Istanbul Medipol University
School of Engineering and Natural Sciences

Graduation Project

2022-2023

2022-2023

PROJECT TITLE
Leakage Detection on Infrared Images of Under Pipes
FACULTY ADVISOR
Prof. Dr. Reda ALHAJJ
TEAM MEMBERS
Mihrimah Göknar

Istanbul Medipol University

School of Engineering and Natural Sciences

Graduation Project

2022-2023

Project Title: Leakage Detection on Infrared Images of Under Pipes	
Faculty Advisor: Prof. Dr. Reda ALHAJJ	
Project Team Members: Mihrimah Göknar	
Sponsor Company (if any):	

BUDGET (TL)	PROPOSED	CONSENTED
IMU FUNDING	93100 TL	
SPONSOR COMPANY FUNDING		
TOTAL	93100 TL	

PROJECT PLANI	PROPOSED	CONSENTED
PROJECT PLAN	02.2022 - 01.2023	
STARTING DATE	26.02.2022	

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

2022-2023

ADVISOR	DEPARTMENT CHAIR
Name: Reda Alhajj	**Name:** Mehmet Kemal Özdemir
Contact Information: Tel : E-mail: ralhajj@medipol.edu.tr	**Contact Information:** Tel : E-mail: mkozdemir@medipol.edu.tr
Signature:	**Signature:**
TEAM MEMBER	**TEAM MEMBER**
Name: Mihrimah Göknar	**Name:**
Contact Information: Tel : 531 834 1149 E-mail: mihrimah.goknar@std.medipol.edu.tr	**Contact Information:** Tel : E-mail:
Signature:	**Signature:**

SPONSOR COMPANY
Name:
Contact Information: **Tel:** **E-mail:**
Signature:

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

2022-2023

Project Title: Leakage Detection on Infrared Images of Under Pipes
Faculty Advisor: Prof. Dr. Reda Alhajj
Project Team Members: Mihrimah Göknar
Project Group Title:

ABSTRACT

Leakage in pipelines carrying water or hydrocarbon fluids is one of the risks that can result in serious injuries, environmental disasters, and a significant economic effect. A preventive, non-destructive inspection method can help to avoid a situation like this. Infrared technology (thermal imaging technology) is a non-destructive assessment for detecting and assessing the severity of a leak by employing sensors to identify the wavelength of infrared light emitted from an object's surface. Using infrared image analysis, this research proposes a hybrid model containing 2 step control with Decision Tree (DT) model on the metadata and with Convolutional Neural Network (CNN) on the thermal image for the leak detection. In addition, the leak propagation analysis is done on a set of images with time information to assess the leakage severity. The analysis done on the IR image dataset and its metadata provide information such as leak detection, pipe failure condition, and leak propagation. With the hybrid model, the background information of the pipes on a network will be analyzed first for the pipe failure/leak possibility. Afterwards, the thermal image of the detected pipe can be acquired with a thermal camera implemented drone without needing an open inspection and non-destructively.

Keywords: Leakage Detection, Image Processing, Infrared Technology, Machine Learning, Metadata, CNN, Decision Tree

BUDGET PROPOSED– (TL)

Table 1: Budget Proposed

	ITEMS				
	PEOPLE	**MACHINE INSTRUMENT**	**MATERIALS**	**SERVICE**	**TRAVEL**
IMU FUND	**75000 TL	***17000TL		****1100 TL	
SPONSOR COMPANY FUND					
TOTAL	75000TL	17000TL		1100TL	

BUDGET APPROVED– (TL)

Table 2: Budget Approved

	ITEMS				
	PEOPLE	**MACHIN INSTRUMENT ***	**MATERIALS ***	**SERVICE**	**TRAVEL**
IMU FUND					
SPONSOR COMPANY FUND					
TOTAL					

Table of Contents

Table of Figures

Table of Tables

ABBREVIATION LIST

IR: Infrared Radiation
IRT: Infrared Technology
NDT: Non-Destructive Testing
GUI: Graphical User Interface
PCA: Principal Component Analysis
CNN: Convolutional Neural Networks
DT: Decision Tree

1. OBJECTIVE OF THE PROJECT

Natural gas and fluids such as water and hydrocarbon fluids are frequently transported via pipelines spanning millions of miles all across the world and pipeline structures built to resist various environmental situations, guaranteeing secure and reliable transfer from the starting point to the ending station. On the other hand, pipeline leaks are among the leading causes of immense losses for operators and the environment. Failures in pipeline can cause major environmental catastrophe, financial losses and human casualties. The demand for these resources is rapidly increasing as a result of population expansion, economic growth, and changing consumption habits, and this demand is predicted to rise dramatically over the next 20 years [1] (as cited in Tindall et. All, 2010). Therefore, the detection of the pipeline leakages is a great concern. This project will focus on the detection of leakages in buried pipes with the help of Infrared Technology (IRT). IRT stands for infrared imaging technology, which use infrared cameras to detect temperature changes in the pipe surroundings between the wavelengths of 900 and 1400 nm[2]. By detecting warm and cool locations shown in the thermal image with the various colors of the environment, the object captured with the thermal camera can be examined to recognize abnormalities in the pipe area. The implementation of an IRT system for pipe status surveillance enable for the early detection of malformations in the pipeline network, reducing gas, water, oil waste along with preventing hazardous effects on the living creatures and environment. Consequently, the aim of this research is to identify potential leakage, anticipate the severity of the leak, evaluate the feasibility of repairing the leakage and calculating the expected loss.

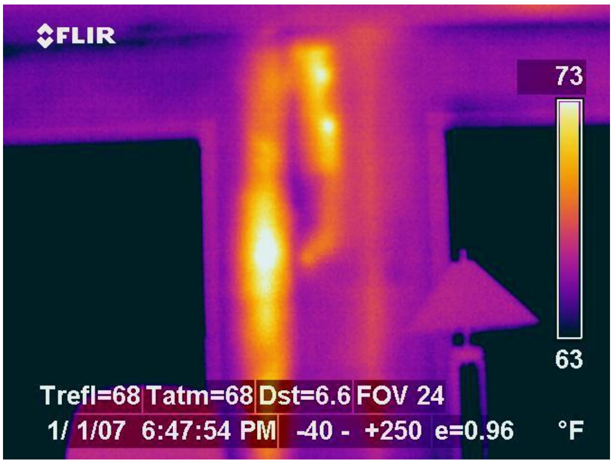

Figure 1: An example IR image from FLIR camera

2. LITERATURE REVIEW

The goal of this study is designing a model for the leakage detection on pipelines and doing a feasibility research based on thermal camera images. Within the scope of this purpose, literature review is divided into three main sections. The initial focus of this study is the leakage detection. The identification of leakage in infrared images is the subject of the second focus. The third and final focus is the predictive modelling on leakage on the pipeline systems. The prior literatures will be summarized based on these study focuses.

2.1 Leakage Detection

Hydrocarbons provide 84% of the world's energy, but they are non-renewable. Additionally, water is not a non-renewable resource, nor is it completely renewable and global water use currently corresponds for 10% of the total consumption [3]. Both of these sources are mostly transported via pipelines that reach millions of miles around the world to fulfil the world's need. Water is vital for all living creatures, and due to poor leak detection in world water utilities, detection of water leaks is essential, as with an estimated 1/3 of world water utilities wasting about 40% of the water due to leakage [4] (as cited in Gupta, 2017). Water pipeline leak detection is very important for water delivery networks to operate safely and efficiently while conserving water resources.
Crude oil has been the most important source of energy on the planet since the mid-1950s. It has been used in different areas from being an energy resource, transportation to industry. Natural gases have been also used from being an energy resource to heating. According to GlobalData statistics, a total length of 2,034,065.0 kilometers is estimated for the global transmission pipeline network. Over 379,000 km of crude oil pipelines are present, whereas over 267,000 km of petroleum products pipelines exist. Natural gas pipelines account for about 1,300,000 kilometers, while natural gas liquids account for approximately 92,000 kilometers [5] (as cited in GlobalData, 2019). Considering most of these pipes are buried under earth, the major causes of the leakages are the corrosive properties of the soil, temperature and pressure, poor material quality in pipes, inability to follow acceptable laying methods for pipes, human damage and geological changes. Pipeline leak localization and detection using several approaches has been a major focus of study to avoid this issue and maintain a secure and dependable pipeline system.
Acoustic emission, fiber optic sensor, vapor sampling, pressure point analysis, ground penetration radar, dynamic modelling, negative pressure wave, infrared thermography, mass-volume balance, and digital signal processing are some of the existing methods on leakage detection. Further researches were performed to categorize these approaches according to their technological character, resulting in the categorization of leakage detection systems in 3 major categories: internal, non-continuous, and external methods [5].

Table 3: Literature Review Related to Leakage Detection

STATEMENT	REFERENCE
In this article, a leakage triggered networking method is developed [1] to decrease energy consumption since real-time observation of pipelines with numerus sensors is energy consuming and affects network lifetime. Their method first initialized the network then used triggered networking with 3 types of controlling frames: active frame, wave frame	(Liu et al., 2019)

and join frame. They employed intrinsic mode function, approximate entropy, and principal component analysis (PCA) which utilize SVM as a classifier to detect leakages.	
In this study, a comparative performance analysis is carried out to help determine which leak detection technology is best for a given operational environment [2]. Aside analysing different leakage detection methods, with the strengths and disadvantages, research gaps and unresolved concerns for the progress of dependable leakage detection systems are addressed.	(Adegboye et al., 2019)
A system with an Machine Learning (ML) algorithm on water distribution systems within a wireless sensor network is developed for detecting precise location of water leakage [4]. The system uses multiple sensors with real-time data collection where the system is divided into multiple modules both on hardware and software parts. Data collection is done from the hardware nodes; data processing and the user interaction is done from the server. The system consists of network of sensors which creates Wireless Sensor Networks (WSN) that is capable of sending data, where it gets data from sensors, to cloud. To send the data outside of the network, an NB-IoT, a standard-based low-power wide-area network, is used to allow devices to connect through mobile phone signal. After that user could see data collected from sensors as well as the current condition of the system where ML algorithms, as well as previous data and situations, are used to assess it.	(Alves Coelho et al., 2020)
The study examines important leak detection methodologies and regulations for oil and gas pipelines, and discusses the possibilities for developing a reliable real-time leak localization and detection system for surface pipelines using data analytics tools and mathematical modeling, also provides some selected case studies [5].	(Idachaba & Rabiei, 2021)

