**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| PRELIMINARIES |  | Page |
| Table of Contents |  | i |
| List of Tables |  | ii |
| List of Figures |  | iii |

[**Chapter 1** 5](#_Toc78489704)

[**Introduction** 5](#_Toc78489705)

[Background of the Study 5](#_Toc78489706)

[Objectives of the Study 7](#_Toc78489707)

[General Objective: 7](#_Toc78489708)

[Specific Objective: 7](#_Toc78489709)

[Scope and Limitation of the Study 7](#_Toc78489710)

[Significance of the Study 8](#_Toc78489711)

[**Chapter 2** 9](#_Toc78489712)

[**Conceptual Framework** 9](#_Toc78489713)

[Review of Related Literature and Studies 9](#_Toc78489714)

[Petrographic Thin Section 9](#_Toc78489715)

[Petrographic Thin Section Analysis 12](#_Toc78489716)

[Machine Learning 15](#_Toc78489717)

[Learning Types 15](#_Toc78489718)

[Machine Learning Algorithms for Image Classification 16](#_Toc78489719)

[Database 19](#_Toc78489720)

[Programming 19](#_Toc78489721)

[Networking 20](#_Toc78489722)

[Systems Analysis and Design 21](#_Toc78489723)

[SDLC models 23](#_Toc78489724)

[Steps in SDLC 23](#_Toc78489725)

[Software Engineering 26](#_Toc78489726)

[Evaluation System 27](#_Toc78489727)

[Software Requirements 28](#_Toc78489728)

[SQLite 30](#_Toc78489729)

[Python 30](#_Toc78489730)

[Hardware Requirements 35](#_Toc78489731)

[Conceptual Model of the Study 40](#_Toc78489732)

[Operational Definition of Terms 43](#_Toc78489733)

[**Chapter 3** 45](#_Toc78489734)

[**Methodology** 45](#_Toc78489735)

[Project Design 45](#_Toc78489736)

[Project Development 52](#_Toc78489737)

[Phase 1 53](#_Toc78489738)

[Phase 2 53](#_Toc78489739)

[Phase 3 53](#_Toc78489740)

[Phase 4 53](#_Toc78489741)

[Phase 5 54](#_Toc78489742)

[Phase 6 54](#_Toc78489743)

[Operation and Testing Procedures 54](#_Toc78489744)

[Operation Procedure 54](#_Toc78489745)

[Testing Procedure 54](#_Toc78489746)

[Evaluation Procedure 55](#_Toc78489747)

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| Table | Title | Page |
| 1 | Test Procedure for Optimizing Processing Time | 55 |
| 2 | Test Procedure for Reliable and Accurate Output | 56 |
| 3 | Five-Point Likert Scale | 57 |
| 4 | Scale Range and Its Descriptive Rating | 58 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Figure | Title | Page |
| 1 | A thin section on a petrographic microscope stage | 10 |
| 2 | Thin Section | 11 |
| 3 | Standard Thin Section Components | 11 |
| 4 | Petrographic Thin Section Images | 12 |
| 5 | Rank of Popular Programming Language | 32 |
| 6 | Conceptual Model of the Study | 40 |
| 7 | Petrographic Thin Section Analysis Workflow | 47 |
| 8 | Context Diagram of Petrographic Thin Section Analyzing System | 48 |
| 9 | Petrographic Thin Section Analyzing System with its Basic Subprocesses | 49 |
| 10 | Petrographic Thin Section Analyzing System | 50 |
| 11 | Flow Chart of Building Machine Learning Analytical Model | 52 |
| 12 | Project Development Phases | 55 |

# **Chapter 1**

# **Introduction**

## Background of the Study

Petrographic thin section is a type of laboratory preparation of a sliced portion of rock with thickness of 0.02 - 0.03 millimeter placed between clear glass plates. It is commonly used in geology, mineralogy, or petrography. Typically, a petrographic thin section is used to analyze a sample of rock. Several methods are used to perform an analysis on a petrographic thin section and the most important and widely used is the manual method performed by a geologist ([arXiv:2003.10437v1](https://arxiv.org/abs/2003.10437v1) [eess.IV]). Skilled and experienced petrographers use a microscope to examine the thin section to classify rocks, identify minerals, and extract essential pore information. These experts using equipment and methods whose task involves manual process, whereas individual observation may vary, could possibly lead to misclassification by the observer (Albar, A. Osman, M.H., Abdullah, M.S. and Ismail, B.N., 2013).

