN outperformed RCNN in terms of recognition speed and accuracy	Using deep learning **RCNN-based Defect** Recognition	Using different surface defect detection and study case

4.	Navpreet S. Kapoor ., Mansimar A ., Priyanshu .(2023). Real Time Face Detection based Automobile Safety System using Computer Vision and Supervised Machine Learning. ISBN 978-1-003-39365-8 (ebook master)	the leading automobile safety technology, offers a detection and prevention system, and describes its specific working methods	The method is using **Supervised Machine Learning**	Same to analyze the performance of the proposed model, various evaluation Parameters and classification
5.	Sanoar H., Vijayan A. (2023). Texture Feature Descriptors for Analyzing Facial Patterns in Facial Expression Recognition System. ISBN 978-1-003-39365-8 (ebook master)	a multi-class classification technique using support vector machine classifier is employed to perform **classification** task to detect the type of facial expression in the image.	Using multi-level feature computation, various global and local **feature** representation schemes	Different study case and purpose detection
6.	Chiara Pero . (2023). A Brief Overview of Recent Techniques in Crowd Counting and Density Estimation. ISBN 978-1-003-39365-8 (ebook master)	Estimating **crowd counting** from images	The method for using crowd counting, density estimation, crowd analysis	Using basic CNNs for crowd counting based on the multi-stage

7.	Prithwish J., Partha P. (2023). Recent Trends in 2D Object Detection and Applications in Video Event Recognition. ISBN 978-1-003-39365-8 (ebook master)	the usefulness of using frame-wise object detection in videos for the task of **activity recognition** and video event classification	Using deep learning for object detection serve.	
8.	Gaurish G., Shailendra T., Shivendra S. (2023). Survey on Vehicle Detection, Identification and Count using CNN-based YOLO Architecture and Related Applications. ISBN 978-1-003-39365-8 (ebook master)	Talking about the detailed **analysis of detection**, identification, and count of vehicles in an image, using convolutional neural networks CNN)- based YOLO (You Only Look Once) architecture	Using convolutional neural network (CNN) for object classification and machine learning	Different study case and identification Architecture.
9.	Sariva S., Rajneesh Rani. (2023). A Comprehensive Review on State-of-the-Art Techniques of Image Inpainting. ISBN 978-1-003-39365-8 (ebook master)	a functional computer vision research problem with the goal of improving an image quality by removing unwanted details, adding missing elements, and presenting the image in a way	Using **Deep learning** need a significant amount of training time to prepare deep learning models	Use of deep learning technologies with image inpainting techniques

		that appeals to the human visual system		
10.	Padmaja P., Swathi B., An Extensive Study on Object Detection and Recognition using Deep Learning Techniques	discusses object detection and recognition and their applications and different algorithmic approaches	Using RCNN, **image classification**, feature extraction, neural networks.	Different study case and identification Architecture.

2.3 Framework of Thought

The framework of thinking in quantitative research has an important function in helping researchers determine theories, concepts and postulates which will later be used as the basis for research. Apart from that, the thinking framework also helps researchers in getting mature concepts which are then used to explain each problem in the research. The thinking framework also makes research easier by providing clear direction and focus and connecting research elements, such as variables, theories and facts, making it easier for researchers to understand the relationship between these elements. By using a thinking framework, researchers can obtain more accurate and structured research results and gain a better understanding of the theory and variables studied. (Nam Bui et al., 2020) A thinking framework also helps readers understand research more easily and clearly and avoid mistakes in arguing.