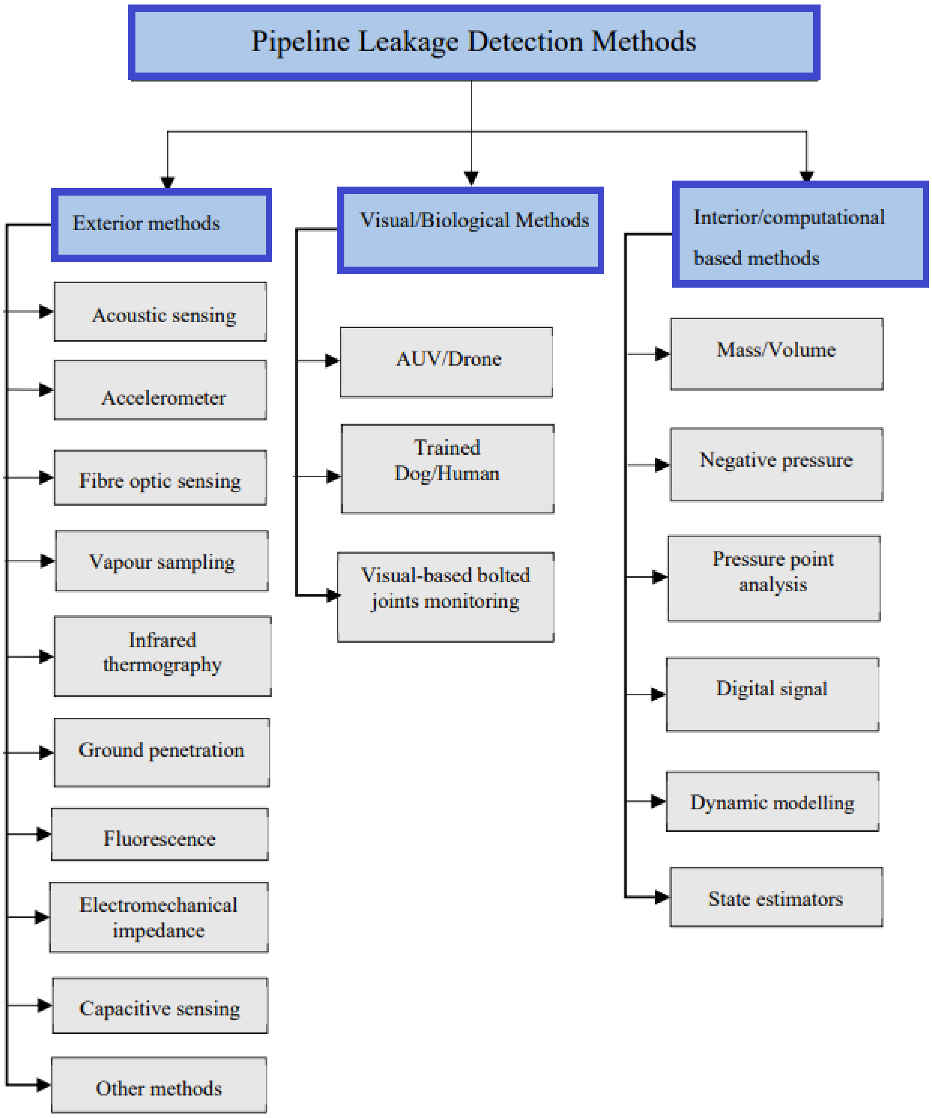

Figure 2: Different methods for detecting pipeline leaks [2]

2.2 Leakage Detection on Infrared Images

A thermal camera is among the tools that can be employed as non-contact inspection tool. This is an infrared technology that measures temperature and the energy radiating from an item using the principle of Infrared Radiation (IR). The thermal image processing technology enhances the pertinent information in the image, allowing important image details to be seen. Thermal cameras can sense the presence of an object from any angle and at any location, resulting in a temperature difference seen as a color bar. Its distinguishing feature is non-destructive evaluation, which doesn't alter the temperature profile of the measured object while offering a, high sensitivity, quick response time, wide temperature range, high precision, long lifespan, low power consumption and ease of use.

Table 4: Literature Review Related to Leakage Detection on Infrared Images

In this paper, an image processing technique is provided to extract the leakage areas in the pipes on the IR image [6]. They identified the factors that affect thermal contrast. Then, analysis and interpretation of data collected from the infrared camera done for detecting and finding leaks. Comparison between the detected and original areas is done and the validation of the intended method for pipeline leaking is done.	(Manekiya & Arulmozhivarman, 2016)
In this article, a case study is provided to emphasize the advantages of thermography, in which thermal photos were acquired in a specific place under various conditions containing a buried pipe. [7]. A thermographic camera was used to take a series of photographs in the same location where a hidden pipe was known to be exist. Environmental data variables were also collected in order to create a database with all of the different situations. Different criteria were considered to contrast the photographs, such as temperature variation because of the season change, weather in the day (cloudy, sunny, etc.) and, finally, occasional components, such as objects or people causing temperature scale variation. Various tests were carried out in order to obtain images of the buried pipe.	(Carreño-Alvarado et al., 2014)
By testing series of experiments on 4 kinds of stainless and carbon steel pipes on the constructed large pipe testing installation, they showed on the study that infrared thermography can detect corrosion pits or wall loss flaws which are greater than a diameter of 10 mm × 40% of wall thickness [8]. They also came to the conclusion that the sensitivity of infrared thermography testing is far more than the acknowledged deficiencies for safe pipe operation (10 mm diameter and 80 percent wall loss defect is safe for all the four pipes).	(Shen & Li, 2007)
They provide a system which is capable of distinguishing between two types of gas conditions: normal and abnormal [9]. They provided a GUI for their system which is capable of classifying situations in which maintenance is required by following these steps: image pre-processing (converting to grayscale and filtering), image processing (applying threshold and binarization), feature extraction, and finally classification.	(Jadin & Ghazali, 2014)
A method for measuring leak flow as a function of measured ultrasonic or temperature change is proposed in this study [10]. They concluded that ultrasound should be used for leakage detection for all orifice diameters, but only for quantification purposes for smaller leaks, namely leaks with sound levels up to 74 dB, because the features they used for leakage detection, sound generated by air leak and change in air temperature, provided a certain range for the leak to be identified. An IR camera should be used to assess leakage if the sound level is greater than 74 dB, as it will provide a more accurate estimate of leakage. As a result, measuring larger losses due to leakage through orifices 1.3–2.0 mm with an ultrasound detector is impossible, but it is doable using an infrared camera.	(Dudić et al., 2012)

In this article, the general structure of an IR camera is described, as well as the image processing activities conducted in the observation of thermal cameras used in security systems, detection, and identification [11]. The authors devised methods for recognizing and tracking objects on thermographic pictures, and developed an object tracking approach, since the detection algorithms used for tracking and vision systems working in visible light can't be directly applied to the analysis of thermal images. The method enables for accurate tracking of moving objects in thermographic images, and proved to be specifically vigorous to low image resolution, which is common of infrared cameras, as well as blurred object edges.	(Sosnowski et al., 2018)
The methodology of this study is processing real-world photos taken as part of a lab experiment [12]. The primary objective of this study's first phase was to qualitatively test this algorithm's behavior in order to accurately locate and detect water leaks. According to the literature, the q-sigmoid function's parameters were determined empirically.	(Carreño-Alvarado et al., 2014)
In this research, authors aim to investigate the possibilities of getting thermographic pictures to create visualization approaches for pipelines as a potential leak detection method [13]. They analysed the data, which they acquired on a controlled lab environment by checking the camera and environmental properties, with different visualization approach for leak detection.	(Pauline et al., 2020)

2.3 Predictive Modelling on Leakage on the Pipeline Systems

The ability of predictive models to accommodate various levels of uncertainty is the primary motivation of their construction. Pipeline operation, pipe failure, and pipe assessment are all intrinsically dangerous. Pipelines are subjected to a variety of risky and uncertain conditions throughout operation, resulting in deterioration and eventually failure. Furthermore, when pipelines are subjected to operational evaluation, ambiguous scenarios such as determining the precise location of the failure, the true reasons of failure, and correct assessment methodologies are encountered. As a result, predictive models are used to estimate several aspects of pipeline failure, such as pipeline risk, deterioration rates, safety evaluation methods, failure likelihood, and system reliability, among other things.

Table 5: Literature Review on Predictive Modelling

In this study, an extensive review on how probabilistic models have been used to forecast pipeline condition in various ways is presented [14]. The authors deduced that predictive models are depicted as highly data driven, requiring a wide range of data for prediction purposes from the various evaluations completed for the study. Since mostly data is only available on a limited basis, when limited but accurate or partial data is available, it is safer to rely on expert or engineering knowledge to	(Ogutu et al., 2017)

improve modelling. On this context, the authors focused on Bayesian Networks (BNs), which are graphical models used to display knowledge about uncertain domains or used for reasoning under uncertainty, as one feasible technique for such a scenario. Consequently, they conclude that pipe age and pipe break rate are the leading causes of failure and in addition, they made some suggestions for future research work.	
In this article, the authors examined the capacity of 6 different ML methods for localization of water leaks with the flow and pressure data[15]. Firstly, by using ERANET software they generated pressure and flow data of water with 251 leak scenarios. Then they analysed 3 supervised ML methods (logistic regression, decision tree, and random forest), and 2 unsupervised methods, by using the generated data. They acquired excellent results with the supervised ML methods, and with ANN method for the localization of water leaks in the water network. However, overlapping clusters made it difficult to leverage the clustering potential of unsupervised approaches in leak localization. Secondly, they used offline water flow data of campus of Lille University recorded in 2015. Due to lack of pressure data, they had difficulties in determining the position of the leaks.	(Mashhadi et al., 2021)
The authors analysed different statistical and ML methods to offer information to operators for selecting an appropriate pipe failure model based on data availability and network characteristics [16]. The analysis of statistical models revealed that when information about variables other than pipe features is unavailable, the cluster-based prediction technique minimizes the prediction error for pipe failure. They concluded that Poisson Regression was the best suited among the statistical model and all approaches in ML models performed well, with the exception of ANNs.	(Giraldo-González & Rodríguez, 2020)
Based on photos obtained by an Unmanned Aerial Vehicle (UAV), the authors presented a system for automatically identifying leaks in underground pipes of district heating networks [17]. They conducted their research with 2 different approaches for both 16-bit and 8-bit image data, which is converted with Dynamic Range Reduction (DRR) operator. First approach is deep learning based which is a CNN model, and second approach is ML classifiers. They used 8 different ML classifiers on which 4 of them linear and other 4 are nonlinear models. According to the results, they concluded that CNN gives best result for both 8- and 16-bit data but requires high computational resources; while, Adaboost achieved the best results between ML models for both 8- and 16-bit data.	(Hossain et al., n.d.)

In this article, an approach that combines image processing with a metaheuristic-optimized SVM model is proposed [18]. Texture-based features are retrieved from image samples using statistical measures of color channels, Gray-Level Co-Occurrence Matrix (GLCM), and Gray-Level Run Lengths in the image processing section (GLRL). The model divides data samples into corrosion and non-corrosion classes using SVM. To improve the SVM-based training and prediction phases, the DFP metaheuristic is applied.	(Hoang & Tran, 2019)
In this study, authors focus on detecting leaks in Water Supply Networks (WSN) by utilizing 2 machine learning methods [19]. One of them is Artificial Neural Network (ANN), a supervised model, which successfully classified the leaking versus non-leaking cases while having the burden of requiring balanced data for leaking and non-leaking conditions. Another one is Autoencoder Neural Networks (AE), an unsupervised model, which further improved for the detection of leaks with unbalanced data.	(Fan et al., 2021)

Additionally, the following articles were taken into consideration for the hybrid model implemented for this research. Metadata enhanced image classification and multi criteria decision method approaches were used for creation and evaluation of the pipe's metadata. Based on the research article El Chanati and others conducted which analyse four different multi-criteria decision-making techniques to create performance assessment models for water pipelines, feature vector information of water pipeline and the feature weights calculated with AHP technique are created for the fictional metadata of pipes. Based on the result of the second article (Thomas,2021), which concludes as adding metadata to the image features increases classification performance, a hybrid model is built for thermal images and metadata of pipes to increase the performance of classification.