The oil and gas industry has a huge benefit from rock classification as a result from analyzing a petrographic thin section. Thin section images can provide abundant petrographic information but a common issue in petrograph is that it calls for long man hours and can be a source of interpretation error. It has been foreseen that there is a need to use computer algorithms thatwill facilitate the analysis of uninterpreted images that have been neglected due to a limited number of experts.

Machine learning is a subfield of artificial intelligence that trains a machine how to learn. The main point of machine learning is to automate a task by turning sample or historical data, commonly called “training data”, into an analytical model. Machine learning has been used in several industries to perform image processing, prediction, and classification. Attempts have been made to clearly explain in an automated way the rock and mineral information from the thin section image by using computer algorithms. This approach demonstrates enhanced efficiency, accuracy, and objectivity when compared to the traditional manual analysis ([arXiv:2003.10437v1](https://arxiv.org/abs/2003.10437v1) [eess.IV]).

Petrographers handling thin sections are challenged by laborious tasks and interpretation errors during the analysis. In this light, the need to create a system that will make the process of analyzing thin section images faster and more accurate. Machine learning algorithms and a reasonable amount of training data can develop an efficient analytical model. Studies conducted to analyze petrographic thin section images using machine learning have positive results. Developed machine learning models and available algorithms in image preprocessing can be used to train and continuously improve the petrographic thin section analytical model for this project.

## Objectives of the Study

### General Objective:

The main objective of the study is to develop a system that will aid in analyzing petrographic thin section of a rock sample using machine learning.

### Specific Objective:

1. To design the system with the following functionalities.

a. System with optimized processing time

b. Reliable and accurate output

2. To create a system as designed.

3. Test the performance of the system.

4. To evaluate the system performance based on ISO 25010 Evaluation of Software/System criteria.

## Scope and Limitation of the Study

The study will focus on developing a system for petrographic thin section analysis. The project will be using petrographic thin section samples images to build an analytical model that will be used for analysis. The development of the analyzing system will focus on the rock classification and only on the specific identification of features or information of a thin section that have big challenge to the process of analyzing thins section samples.

The development of the project will require a large amount of data to build a reliable model which will be determined during the training, validating and testing process. Public data will be used during the development phase of the system due to the high confidentiality nature of the data where this study will be applied.

## Significance of the Study

This part of the study will present the importance of the project and the benefits from its result. The development of analyzing system for petrographic thin section using machine learning will have a positive impact in the field of petrology and technology of computer vision.

Petrographic thin section images can provide information that is immensely beneficial but requires long man hours of work doing the analysis. Thus, petrographers will benefit from this project that will enhance the process of petrographic thin section analysis. Organizations that perform petrography can also benefit from this study considering the limited numbers of skilled and experienced petrographers and expensive equipment.

The study can also contribute to the exploration and reservoir characterization in the oil and gas industry facing challenges on the increasing number of petrograph samples that requires consistent and efficient analysis.