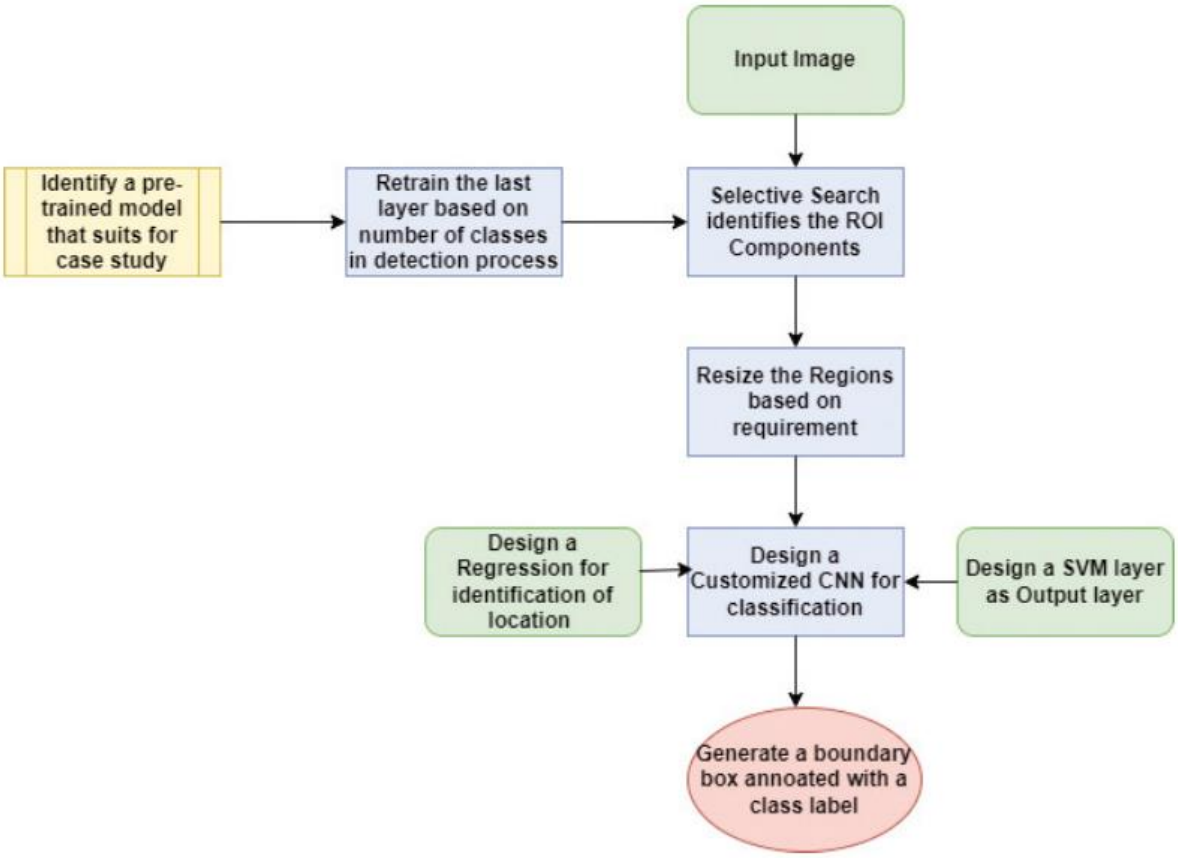


Figure 2 Object detection process using traditional RCNN approach(Naaz Mir et al., n.d.)

Initially, the input image is passed to the RCNN model, and then the selective search approach generates candidate regions by performing a sub-segmentation process. Then, to reduce the number of regions, it applies a greedy approach recursively to combine similar regions. Then combined region are known as “Proposal region”. All the region identified have different spatial coordinates, with which the CNN model cannot work because CNN needs a standard size as input(Haurum & Moeslund, 2020)

Hence, the model needs to apply geometric transformation functions like rescaling or resizing to get all the images of the same size. Finally, the retrained model in which the last layer are design using SVM for classification is applied as the last but one layer with its mathematical notation to minimize the distance between the point and find as many point as possible. This shown in eqn 1 and 2:

$$P^* = agrg_p\left(max\left(min\left(dist_H(\emptyset(X_n))\right)\right)\right)$$

$$Dist_H(\emptyset(X_n)) = \frac{P^T(\emptyset(X_0)) + b}{\sqrt{P_1^2 + P_2^2 \ldots + P_n^2}}$$

Here $P^T(\emptyset(X_0)) + b$ represents a hyperplane equation to split the data points in the crowd into positive and negative classes. The last layer in CNN is designed using linear regression to find the coordinates associated with the object using the eqn 3:

$$ind\_var_1 = f(Feature_i, \beta) + Error\_rate,$$

Where $f(Feature_i, \beta)$ represents a mapping function between the features of the image and unknown parameters. The below section discusses various applications of the detection mechanism.

1. Tracking an object plays a vital role in the security aspects by checking the live video streaming continuously
2. Counting of things or person through images and videos to make the checking and packing as quickly as possible
3. Vehicle and person detection to identify things while unexpected things happen.
4. Identification of number plates at roll gates in the process of collecting toll passes, finding the rash driving to impose fines on particular vehicles, to

reduce the traffic at toll gates by automatic detection of Fast Scan tags, a major application of OCR

5. Self-driving cars are popular in abroad countries where cars can automatically identify the persons crossing the roads, signal at the traffic junction

2.4 Operational Hypothesis

This operational hypothesis is formulated based on the theory and research objectives to be achieved. In addition, this operational hypothesis can also be measured empirically by using measuring instruments that are in accordance with the research. In this study, the operational hypothesis is used to examine the relationship between the variables involved in deep learning-based CCTV sensor systems for identifying anomalies in crowds.

The operational hypotheses that researchers can observe are as follows:

- There is a positive relationship between the accuracy of deep learning-based CCTV sensor systems and the ability to identify anomalies in crowds.
- There are significant differences in the accuracy of deep learning-based CCTV sensor systems between different lighting conditions.
- There are significant differences in the accuracy of deep learning-based CCTV sensor systems between crowds of different sizes and densities.