Table 6: Additional Review

In this article, the creation of performance assessment models for water pipelines is discussed by taking into account the physical, environmental, and operational elements that affect the performance of water pipelines [20]. The models were created using information gathered from questionnaires provided to Qatari water pipeline experts. The primary goals of the current study are identifying and investigating factors influencing the performance of water pipelines; calculating the weights of importance of the identified factors using four different multi-criteria decision making (MCDM) techniques; developing four performance assessment models for water pipelines; and comparing the outcomes from each developed model.	(el Chanati et al., 2016)
The author provides a comparison of how well deep neural networks (6 different deep CNNs) perform in classification tasks when using only image features versus when these attributes are paired with patient metadata [21]. The research is comprised of image processing techniques and utilizing transfer learning while including and not including metadata, to classify the disease of patient. Also, it includes a	(Thomas, 2021)

comparison of the results with and without data augmentation during the training of the model.	
This study provides a summary of the literature on the Decision Support System (DSS) used for water supply system leak detection and control by dividing the research on DSS for leak localization and control into two primary groups: DSS employing wireless sensor networks and DSS using machine learning approach. The results obtained from this literature researches are examined and then the future research is discussed [22].	(Shabangu et al., 2020)

3. ORIGINALITY

In the literature, there are many studies conducted on various pipeline leak detection methods. As mentioned in the literature, all of the methods have strengths and weaknesses coming with them. Some even can identify leak location while some have a great sensitivity to environmental variables. In addition, every method has a different performance characteristic such that some can detect the leakages on gas pipes, some on liquid pipes and some on both. Also, false alarm rates of each method are differed.

After careful examination of the literature, the method to be studied, Pipeline Leakage Detection and Investigation on IR images, will have some significant distinctions from the research done. In this project, apart from the pipe leak detection, it is also aimed to do detection investigation such that; anticipating the amount of leakage and the severity of the case, is it feasible to fix, the prediction of how the case will evolve and the expected loss from the leak.

4. SCOPE OF THE PROJECT AND METHODS

IRT has been used for leak detection on water, gas, and hydrocarbon fluid pipelines of any kind and size. IRT enables temperature measurement that is contactless, non-invasive, real-time, and distributed throughout a continuous under pipe zone [2]. It is easy to use and does not require specially trained or experienced personnel. However, it is very expensive to build an infrared camera system with high resolution.

On the other hand, although IRT may remotely assess temperature distribution of an object and demonstrate a visual representation with distinct colors showing the degree of the measured data in that place, it can be easily affected by environmental changes (living creatures, weather, etc.) when the area of study cannot be completely isolated [7].

Considering these aspects, the data for an under-pipeline area/system (preferably both isolated and not isolated to achieve consistent results) monitored with IRT will be tried to acquire, the prediction model for the leakage detection with size, location information on the images will be tried to be done with ML methods which may require different image processing methods. If the data can be retrieved, the test results and their occurrence with different times and weather conditions can be taken into account to analyze whether there is a pattern or conditional properties that affect leakage on the pipes. After that, from the acquired knowledge from the IR image, the anticipated leakage amount will be tried to be predict. Also, the prediction on the course of the leakage: How long the pipe can handle, whether the leak can be fixed or whether the entire pipe needs to be changed, and the expected loss, will be tried to obtained with the help of a feasibility study.

With the proposed method, the input metadata and thermal images will go through various stages which is described as shown in the **Figure *3***.

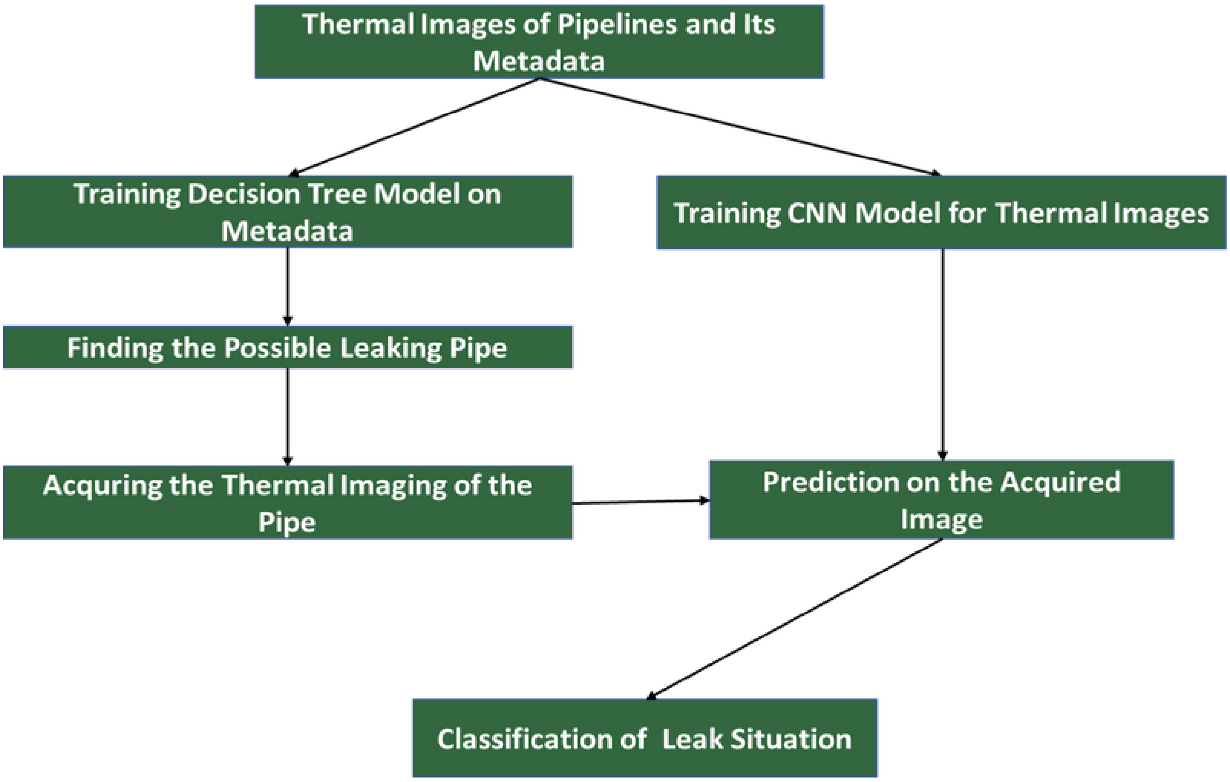

Figure 3: Steps for Hybrid Model of Leakage Detection

4.1 Thermal Images of Pipelines

This is the first step of the proposed methodology. The data containing thermal camera images of pipes, is taken from a private source. It consists of 2 class of images as Leak and No Leak. In addition, there is knowledge of time taken of a leaking set of images. These leaking images that contains time information is used for leak propagation analysis and will be analysed in the next sections.

4.2 Metadata of Pipes

Based on the information provided by El Chanati et. all (2015) [20], the metadata of the pipes is created fictionally. They developed water pipeline performance assessment models while taking into account the operational, environmental, and physical factors that influence the performance of water pipelines. Due to the characteristics of the pipe and installation methods, static elements like material, diameter, and installation method remain constant over time. Age, soil temperature, dynamic loads, and other dynamic characteristics are connected to the environment around the pipe. The operational factors, such as flow rate, operating pressure, and replacement rate, show how the pipe is used during its service life. By taking the article as a base, following factors are selected for fictional data set.

Table 7: Factors Affecting Water Pipeline Performance [20]

Main Factors	Sub Factors	Description
Physical (P)	Age (AG)	Over time, the effects of pipe deterioration become increasingly noticeable.
	Material Type (MT)	Different failure scenarios are linked to various types of pipeline material.
	Size (SZ)	The dimensions of a pipe also include thickness and length. Longer pipes are vulnerable to higher rates of degradation, and vice versa.
	Installation Quality (IQ)	Analyse whether the installation followed the instructions and met the requirements. Poor installation quality results in a high breakage rate, and inadequate pipe bedding may cause pipe failure.
Environmental (E)	Surface Type (SR)	It is important to consider whether the installation zone for a pipeline is residential, industrial, or in a school, etc. Pipes located in residential regions may be covered by different surface types, making them more susceptible to certain conditions than those found in industrial or urban settings (e.g., asphalt, seal, unpaved etc.). City pipelines beneath highways are vulnerable to considerable traffic, which increases the pipe's dynamic stress.
	Ground Water Level (GW)	This aspect relates to how much groundwater has an impact on the pipe. The soil resistivity, which is inversely related to the rate of corrosion, is influenced by the amount of water in the soil. If there are salts and other corrosive materials in the ground water, that could immediately cause the pipe to corrode.
	Soil Type (ST)	Refers to the kind of soil that is directly in contact with the surface of the pipe; various soils have various effects on the pipe. Some soils are corrosive, while others undergo substantial volume changes in reaction to variations in moisture, which alter pipe loading. Pipe deterioration may be spurred on by the solvents and hydrocarbons present in the soil.
Operational (O)	C-Factor (CF)--Pressure/flow velocity/ Hazen-Williams coefficient (CF)	The pressure brought on by transients in the water distribution system may result in the failure of pumps and other equipment, system wear, or pipe ruptures. When flowing between pipes with differing diameters, high velocity water notably corrodes the internal pipe walls and creates disruptions. Hazen-Williams coefficient includes contributions from pressure and flow velocity (C-Factor). The smoothness of a pipe's inside is measured using the C-factor. Briefly, high C-Factor indicates, smoother pipe, higher carrying capacity, and lower amount of energy lost through friction as the water flows through the pipe.

	Leakage/Breakage Rate (BR)	Water leaking from the pipe through surface fractures is known as leakage. Water leaks enlarge cracks, raise soil moisture levels nearby, and increase the likelihood of exterior corrosion. The stress distribution on the pipe may alter and as a result, finally resulting in pipe breaking. The number of breaks/km/years is known as the breakage rate. A pipeline with a high breakage rate is likely to be in bad condition, hence repair is necessary.
	Water Quality (WQ)	While being moved, chemicals and other materials (such as salts and micro-bio-organisms) degrade the quality of the water while also having the potential to erode the pipe's internal surface and lead to pipe breakage.

According to the questionnaire they made, a pairwise comparison between them was conducted using Saaty's (1980) [23] fundamental pairwise comparison scale to establish the relative relevance of the chosen criteria. The questionnaires were created to make it easier to determine the level of significance of each aspect in relation to a certain goal. An assigned value of "10" indicates that a certain component has absolute priority over the comparison factor, whilst assigning a value of "1" denotes that the two factors under consideration have "equal" importance with respect to the defined goal.