# **Chapter 2**

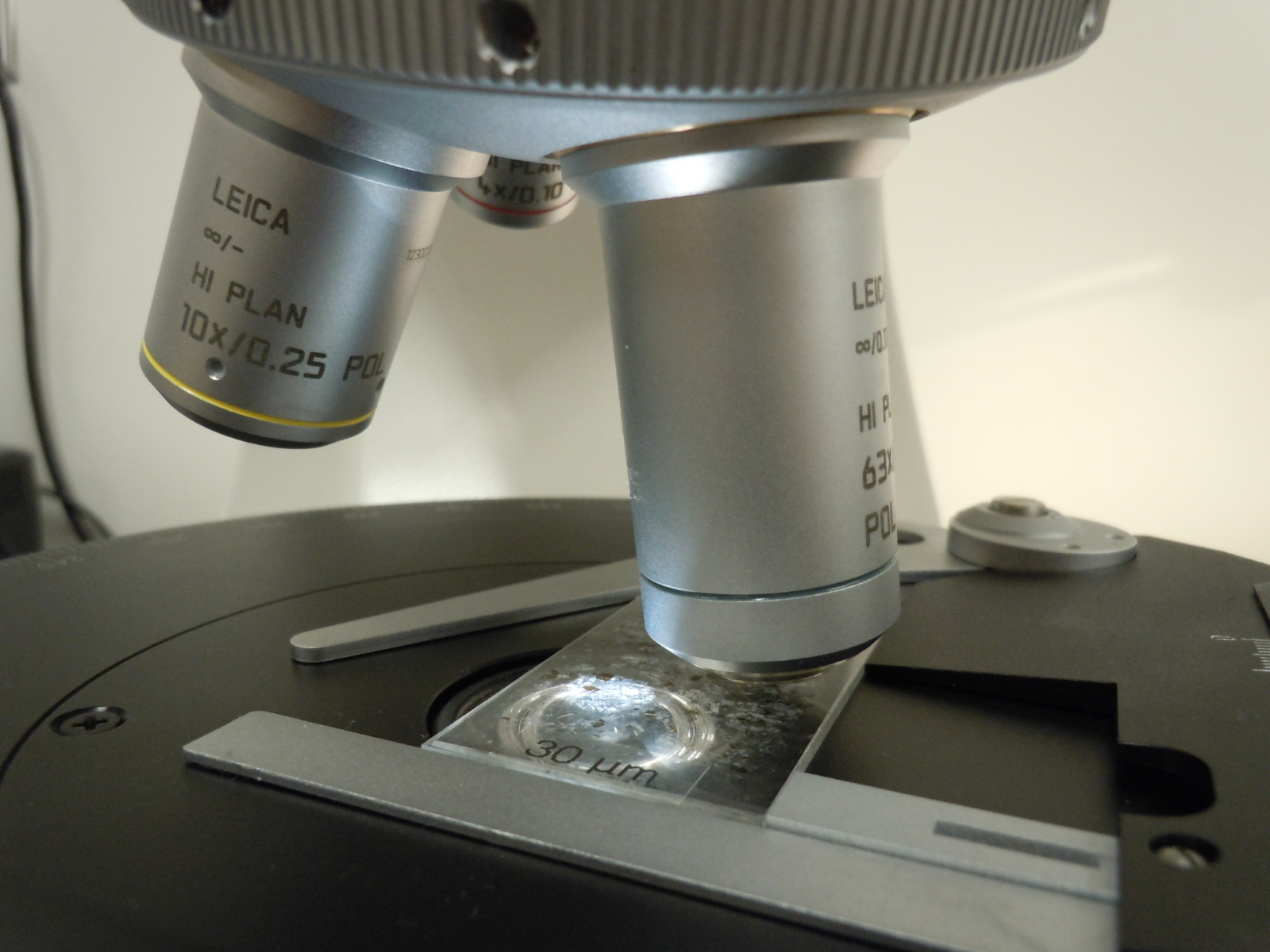
# **Conceptual Framework**

## **Review of Related Literature and Studies**

### Petrographic Thin Section

A petrographic thin section is a sliced portion of rock 0.02 to 0.03 millimeters thick, cemented for study between clear glass plates (Dictionary, Merriam-Webster). Petrography is the description and systematic classification of rocks. Petrographic thin section is a laboratory preparation of a mechanically ground optically flat fragment or slice of rocks to a thickness of 0.03 millimeters (30 µm).

Petrology is a scientific study of rocks that deals with their composition, texture, and structure; their occurrence and distribution; and their origin in relation to physicochemical conditions and geologic processes (Encyclopedia Britannica, 2018). It is concerned with all three major types of rocks, igneous, metamorphic, and sedimentary. Petrology includes the subdisciplines of petrography and experimental petrology. Experimental petrology involves the laboratory synthesis of rocks for the purpose of ascertaining the physical and chemical conditions under which rock formation occurs. Petrography is the study of rocks in thin section by means of a petrographic microscope (i.e., an instrument that employs polarized light that vibrates in a single plane). Petrography is primarily concerned with the systematic classification and precise description of rocks. A person who studies petrography is called a petrographer. Thin section petrography started in the later part of the 18th century. A Scottish scientist Willian Nicol created the first thin sections by producing samples of fossilized wood that were thin enough to be transparent under a microscope. He then examined the sections to determine the species of tree from which they originated (Sorby 1882, 101; Humphries 1992, 2; Croft 2006, 20, 33). Nicol also contributed to the creation of the first polarizing microscope in 1828 by inventing a prism that enabled the transmission of a single beam of plane-polarized light. Two of these “Nicol” prisms were first applied to a microscope in 1834 (Sarah E. Peterson and Philip P. Betancourt, 2009).



Source: https://viva.pressbooks.pub/petrology/chapter/thin-sections/

Figure 1.0 A thin section on a petrographic microscope stage.

Preparation of petrographic thin section of rock are not done in the same way but the commonly used are the standard thin section and polished thin section. It is mounted between a thin glass slide and mostly translucent to transparent so that optical properties can be studied under a microscope. The general procedure for creating thin sections has remained nearly the same since the modern version was created nearly 200 years ago (Elizabeth Johnson, Juhong Christie Liu, and Mark Peale, n.d.).

A standard thin section is generally fixed with epoxy glue between a glass slide and a slipcover before analysis. You can use this section for many purposes. For instance, you might look for the hints of color or texture if the thin section has been stained. Or you might assess a rock’s thermal maturation with Thermal Alteration Index slides, or TAI slides; or using a Vitrinite Slide in reflected light. Polished thin section, if you have an opaque metallic mineral, you’ll most likely be using reflected-light microscopic analysis, for which you will need a thicker section polished on at least one side. Doubly-polished sections, or sections polished on both sides, typically provide a sharper image and greater detail than a section polished on one side alone. Sometimes, you might have doubly-polished sections prepared for fluid inclusion so that you can have them examined for microscopic bubbles inside the mineral grains as the mineral was initially forming.

|  |  |
| --- | --- |
| Source: http://www.nationalpetrographic.com/different-types-of-thin-sections  Figure 2.0 Thin Section | Source: https://viva.pressbooks.pub/petrology/chapter/thin-sections/  Figure 3.0 Standard Thin Section Components |

Petrographic Thin Section Image or Photograph - Petrographic thin section is prepared to study the optical properties of the minerals in the rock. Photograph of rock from thin section are called photomicrograph, prepared using photo microscopy using optical microscope. Basically, the process is done by connecting a digital camera to a laboratory microscope providing the capacity to take picture with high magnification.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Source: https://www.corelab.com/ps/thin-section-petrography  Figure 4.0 Petrographic Thin Section Images | | |

### Petrographic Thin Section Analysis

Micromorphology, or thin-section analysis, is the microscopic examination of the composition and structure of sediments. It was originally developed in soil science from the early 1960’ on the concepts of microstructures of fine-grained sediments, "Micromorphology is the branch of soil science that is concerned with the description, interpretation and, to an increasing extent, the measurement of components, features and fabrics in soils at a microscopic level" (Bullock et al, 1985). Analysis of thin section uses standard equipment like optical microscope and techniques to highlight textural and structural characteristics of sample rock. Analyses results on thin section helps in exploration and reservoir characterization in oil and gas industries.

Petrographic analysis provides a detailed description of the texture (grain size, sorting, and grain contacts), sedimentary structures (laminations, bioturbation), framework grain composition, authigenic minerals, and types and distribution of microporosity seen in a thin section. Detailed petrographic techniques can be used in porosity modeling programs and analysis with ultraviolet light can be useful in delineating features that are too small to be easily recognized with standard petrographic analysis. Petrographic analysis can be used to evaluate the pore system in a reservoir rock. The occurrence and distribution of pore types can be identified, and pore structure analysis can determine the ratio of primary intergranular pores to secondary leached pores and provide critical data for evaluation the efficiency of the pore network.