Table 8: Factors' Average Effect Values (EV_i) [20]

Factor	**Characteristic**	**Average effect value (EV_i)**
Age (AG)	Old (>70 years)	1.0
	Medium (30–70 years)	5.0
	New (<30 years)	9.5
Material type (MT)	Asbestos	5.5
	Cast Iron	5.5
	Concrete	6.5
	Ductile iron	7.5
	PVC	9.5
Size (SZ)	Small (<200mm)	3.0
	Medium (200-350 mm)	6.5
	Large (>350 mm)	9.5
Installation quality (IQ)	Poor	2.5
	Fair	6.0
	Good	9.5
Surface type (SR)	Asphalt	4.5
	Seal	6.5
	Foot path	7.0
	Unpaved	7.5
Ground water level (GW)	Shallow	2.0
	Moderate	5.5
	Deep	9.5
Soil type (SL)	Aggressive	1.5
	Moderate	5.0

	Non-aggressive	10.0
C-factor (CF)	Low (<41)	3.5
	Medium (41-101)	6.0
	High (>101)	10.0
Breakage rate (BR)	High (>0.5 breaks/km/year)	1.5
	Medium (0.1–0.5 breaks/km/year)	4.5
	Low (<0.1 breaks=km=year)	10.0
Water quality (WQ)	Poor (high impurities level)	2.0
	Fair (medium impurities level)	5.0
	Good (low impurities level)	10.0

For the fictional metadata 1000 pipe with 10 feature factors specified as above ***Table 8***. Each feature is created randomly according to the specified characteristics of the factor, and the specific expected value of that character-factor is given respectively. For example, for the age assignation, if the random assignment given as 'medium' then the age is randomly assigned between 30-70. For each factor this is applied. For the random assignment of Breakage Rate, randomness given as uniform distribution according to specified brackets.

For the performance index calculation, the final factors' average weights (W_i) taken from the article and AHP weights is selected.

Table 9: Final factors' average weights (W_i) [20]

%	Age (AG)	Material type (MT)	Size (SZ)	Installation quality (IQ)	Surface type (SR)	Ground water level (GW)	Soil type (SL)	C-factor (CF)	Breakage rate (BR)	Water quality (WQ)
AHP	9.22	9.8	8.54	11.94	8.08	11.65	11.98	10.38	9.15	9.25

Performance index of each pipe is calculated by multiplying pipeline effect values, according to their specific pipeline characteristics, and AHP weights respectively.

Table 10: Example Pipe for Performance Index Calculation

Factor	Pipe-x	Effect Values	AHP weights
AG	80	1.0	9.22
MT	Asbestos	5.5	9.8
SZ	150	3.0	8.54
IQ	Poor	2.5	11.94
SR	Asphalt	4.5	8.08
GW	Shallow	2.0	11.65
SL	Aggressive	1.5	11.98
CF	30	3.5	10.38
BR	1	1.5	9.15
WQ	Poor	2.0	9.25

For example, according to ***Table 10***, performance index of pipe-x is calculated by adding the multiplication of effect values and AHP weights and dividing the sum by 100, as 2.65.

After these calculations, to train the decision tree model, a feature column of ‘Failure Probability’ is added according to the exponential distribution when,

$$Mean = \frac{age}{break\ rate} \text{ and } Probability \approx Exp(age, mean) \tag{1}$$

Finally for binary classification a label column named ‘Failure’ is added and by giving a threshold of 0.47; it is calculated as ‘Y-yes’ if it the Failure Probability > 0.47.

4.3 Training Decision Tree Model on Metadata

In this part, decision tree model is built to classify the metadata as ‘leaking’ or ‘not leaking’.
The non-parametric supervised learning approach used for classification and regression applications is the DT. It is organized hierarchically and has a root node, branches, internal nodes, and leaf nodes. It uses a divide and conquer technique by using a greedy search to find the ideal split points inside a tree. When most or all of the records have been classified under distinct class labels, this splitting procedure is then repeated in a top-down, recursive fashion.
Although there are other approaches to choose the optimal attribute at each node, the Gini impurity and information gain methods are the two that are most frequently used as a splitting criterion in DT models. They help in assessing the effectiveness of each test condition and its capacity to categorize samples into a group.

4.3.1 Information Gain and Entropy

Without initially considering entropy, it is challenging to understand information gain. Information theory gave rise to the idea of entropy, which quantifies the impurity of sample values.
Entropy values range from 0 to 1. Entropy will be equal to zero if the data set S contains only samples that fall into a single class. Entropy will be at its peak at 1 if half the samples are assigned to one class and the other half to a different class. The characteristic with the least degree of entropy should be utilized to choose the best feature to split on and identify the best DT. Entropy before and after a split on a particular property are different, and information gain is that difference.
Since it performs the best at categorizing the training data in accordance with its target classification, the attribute with the maximum "information gain" will result in the best split.

4.3.2 Gini Impurity

Gini impurity is the likelihood that a randomly chosen data point in the dataset would be wrongly classified if it were labelled using the dataset's class distribution. Similar to entropy, the impurity of a set, S, if it belongs to one class, is zero.

The training of the metadata is done on a DT where selection criteria is set as ‘entropy’ and the maximum depth of tree is set as ‘5’.

4.4 Training CNN with thermal pipe data

The thermal camera images of pipes are fed into a Convolutional Neural Network which have 2 class as ‘Leak’ and ‘No Leak’.

A CNN consist 5 sections: input layer, convolutional layers, activation function layer, pool layer and fully connected layer.

Input Layer: The image's raw input is contained in this layer, which has the following dimensions: width 254, height 254, and depth 3.

Convolution Layer: Using the dot product between each filter and each picture patch, this layer calculates the output volume.

Activation Function Layer: The output of the convolution layer will be subjected to an element-wise activation function in this layer. The following activation functions are frequently used: Tanh, Leaky RELU, RELU: max(0, x), Sigmoid: 1/(1+e-x), etc. The volume stays the same.

Activation function of input and convolutional layers selected as ‘Relu’.

Pool Layer: This layer is regularly added to convolutional networks, and its primary purpose is to lower the volume, which speeds up computation, saves memory, and also guards against overfitting. Max pooling and average pooling are two popular varieties of pooling layers.

Fully-Connected Layer: This layer is a standard neural network layer that computes the class scores using the input from the layer before and outputs a 1-D array with a size equal to the number of classes which is 2 in our case.

The activation function of final layer with 2 dense neurons, is set as ‘sigmoid’ and ‘softmax’. A set of training cases are done. With 2 different output activation and increasing epoch number, the accuracy and loss plots are examined.

4.5 Feasibility Analysis – Leak Propagation Analysis

For the analysis of leak propagation, the set of thermal images which have time information are taken into consideration. The images firstly masked with k-means method by choosing k=2 to discriminate the leak segment. Then, the area (calculated as total pixels that are below the set threshold) of leak segment is calculated for each image by looping through the masked images and added to the ‘area list’ while adding the time of the respective image to the ‘time list’. Then, the area list and time list are fitted on linear regression. As a result, the regression plot and the regression equation are acquired.

6. PROJECT TARGETS AND SUCCESS CRITERIA

1. Selecting the most suitable method for the problem and necessary properties for the model by making a literature review on leakage detection and ML methods.
2. Determining the parameter structure with the data obtained.

For example, it is better to acquire IR image data for both isolated and not isolated case since the parameters or pre-processing of the images will be changed for the prediction model.

3. Applying image processing to the acquired data.

Normalizing the IR images to have more precise and stable temperature ranges.

4. Creating, training the prediction model.

Having successfully detected leakage areas.

5. Reaching at least 75% test accuracy on the model.
6. Interpreting the amount of leakage from the detected area on IR image.

This will depend on the knowledge comes with the IR images since there are many things that can affect interpretation, like the distance which the IR camera took the image and the area of leakage on that image is highly correlated with the amount of flow(waste) from the leak.

7. Interpreting the feasibility of fixing the pipe.

This also depends on the knowledge comes with IR image such as the age and condition of the pipe on the image.

8. Providing expected loss of the leak.
9. Validating the model.

With the images that have strong knowledge behind, the prediction model outcomes will be validated.

10. Creating an article ready for submission to an international/national journal or conference.

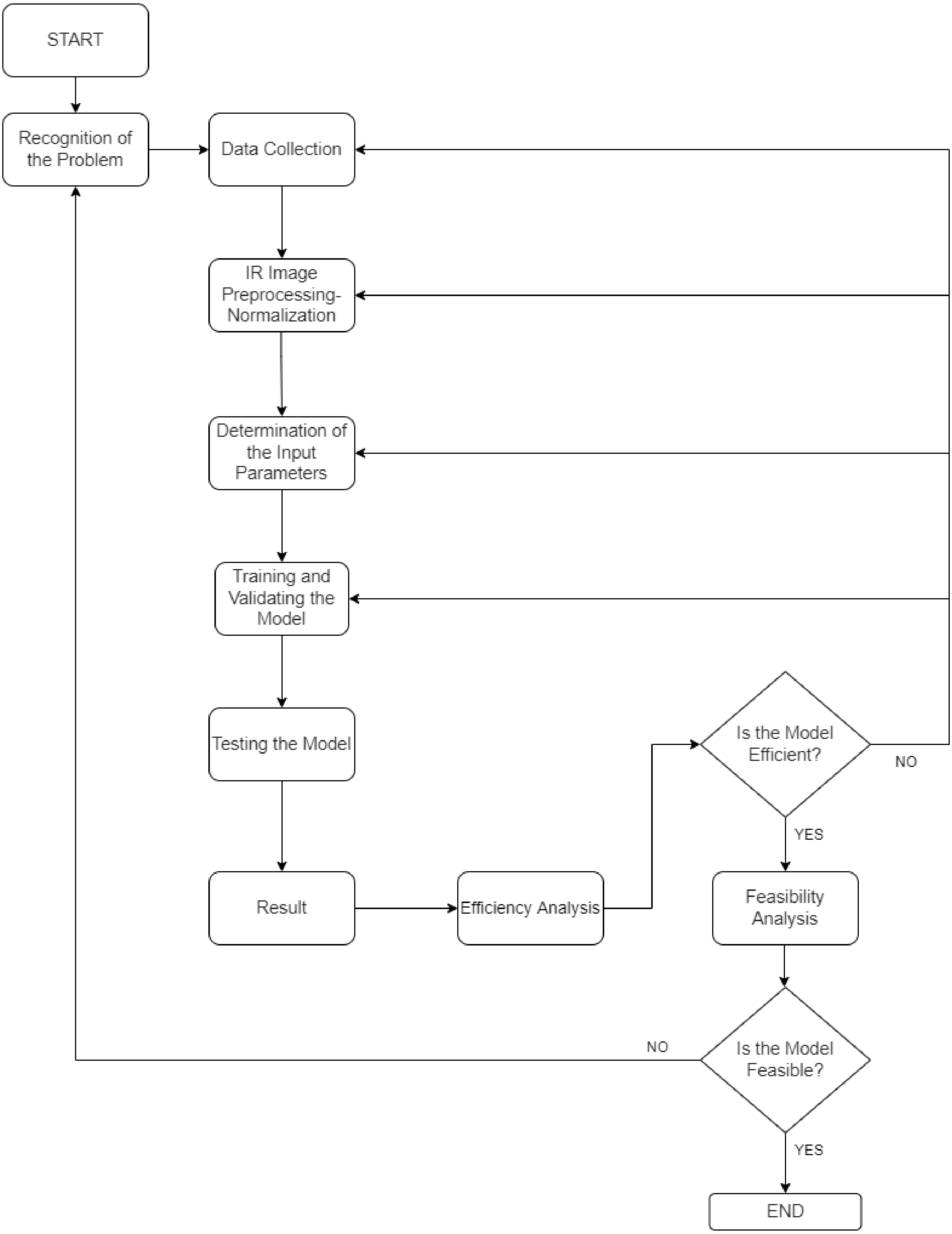

Figure 4: Flow Diagram

7. RISK AND B PLAN

Risk Management

Table 11: Risk Management

No	Biggest Risks	Risk Management (Plan B)
1	Unable to find IR image data.	Searching and requesting (if applicable) the IR image of pipes from DSİ (General Directorate of State Hydraulic Works) and BOTAŞ (Petroleum Pipeline).
2	Having no knowledge or restricted knowledge for the IR image data such as pipe condition and camera distance.	Trying to acquire the knowledge or making logical improvisions.
3	Considering the data obtained, chosen prediction method may be insufficient.	Developing innovative approaches to create a new solution.
4	Getting low test accuracy from the model.	Trying to acquire more data and treating the missing values(information). Applying feature transformation and selection methods to data. Tuning the algorithm or changing the algorithm.