Classification of rocks as one of the fundamental tasks in a geological study requires a human expert to examine sampled thin section images under a microscope (S. Joseph, H. Ujir and I. Hipiny, 2017). In a study conducted by the same authors titled "*Unsupervised classification of Intrusive igneous rock thin section images using edge detection and colour analysis*", where they used edge detection and colour analysis (histogram) as part of their method in the rock classification that yielded successful results as high as 90% to 100%.

Petrographic microscopy is time-consuming and iterative (S. Joseph, H. Ujir and I. Hipiny, 2017). This inherent drawback is something that has been with petrographic thin section microscopy study despite of its long tradition with the applied and academic geosciences. A geoscientist can straightforwardly identify a rock’s constituent mineral phases, quantify fabric parameters and infer a rock’s genesis using thin sections and a petrographic microscope. Mineral and rock classification use point counting method in a petrographic thin section to find out the percentage of each mineral. Although point counting is very time consuming, it is of common use in several domains including geology, biology, medicine, and materials sciences, among others (S. Joseph, H. Ujir and I. Hipiny, 2017). Point counting in thin sections is normally conducted through mechanical or electromechanical devices attached to a microscope; such devices are very expensive, offer limited functionality, and are very time consuming (Martín L. Larrea, Silvia M. Castro & Ernesto A. Bjerg, 2014).

It requires a human expert with substantial knowledge and experience in combining multiple petrographic and microscopy classification criteria, e.g., texture, color, cleavage, twinning and tartan pattern, to perform the point counting. Due to this, we argue that an unsupervised classification of the rock thin sections to replace the human operator is therefore necessary (S. Joseph, H. Ujir and I. Hipiny, 2017).

Petrographic thin section analysis to classify rocks can be accomplished via the characterization of different minerals in rocks, which is performed by using methods such as listed below (C. Su, S. Xu, K. Zhu, X. Zhang, 2020).

* polarized light microscopy,
* X-Ray Diffraction (XRD),
* X-Ray Fluorescence (XRF),
* Atomic Absorption Spectroscopy (AAS),
* Electron Micro Probe Analyzer (EMPA),
* Scanning Electron Microscopy-Energy Dispersive X-ray spectroscopy (SEM-EDX),
* Transmission Electron Microscopy (TEM).

Among the mentioned methods used for petrographic thin section analysis, the most important and widely used methodology is the manual analysis conducted by geologists on the image of the petrographic thin section, which is obtained by using the polarizing microscope. Given the methods, combined with the expertise of an experienced geoscientist, manual analysis is exposed for potential errors and inconsistency on the analysis result[.](https://arxiv.org/pdf/2003.10437.pdf).) Since this process is done manually, it is more inclined towards qualitative analysis rather than quantitative, as individual observation may vary and this could possibly lead to misclassification by the observer (C. Su, S. Xu, K. Zhu, X. Zhang, 2020). Even experts, aided with microscope, have problems with automation of mineral classification process using a microscope (S. Joseph, H. Ujir and I. Hipiny, 2017).

### Machine Learning

Machine learning is the field of scientific study that concentrates on induction algorithms and on other algorithms that can be said to ``learn.'' (R. Kohavi and F. Provost,1998). In the field of machine learning, an induction algorithm represents an example of using mathematical principles for the development of sophisticated computing systems, induction algorithms are often thought of as a form of “decision support” (Justin Stoltzfus, 2018). Machine learning looks closely on computer algorithms that improve automatically through experience, (T. M. Mitchell, 1997) it is a branch of artificial intelligence (AI) that focuses on the study of computer algorithms that use data to improve a system through experience. Machine learning algorithms use historical data, also known as “training data”, to build an analytical model to make a prediction without being explicitly programmed. Generally, an algorithm is a finite mathematical step to solve a specific problem. In the world of computers, an algorithm is the set of instructions that defines not just what needs to be done but how to do it, combined with other algorithms produces a computer program.