Table 12: Observed Risk and the Management

No	Biggest Risks	Risk Management (Plan B)
1	Unable to find IR image data.	The risk did not occur.
2	Having no knowledge or restricted knowledge for the IR image data such as pipe condition and camera distance.	The metadata of the pipes are created by taking reference the article written by (el Chanati et al., 2016) fictionally.
3	Considering the data obtained, chosen prediction method may be insufficient.	For the leak propagation- feasibility analysis, the segmentation of leak area ,to put in the regression, is firstly tried on the method of feature extraction and RandomForest model but this did not give appropriate leak segmentation since the area of leak with time was not increasing. Therefore, before

		thresholding to calculate leak area, k_means method is used for leak area segmentation.
4	Getting low test accuracy from the model.	The risk did not occur. Both models achieved more than specified 75% accuracy success criteria that DT: 76% and CNN: 95%.

Istanbul Medipol University

School of Engineering and Natural Sciences

Graduation Project

2021-2022

8. WORK TIME PLAN OF THE PROJECT

Table 13: Work Plan

No	Work and Activity	Responsible Group Member	Timeline													
			16.week	17.week	18.week	19.week	20.week	21.week	22.week	23.week	24.week	25.week	26.week	27.week	28.week	29.week
2	**Study on the Model**									Presentation Week						Presentation Week
2.1	Literature Review on Tabular/Meta Data	Mihrimah Göknar														
2.2	Fictional Meta Data Creation	Mihrimah Göknar														
2.3	Model Selection and Training on Meta Data	Mihrimah Göknar														
2.4	Building CNN Model	Mihrimah Göknar														
3	**Thermal Images of Pipes**															
3.1	Training CNN Model on Thermal Images	Mihrimah Göknar														
3.2	Connecting 2 models for Final Result	Mihrimah Göknar														
4	**Feasibility Analysis**															
4.1	Analyzing the Leak Situation	Mihrimah Göknar														

9. DEMO PLAN

The demo will be done by testing the hybrid model. Since the methodology randomly generates the testing data set it won't be a problem of certainty and everything will be openly showed in the presentation-demo.

10. FINANCIAL EVALUATION

The requested budget (provided in **Table *1***) has been requested considering the following:

- The amount in 'people' section is equal to 10 months (project duration) salary of a computer engineer (7500TL/month).
- The amount in 'machine instrument' section is for a suitable computer for processing and training the image data: Intel i5 or AMD Ryzen 5 and above, 16GB RAM, at least 512 GB SSD, 1TB HDD for storing data, and a NVIDIA GeForce GTX 1660 GPU for large amount of data processing.
- The amount in 'service' section is for 10 months fiber internet subscription.

For this project the requested budget is not provided. Therefore, I used my own belongings.

11. RESULTS

11.1 Decision Tree Model

A sample metadata structure of pipes with the effect values is given in **Table *14*** and without effect values in **Table *15***. The final shape of the data to use in DT is arranged by encoding the categorical variables of the data which can be seen in **Table *16***.

Table 14: Sample data with respective effect values of each factor

	Age (AG)	MaterialType (MT)	Size (SZ)	Installation Quality (IQ)	SurfaceType (SR)	Ground WaterLevel (GW)	SoilType (SL)	C-Factor (CF)	BreakRate (BR)	WaterQuality (WQ)	Performance Index (PI)	FailureProb (FP)	Failure (F)
pipe1	[43, 5.0]	[Concrete, 6.5]	[184, 3.0]	[Good, 9.5]	[Asphalt, 4.5]	[Moderate, 5.5]	[Non-aggressive, 10.0]	[106, 10.0]	[0.340568568118983, 4.5]	[Fair, 5.0]	6.6031	0.288634	N
pipe2	[11, 9.5]	[Concrete, 6.5]	[370, 9.5]	[Good, 9.5]	[Unpaved, 7.5]	[Deep, 9.5]	[Non-aggressive, 10.0]	[58, 6.0]	[0.03608051795996014, 10.0]	[Good, 10.0]	8.83205	0.035437	N
pipe3	[64, 5.0]	[PVC, 9.5]	[123, 3.0]	[Poor, 2.5]	[Asphalt, 4.5]	[Deep, 9.5]	[Aggressive, 1.5]	[159, 10.0]	[0.05534984937939961, 10.0]	[Fair, 5.0]	6.01225	0.053846	N
pipe4	[53, 5.0]	[Asbestos, 5.5]	[420, 9.5]	[Poor, 2.5]	[Unpaved, 7.5]	[Deep, 9.5]	[Moderate, 5.0]	[196, 10.0]	[0.5564472807172127, 1.5]	[Fair, 5.0]	6.0593	0.426758	N
pipe5	[47, 5.0]	[Asbestos, 5.5]	[403, 9.5]	[Fair, 6.0]	[Foot path, 7.0]	[Moderate, 5.5]	[Aggressive, 1.5]	[19, 3.5]	[0.35507026627865534, 4.5]	[Good, 10.0]	5.6138	0.298876	N

Table 15: Sample of pipe metadata

	Age (AG)	MaterialType (MT)	Size (SZ)	Installation Quality (IQ)	SurfaceType (SR)	Ground WaterLevel (GW)	SoilType (SL)	C-Factor (CF)	BreakRate (BR)	WaterQuality (WQ)	Performance Index (PI)	FailureProb (FP)	Failure (F)
pipe1	43	Concrete	184	Good	Asphalt	Moderate	Non-aggressive	106	0.340569	Fair	6.6031	0.288634	N
pipe2	11	Concrete	370	Good	Unpaved	Deep	Non-aggressive	58	0.036081	Good	8.83205	0.035437	N
pipe3	64	PVC	123	Poor	Asphalt	Deep	Aggressive	159	0.05535	Fair	6.01225	0.053846	N
pipe4	53	Asbestos	420	Poor	Unpaved	Deep	Moderate	196	0.556447	Fair	6.0593	0.426758	N
pipe5	47	Asbestos	403	Fair	Foot path	Moderate	Aggressive	19	0.35507	Good	5.6138	0.298876	N

Table 16: Encoded metadata

	Age (AG)	Size (SZ)	C-Factor (CF)	BreakRate (BR)	PerformanceIndex (PI)	FailureProb (FP)	MaterialType (MT)	InstallationQuality (IQ)	SurfaceType (SR)	GroundWaterLevel (GW)	SoilType (SL)	WaterQuality (WQ)	Label
pipe1	43	184	106	0.340569	6.6031	0.288634	2	1	0	1	2	0	0
pipe2	11	370	58	0.036081	8.83205	0.035437	2	1	3	0	2	1	0
pipe3	64	123	159	0.05535	6.01225	0.053846	4	2	0	0	0	0	0
pipe4	53	420	196	0.556447	6.0593	0.426758	0	2	3	0	1	0	0
pipe5	47	403	19	0.35507	5.6138	0.298876	0	0	1	1	0	1	0

The first phase of the hybrid model consists of DT pre-decision of leaking by looking at metadata. The decision steps of the final trained DT can be examined in **Figure 6**.
The confusion matrix of DT model and its evaluation metrics are given below.

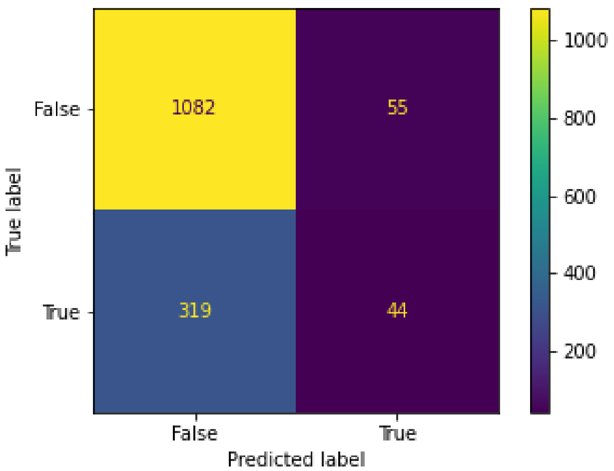

Figure 5: Confusion matrix of DT model

Accuracy measures how often the model is correct.

$$\frac{\text{True Positive} + \text{True Negative}}{\text{Total Predictions}} \tag{2}$$

Precision shows number of the positives predicted, what percentage is truly positive?

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{3}$$

Sensitivity (Recall) shows number the positive cases, what percentage are predicted positive?

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{4}$$

F1 Score: Harmonic mean of Precision and Recall

$$\frac{\text{Precision} + \text{Recall}}{2 * Precision * Recall} \tag{5}$$

Table 17: Metrics of DT

Accuracy	0.7506
Precision	0.4444
Sensitivity	0.1212
F1 Score	0.1905

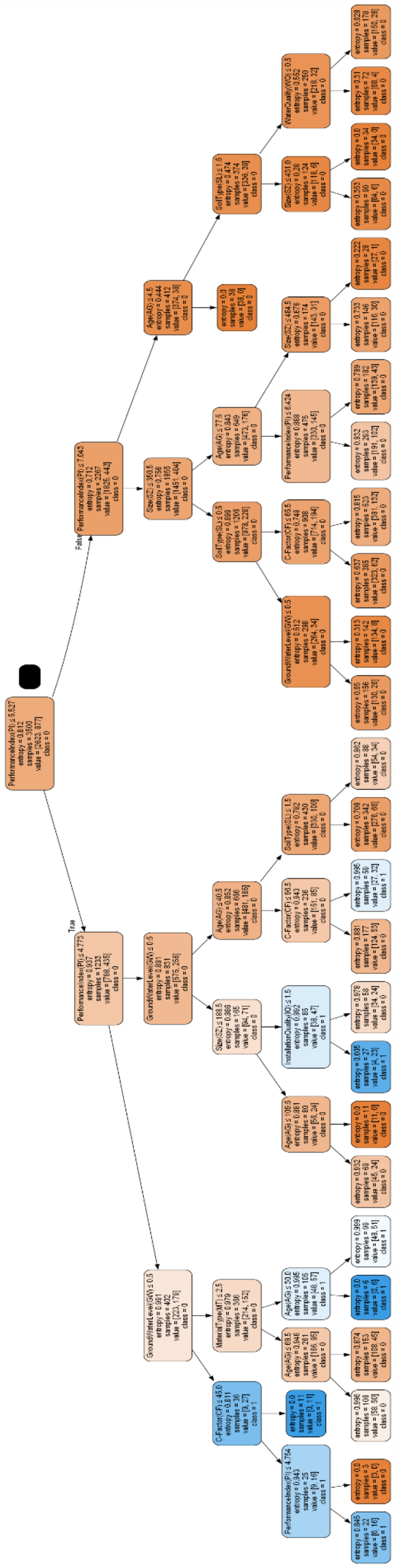

Figure 6: Resolving Steps of Trained Decision Tree

11.2 Convolutional Neural Networks Model

Before evaluating the leak results of DT, CNN model will be explained.
The CNN is built by using TensorFlow.keras library's Sequential model. The model summary is given below,

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 254, 254, 32)      896

 max_pooling2d (MaxPooling2D  (None, 127, 127, 32)     0
 )

 conv2d_1 (Conv2D)           (None, 125, 125, 32)      9248

 max_pooling2d_1 (MaxPooling  (None, 62, 62, 32)       0
 2D)

 conv2d_2 (Conv2D)           (None, 60, 60, 32)        9248

 max_pooling2d_2 (MaxPooling  (None, 30, 30, 32)       0
 2D)

 conv2d_3 (Conv2D)           (None, 28, 28, 16)        4624

 max_pooling2d_3 (MaxPooling  (None, 14, 14, 16)       0
 2D)