### Learning Types

Machine learning approaches, also termed as “learning types” (Shagan Sha, 2020) have three primary broad categories, the Supervised, Unsupervised, and Reinforced Learning. These categories can be grouped based on the learning approach, type of input and output data and problem that they target to solve.

Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs. Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input.

Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.

Performing machine learning involves creating a model, which is trained on some training data and then can process additional data to make predictions. Various types of models have been used and researched for machine learning systems. Usually, machine learning models require a lot of data in order for them to perform well. Usually, when training a machine learning model, one needs to collect a large, representative sample of data from a training set. Data from the training set can be as varied as a corpus of text, a collection of images, and data collected from individual users of a service. Overfitting is something to watch out for when training a machine learning model. Trained models derived from biased data can result in skewed or undesired predictions. Algorithmic bias is a potential result from data not fully prepared for training.

### Machine Learning Algorithms for Image Classification

Image Classification is a fundamental task that attempts to comprehend an entire image as a whole. The goal is to classify the image by assigning it to a specific label. Typically, Image Classification refers to images in which only one object appears and is analyzed. In contrast, object detection involves both classification and localization tasks, and is used to analyze more realistic cases in which multiple objects may exist in an image (<https://paperswithcode.com/task/image-classification>).

This study compared five common machine learning algorithms for performing classification included Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Binary Decision Tree (BDT) and Discriminant Analysis (DA). AlexNet deep learning model was used to build these machine learning classifiers. The building classifiers were implemented and evaluated according to standard performance criteria of Accuracy (ACC), Precision (P), Sensitivity (S), Specificity (Spe) and Area Under the ROC Curve (AUC). The five methods were evaluated using 2608 histopathological images for head and neck cancer. The comparison was conducted using 2 times 10-fold cross validation. For each method, the pre-trained AlexNet network was used to extract features from the activation layer. The results illustrated that, there was no difference between the results of SVM and KNN. Both have the same and the higher accuracy than others were 99.98 %, whereas 99.81%, 97.32% and 93.68% for DA, BDT and NB, respectively. The present study shows that the SVM, KNN and DA are the best methods for classifying our dataset images (Maisun Al Zorgani and Hassan Ugail, 2018). <https://easychair.org/publications/preprint/bRtJ#:~:text=Both%20have%20the%20same%20and,for%20classifying%20our%20dataset%20images>.

Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

In a study conducted by R. Pires De Lima (2019), titled “PETROGRAPHIC ANALYSIS WITH DEEP CONVOLUTIONAL NEURAL NETWORKS”, stated, “machine learning techniques have been successfully applied, with considerable success, in the geosciences for almost two decades. A convolutional neural network (CNN) is a deep learning (DL) method that has been widely and successfully applied to computer vision tasks including object localization, detection, and image classification. DL for supervised learning tasks is a method that uses the raw data to determine the classification features, in contrast to other machine learning (ML) techniques that require pre-selection of the input features (or attributes). In the geosciences, we hypothesize that ***deep learning will facilitate the analysis of uninterpreted images*** that have been neglected ***due*** to a ***limited number of experts***, such as fossil images, slabbed cores, or petrographic thin sections. We use transfer learning, which employs previously trained models to shorten the development time for subsequent models, to address a suite of geologic interpretation tasks that may benefit from ML. Using two different base models, MobileNet V2 and Inception V3, we illustrate the successful classification of microfossils, core images, petrographic photomicrographs, and rock and mineral hand sample images. ML does not replace the expert geoscientist. The expert defines the labels (interpretations) needed to train the algorithm and also monitors the results to address 4 incorrect or ambiguous classifications. ML techniques provide a means to apply the expertise of skilled geoscientists to much larger volumes of data.

Machine Learning in Petrographic Thin Section Analysis - Combining image processing and statistical learning, it has been proven that machine learning can perform automatic quantitative and qualitative analysis of thin sections. Machine learning was incorporated in the methodology of lithology identification, classification and quantification of thin sections samples with fast-accurate model training and generation and have a good classification result.

### Database

A database is an organized collection of data, generally stored and accessed electronically from a computer system. Where databases are more complex, they are often developed using formal design and modeling techniques. The database management system (DBMS) is the software that interacts with end users, applications, and the database itself to capture and analyze the data.

Database design is the organization of data according to a database model. The designer determines what data must be stored and how the data elements interrelate. With this information, they can begin to fit the data to the database model. Database management system manages the data accordingly. Database design involves classifying data and identifying interrelationships. This theoretical representation of the data is called an ontology. The ontology is the theory behind the database's design.

Database for Machine Learning - Recommended database for machine learning are available that can support or allow to build machine learning models right where the data resides. These databases are also ready to support programming language for machine learning. MLDB is an open-source database designed for machine learning. You can install it wherever you want and send it commands over a RESTful API to store data, explore it using SQL, then train machine learning models and expose them as APIs.

### Programming

Computer programming is the process of designing and building an executable computer program to accomplish a specific computing result or to perform a specific task. Programming involves tasks such as: analysis, generating an algorithm, profiling algorithms' accuracy and resource consumption, and the implementation of algorithms in a chosen programming language (Commonly referred to as coding).

Choosing Programming Language for a Project - A computer programmer must know the foundation of computer programming and should start to focus a at least one specific programming language. There are a variety of ways to learn computer coding.

Programming Language for Machine Learning - The complexity of Machine Learning (ML) models and the frame-works people are using to build them has exploded along with ML itself. State-of-the-art models are increasingly programs, with support for programming constructs like loops and recursion, and this brings out many interesting issues in the tools we use to create them — that is, programming languages. Among the selection on different available programming languages, Python ranks as the most preferred tool but when building a complete system, it will require that you have knowledge on other programming language for front, middle, and backend part of a complete system.

### Networking

A computer network is a group of computers that use a set of common. communication protocols over digital interconnections for the purpose of sharing resources located on or provided by the network nodes. The interconnections between nodes are formed from a broad spectrum of telecommunication network technologies, based on physically wired, optical, and wireless radio-frequency methods that may be arranged in a variety of network topologies.

The nodes of a computer network may include personal computers, servers, networking hardware, or other specialized or general-purpose hosts. They are identified by hostnames and network addresses. Hostnames serve as memorable labels for the nodes, rarely changed after initial assignment. Network addresses serve for locating and identifying the nodes by communication protocols such as the Internet Protocol.

Purpose or Use - A computer network extends interpersonal communications by electronic means with various technologies, such as email, instant messaging, online chat, voice and video telephone calls, and video conferencing. A network allows sharing of network and computing resources. Users may access and use resources provided by devices on the network, such as printing a document on a shared network printer or use of a shared storage device. A network allows sharing of files, data, and other types of information giving authorized users the ability to access information stored on other computers on the network. Distributed computing uses computing resources across a network to accomplish tasks.

### Systems Analysis and Design

Systems Analysis and Design (SAD) is a broad term for describing methodologies for developing high quality Information System which combines Information Technology, people and Data to support business requirement. The SAD technique is not only limited to IT systems and can be used to create just about anything, from a family house to the international space station. But there is no silver bullet in simplifying the development of computer systems. This principle is still true today. In other words, there is no single, simple technique that developers can use to ensure successful Information Technology (IT) projects. However, there are development methodologies that can be followed which will greatly assist an IT professional in developing and enhancing systems. A methodology is essentially a procedure to get something done. A development methodology can be thought of as a roadmap. While a roadmap for a traveler will provide the details from driving from point A to point B, a development methodology will provide the IT professional with guidelines for taking a system from conception through implementation and beyond. (Ramakrishnan S, 2012) <https://www.longdom.org/open-access/system-analysis-and-design-2165-7866.S8-e001.pdf>

System Development Life Cycle - The systems development life cycle (SDLC) is the process of determining how an information system (IS) can support business needs, designing the system, building it, and delivering it to users. the fundamental four-phase model (planning, analysis, design, and implementation) common to all information systems development projects. It describes the evolution of system development methodologies and discusses the roles and skills required of a systems analyst (Dennis, Wixom, Roth, 2012, 2019).