 flatten (Flatten)           (None, 3136)              0

 dense (Dense)               (None, 256)               803072

 dense_1 (Dense)             (None, 2)                 514

=================================================================
Total params: 827,602
Trainable params: 827,602
Non-trainable params: 0
```

Figure 7: CNN Model Summary

The data used for training consist of 2 class as 'Leak' and 'No Leak'. Therefore, finale layer of the CNN is dense layer with 2 neurons.

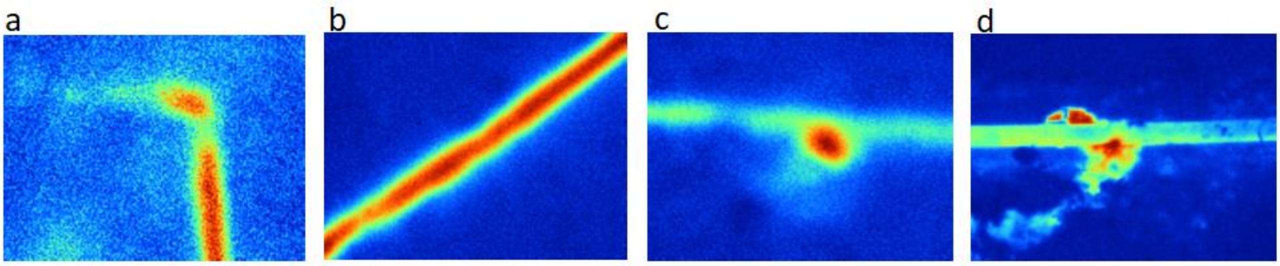

Figure 8: Example Thermal Image data. a-b is No Leak, c-d Leak cases

The images are which show uncertainty and absurd pixel noises are removed from the dataset and with image augmentation techniques the image number is increased. Final training data example is given below figure. [0,1] encoding is the ‘No Leak’ case while [1,0] encoding is the ‘Leak’ case.

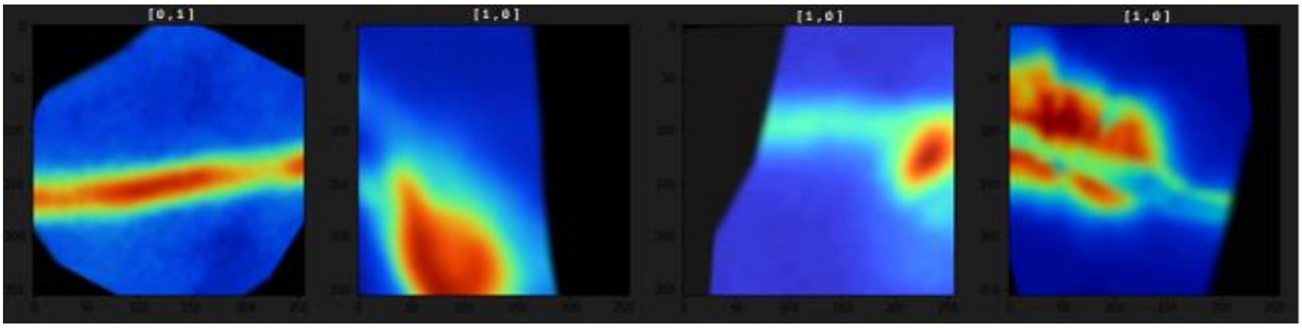

Figure 9: Training Data

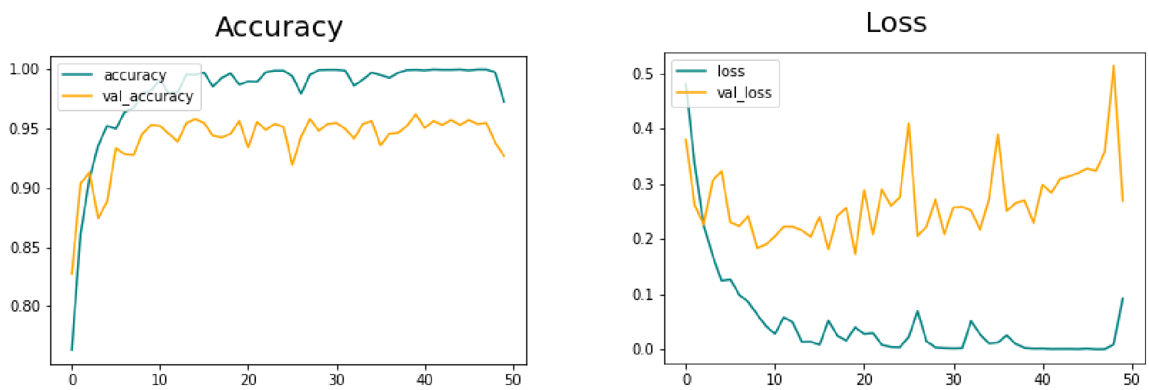

Figure 10: Accuracy and Loss Plot for 50 epochs

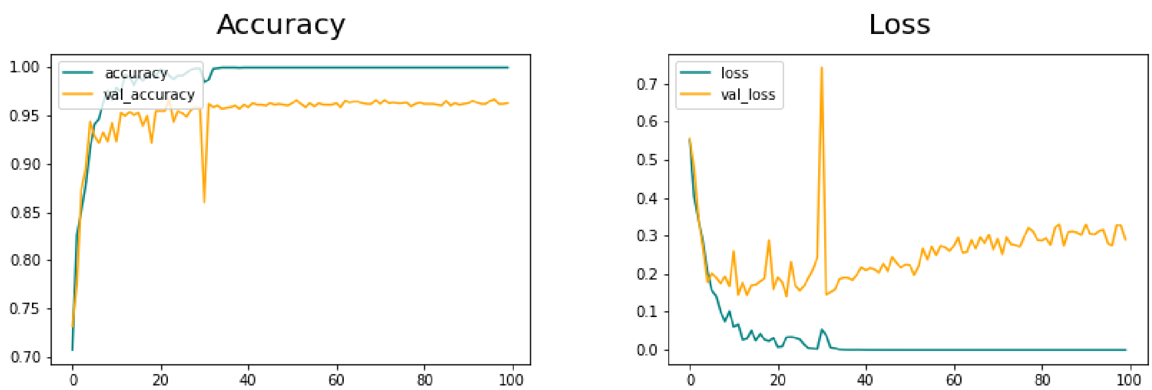

Figure 11: Accuracy and Loss Plot for 100 epochs

The metrics for CNN 100 epochs are:

- Precision: 0.9557
- Recall (Sensitivity): 0.9589
- Accuracy: 0.9572

The leaking classification is decided with a simple method that compares the probabilities of 2 classes and present the result as 'Leaking' or 'Not Leaking'.

11.3 Hybrid Model

After the designation and training of both models separately, the connection is made according to **Figure 12.**

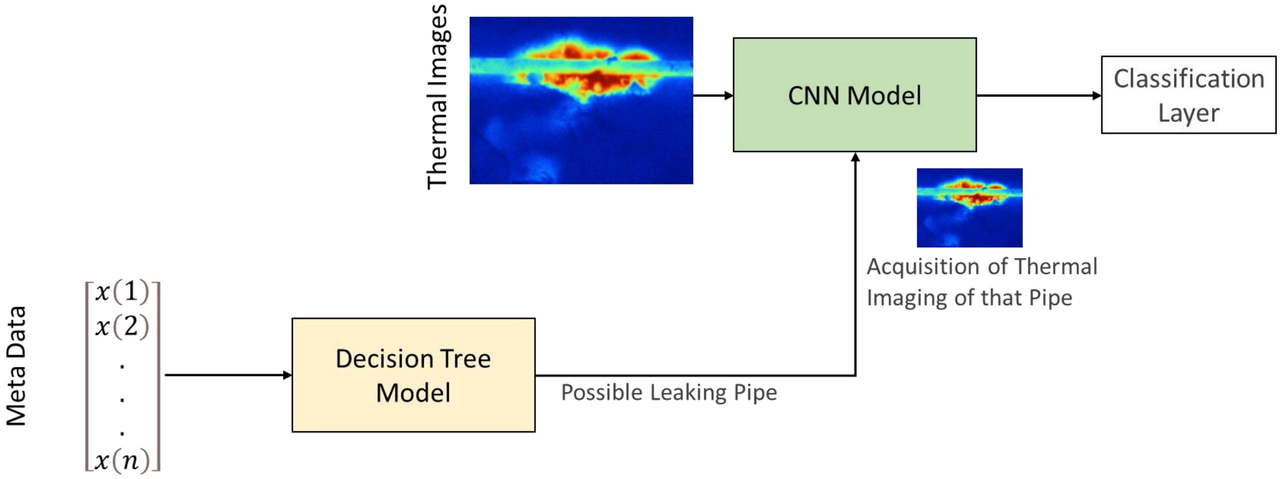

Figure 12: Hybrid Model Visualization

Firstly, the metadata of the pipes in a water distribution network will be sent to DT model and by predicting every pipe one by one it will check the pipe leakage condition if the DT model predicts as 'Leaking' then the thermal image of that specific pipe will be fed into CNN model for further prediction. This way 2 -Factor Classification will be made the correctness of the system will be increased.

For the testing, a network with 500 pipes is created and then fed into hybrid model. 35 out of 500 pipes is predicted as leaking in the DT model and then the thermal images of these pipes fed into CNN model which predicted '25' out of these 35 images as 'leaking'. The reason that 9 images are predicted as 'not leaking' is that the images are randomly assigned from the 5000 images which contains both cases.

Some of the pipes' testing result is given below.

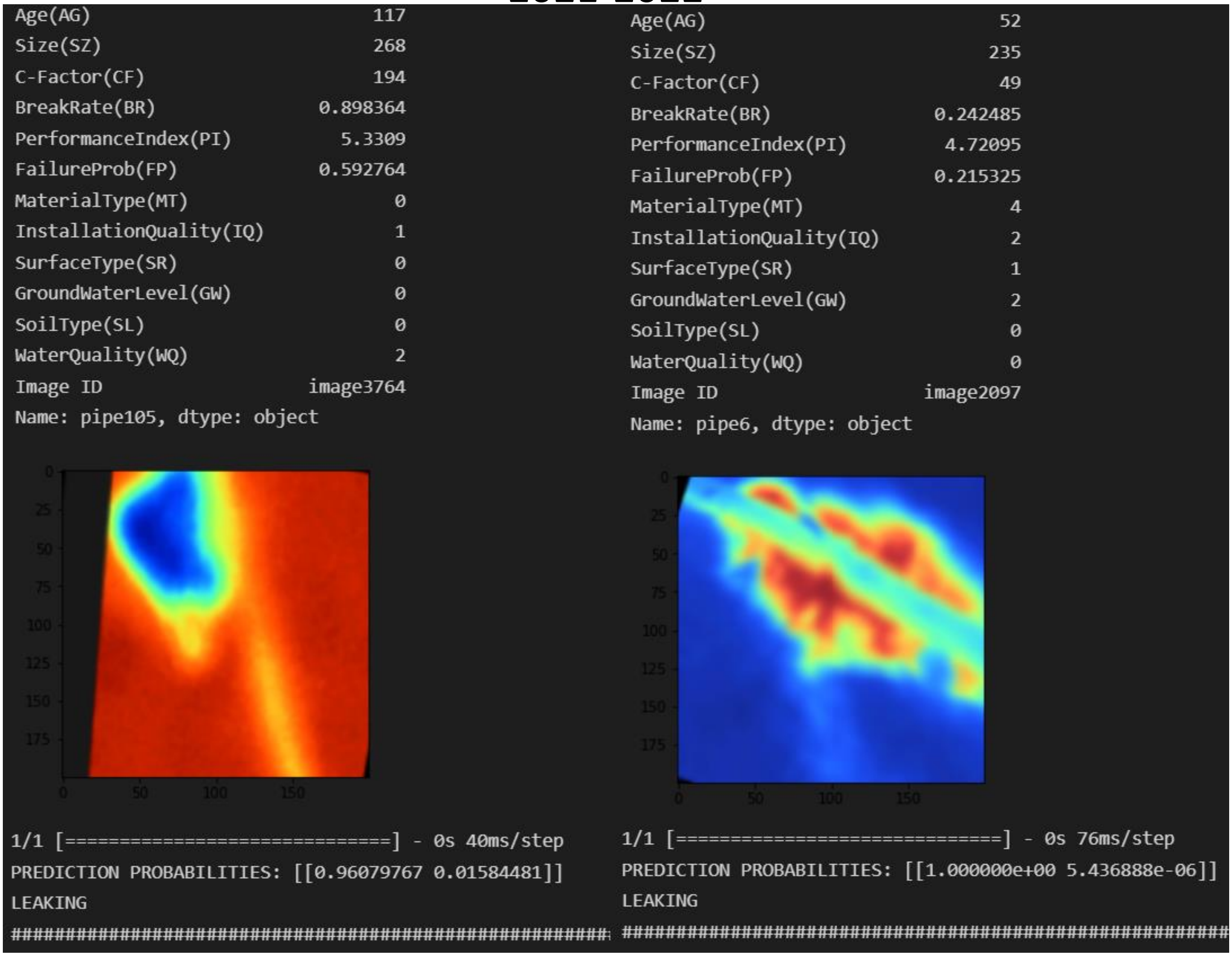

Figure 13: Sample output of leaking case

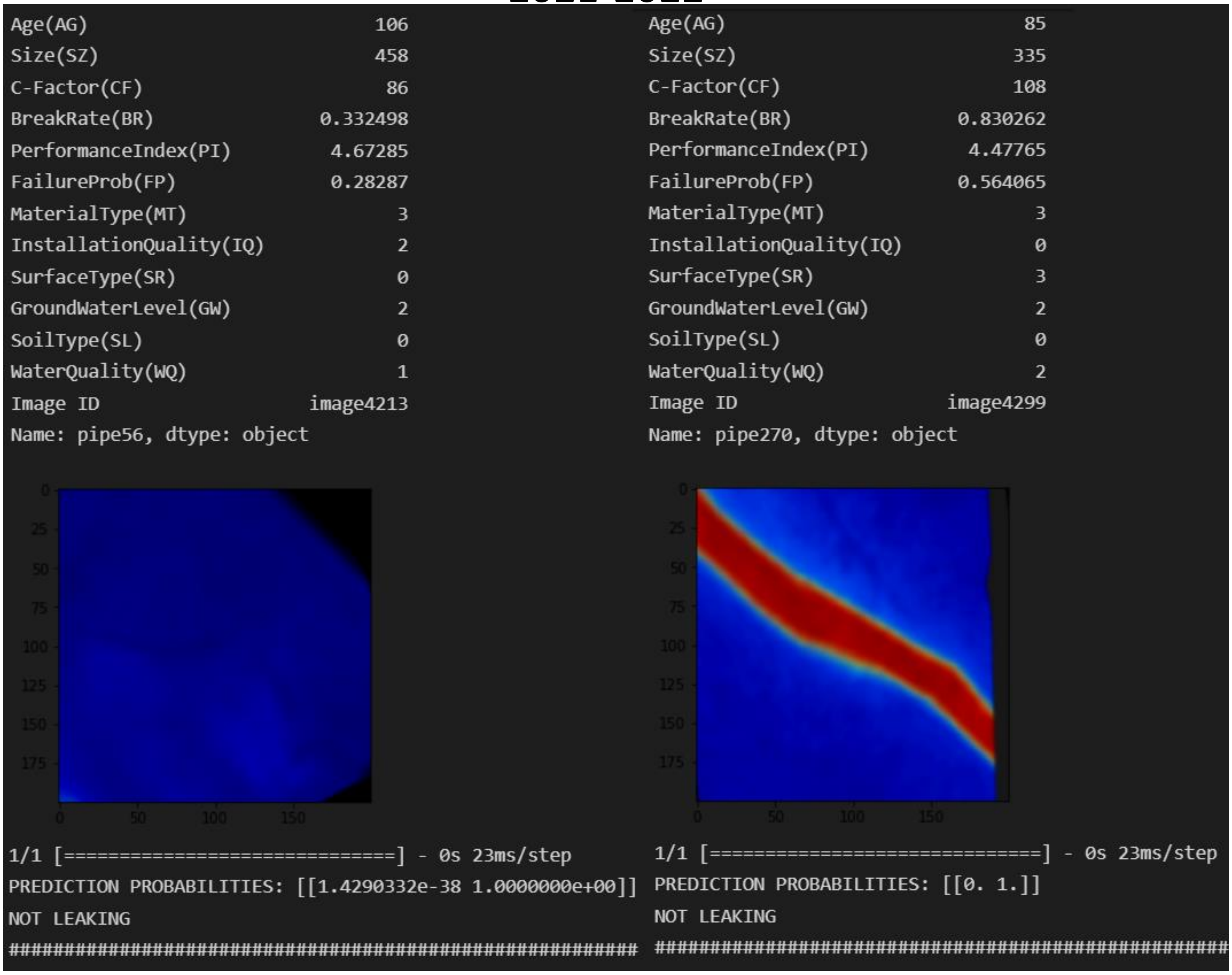

Figure 14: Sample output of not leaking case

One thing to note, in the real world the thermal images of pipes won't be available immediately. The images of the specific pipe will be collected through a thermal camera-built drone and then it will be sent to CNN model. For further analysis, if it is predicted as leaking then the thermal camera can continue taking snaps with time intervals to approximate leak propagation which will be analyzed in the Section **11.4**.

11.4 Feasibility analysis- Leak Propagation

For this thermal data set, a set of images have the time information. Therefore, a leak propagation analysis is made. For the nine images, k-means method is used to initially mask the images in order to identify the leak segment by selecting k=2. The area of the leak segment is then computed for each image by looping over the masked images and added to the 'area list' while the time of the corresponding image is added to the 'time list'. The area of the leak segment is determined as total pixels that are below the set threshold. The k-means applied leak masks and then thresholded images for removing outlier leak areas and other pixel distortions for a clear area calculation can be seen in **Figure 15**.

Following that, linear regression is fitted when the dependent variable is area and independent list is time. The regression plot **Figure 16**, and regression equation are also obtained as in below.

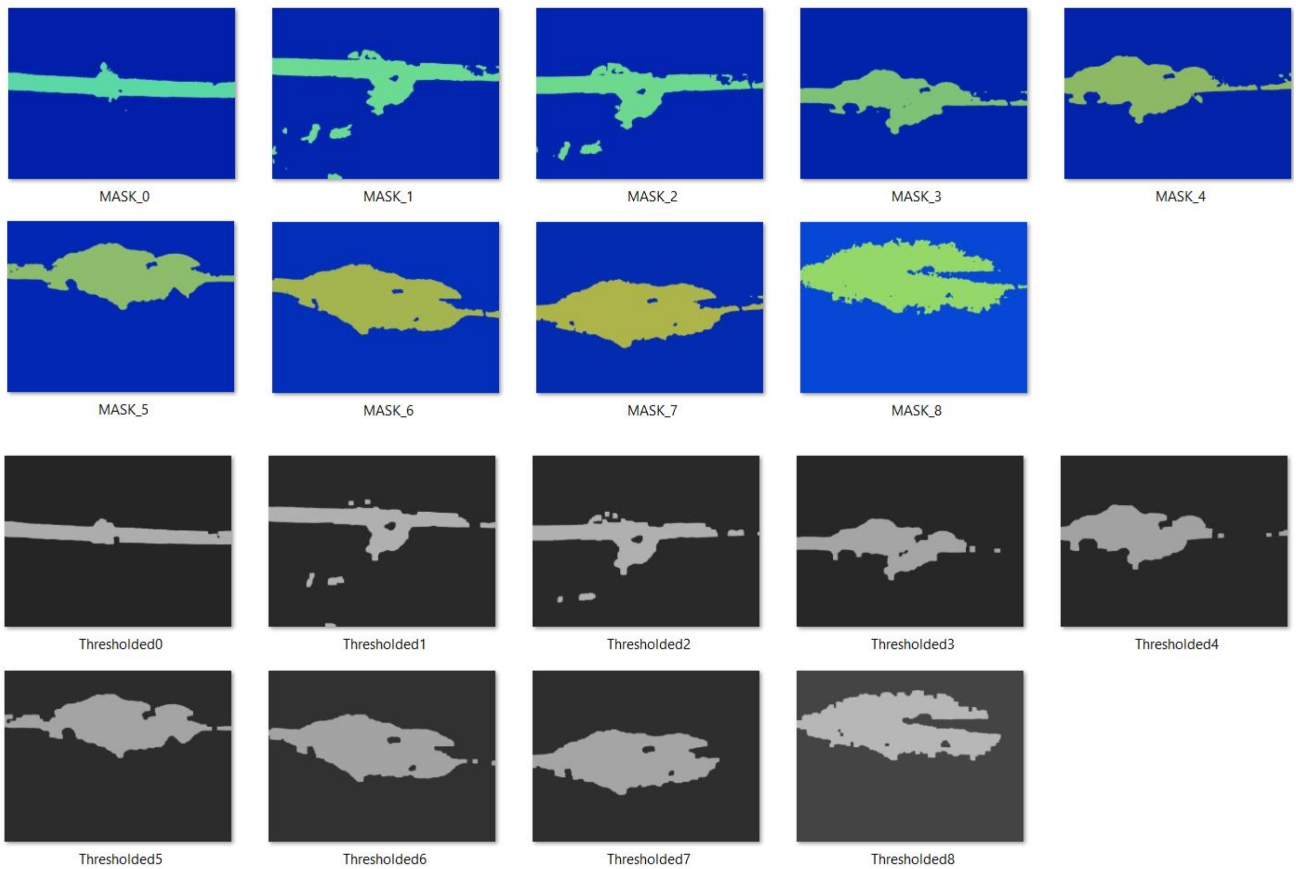

Figure 15: K-means of Leak Images and Thresholded Images

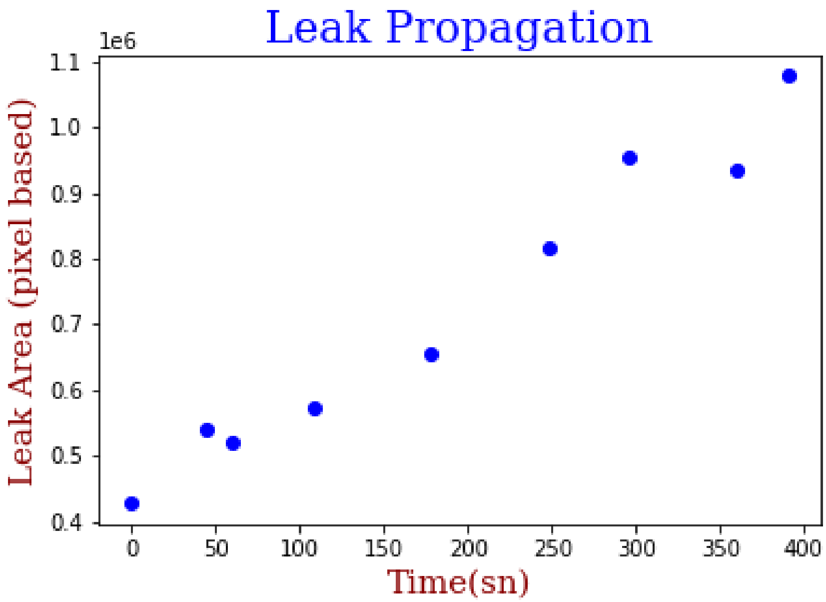

Figure 16: Leak Propagation Plot

Regression equation is obtained as below and 96.7% of the variance in area is explained by time.

- $y = 1577.5702137787778\ x + 427308.4655314177$
- $R^2 = 0.9671932465177697$

12. DISCUSSION

The accuracy scores of both models achieved higher than the goal set in the success criteria. But a method to calculate the overall accuracy of the hybrid model couldn't be found. Therefore, the accuracy criteria scores are evaluated separately. In addition, the feasibility assessment was done up to a content which was limited to leak propagation assessment since the information behind the leak could not be obtained such as water loss or water pressure drop and current of the water. Therefore, the evaluation of water loss and economic loss calculation could not be done. Overall, the success criteria of this project are met 90%. The 10% lack is due to not being able to identify the leak amount. The reason for this is the required information of spent water in the leak was missing. Therefore, the association with spent water and leak area couldn't be achieved.

13. CONCLUSION

If this project could be adapted to real world water, gas and other hydrocarbon fluid networks, it will not only contribute economically (reducing the cost for digging and changing, repairing the pipes), to several institutions that operate and control buried pipes, but also ensure that leaks in pipes are detected early, without the need for open inspection for pipes, and before they become harmful to the environment and humanity.

14. PLAN FOR FUTURE STUDIES

In real life, usually the background information of pipes in a network is not exist or partially exist. Therefore, it is hard to obtain these results for old networks but for new and future networks this information can be kept and updated regularly. For future research, the evaluation of pipes in the hybrid model can be improved by changing some criteria and may be by advancing and changing the model. In addition, as said above in the results section, the real-world implementation of this hybrid model does not keep the thermal images of pipes in the network; it should take the images with a thermal camera implemented drone. Therefore, future studies can focus on this matter.

15. ASSESSMENT OF ENGINEERING COURSES

The courses that I've taken advantage of are Introduction to Machine Learning, Introduction to Deep Learning, Decision Analysis (In the IE education) as technical electives and all the other mandatory lectures have helped me for finding the problem, planning the process, conducting and building the project to some content.

16. REFERENCES

1. Liu, Y., Ma, X., Li, Y., Tie, Y., Zhang, Y., & Gao, J. (2019). Water Pipeline Leakage Detection Based on Machine Learning and Wireless Sensor Networks. Sensors, *19(23)*.
2. Adegboye, M. A., Fung, W.-K., Karnik, A. (2019). Recent Advances in Pipeline Monitoring and Oil Leakage Detection Technologies: Principles and Approaches. Sensors, *19(11)*.
3. Wada, Y. et al. (2016) Modelling global water use for the 21st century: The Water Futures and Solutions (WFaS) initiative and its approaches. *Geosci. Model Dev. 9*, 175–222.
4. Coelho, J. A., Glória, A., Sebastião, P. (2020). Precise Water Leak Detection Using Machine Learning and Real-Time Sensor Data. IoT, 1(2), 474–493.
5. Idachaba, F., Rabiei, M. (2021). Current technologies and the applications of data analytics for crude oil leak detection in surface pipelines. Journal of Pipeline Science and Engineering, *1(4)*, 436–451.
6. Manekiya, M. H., Arulmozhivarman, P. (2016). Leakage detection and estimation using IR thermography, 2016 International Conference on Communication and Signal Processing (ICCSP),1516-1519.
7. Alvarado, E. P. C., Cabrera, D. A., García, R. P., Sebastián, J. I. (2014). Identification of Buried Pipes Using Thermal Images and Data Mining. Procedia Engineering, *89*, 1445–1451.
8. Shen, G., Li, T. (2007). Infrared thermography for high temperature pressure pipe. Insight. *Volume 49*, 151-153.
9. Jadin, M. , Ghazali, K. (2014) Gas Leakage Detection Using Thermal Imaging Technique. 16th International Conference on Computer Modelling and Simulation (UKSim), 302-306.
10. Dudic, S., P., Ignjatovic, I., M., Seslija, D., D., Blagojevic, V., A., Stojiljkovic, M., M. (2012). Leakage quantification of compressed air on pipes using thermovision. Thermal Science, *16*,555–565.
11. Sosnowski, T., Bieszczad, G., Madura, H. (2018) Image Processing in Thermal Cameras. Studies in Systems, Decision and Control, *106*, 35-57.
12. E. P. Carreño-Alvarado, D. Ayala-Cabrera, R. Pérez-García, and J. Izquierdo, (2014) Identification of buried pipes using thermal images and data mining, Proc. Eng., vol. 89, pp.

1445–1451.
13. Pauline, E.; Carreño-Alvarado; Meza, G.R. (2020). Water Leak Detection by Termographic Image Analysis, In Laboratory Tests. *Proceedings* , *48*, 15.
14. Ogutu, G., Okuthe, P.K., Lall, M. (2017). A review of probabilistic modeling of pipeline leakage using Bayesian Networks. J. Eng. Appl. Sci., *12*, 3163–3173.
15. Mashhadi, N., Shahrour, I., Attoue, N., Aljer, A. (2021). Use of Machine Learning for Leak Detection and Localization in Water Distribution Systems. Smart Cities,*4*, 1293-1315.
16. Giraldo, M., & Sánchez, J. P. R. (2020). Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks. *12*.
17. Hossain, K., Villebro, F., & Forchhammer, S. (2019). Leakage Detection in District Heating Systems Using UAV IR Images: Comparing Convolutional Neural Network and ML Classifiers. 27th European Signal Processing Conference European Association for Signal Processing (EURASIP), *140*, 158-164.
18. Hoang, N.D. and Tran, V.-D. (2019). Image Processing Based Detection of Pipe Corrosion Using Texture Analysis and Metaheuristic Optimized Machine Learning Approach. Computational Intelligence and Neuroscience.
19. Fan, X., Zhang, X. & Yu, X..B. (2021). Machine learning model and strategy for fast and accurate detection of leaks in water supply network. J Infrastruct Preserv Resil *2*, 10.
20. Chanati, H. E., El-Abbasy, M. S., Mosleh, F., Al-Derham, H. (2015). Multi-Criteria Decision Making Models for Water Pipelines. American Society of Civil Engineers, *30*,4.
21. S. A. Thomas, (2021). Combining Image Features and Patient Metadata to Enhance Transfer Learning, 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2660-2663.
22. Shabangu , T. H. , Hamam , Y. , Adedeji, K. B. (2020). Decision support systems for leak control in urban water supply systems: A literature synopsis. (n.d.). Procedia CIRP, *90*, 579–583.
23. Saaty, T. L. (1980). The analytic hierarchy process, McGraw-Hill, New York.

17. PROJECT ACTIVITIES AND WORK PLAN

WP No	Detailed Definition of Work and Activity
1	By making a literature review on leakage detection methods and machine learning methods; the most suitable method will be determined for our problem.
2	The data obtained after the literature review will determine the prediction model input and parameter structure for the model. These data will be used for training and design of the model in the future.
3	The image pre-processing on IR images will be done.
4	The prediction model will be built, and training will be done.
5	Interpretation of amount of leakage from the detected area on IR images.
6	Doing feasibility analysis and finding the expected results.

Work package	Target	Measurable outcome	Contribution to overall success (%)
1	Literature Review	Getting Deeper Knowledge on the Model and Leakage Detection	%15
2	Acquiring IR data and determination of model and parameter inputs	Acquisition of data	%15
3	Applying image pre-processing	Preparation of data for model	%5
4	Identifying, training, and designing the appropriate model	Validating the Model	%30
5	Interpretation of leak amount	An important outcome	%25
6	Finding the expected results	Feasibility Analysis	%10

WORK PACKAGE DISTRIBUTION						
Project Member	**WP1**	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**
Mihrimah Göknar	100	100	100	100	100	100
Total	100	100	100	100	100	100

18. CURRICULUM VITAE

Mihrimah Göknar-CV

19. SUPPORT LETTERS (if any)