The systems development life cycle (SDLC) is a conceptual model used in project management that describes the stages involved in an information system development project, from an initial feasibility study through maintenance of the completed application. SDLC can apply to technical and non-technical systems. In most use cases, a system is an IT technology such as hardware and software. Project and program managers typically take part in SDLC, along with system and software engineers, development teams and end-users.

Every hardware or software system will go through a development process which can be thought as an iterative process with multiple steps. SDLC is used to give a rigid structure and framework to define the phases and steps involved in the development of a system.

SDLC is also an abbreviation for Synchronous Data Link Control and software development life cycle. Software development life cycle is a very similar process to systems development life cycle, but it focuses exclusively on the development life cycle of software.

### SDLC models

Various SDLC methodologies have been developed to guide the processes involved, including the original SDLC method, the Waterfall model. Other SDLC models include rapid application development (RAD), joint application development (JAD), the fountain model, the spiral model, build and fix, and synchronize-and-stabilize. Another common model today is called Agile software development.

Frequently, several models are combined into a hybrid methodology. Many of these models are shared with the development of software, such as waterfall or agile. Numerous model frameworks can be adapted to fit into the development of software.

(Dennis, Wixom, Roth, 2012, 2019, Systems Analys & Design, 5th Edition)

In SDLC, documentation is crucial, regardless of the type of model chosen for any application and is usually done in parallel with the development process. Some methods work better for specific kinds of projects, but in the final analysis, the most crucial factor for the success of a project may be how closely the particular plan was followed.

### Steps in SDLC

SDLC can be made up of multiple steps. There is no concrete set number of steps involved. Around seven or eight steps appear commonly; however, there can be anywhere from five upwards to 12. Typically, the more steps defined in an SDLC model, the more granular the stages are.

In general, an SDLC methodology follows these following steps:

1. Analysis: The existing system is evaluated. Deficiencies are identified. This can be done by interviewing users of the system and consulting with support personnel.
2. Plan and requirements: The new system requirements are defined. In particular, the deficiencies in the existing system must be addressed with specific proposals for improvement. Other factors defined include needed features, functions and capabilities.
3. Design: The proposed system is designed. Plans are laid out concerning the physical construction, hardware, operating systems, programming, communications and security issues.
4. Development: The new system is developed. The new components and programs must be obtained and installed. Users of the system must be trained in its use.
5. Testing: All aspects of performance must be tested. If necessary, adjustments must be made at this stage. Tests performed by quality assurance (QA) teams may include systems integration and system testing.
6. Deployment: The system is incorporated in a production environment. This can be done in various ways. The new system can be phased in, according to application or location, and the old system gradually replaced. In some cases, it may be more cost-effective to shut down the old system and implement the new system all at once.
7. Upkeep and maintenance: This step involve changing and updating the system once it is in place. Hardware or software may need to be upgraded, replaced or changed in some way to better fit the needs of the end-users continuously. Users of the system should be kept up to date concerning the latest modifications and procedures.

Other steps which may appear include project initiation, functional specifications, detailed specifications, evaluation, end-of-life and other steps that can be created by splitting previous steps apart further.

1. Advantages and disadvantages of SDLC
2. Benefits of abiding by a clearly defined SDLC model include:
3. Having a clear view of an entire project, workers involved, estimated costs and timelines.
4. Gives project managers a projected base cost of the project.
5. Goals and standards are clearly defined.
6. Developers can move back a step if something does not go as expected.
7. Disadvantages, however, can include:
8. Due to assumptions made at the beginning of a project, if an unexpected circumstance complicates the development of a system, then it may stockpile into more complications down the road. As an example, if newly installed hardware does not work correctly, then it may increase the time a system is in development, increasing the cost.
9. Some methods are not flexible.
10. It can be complicated to estimate the overall cost at the beginning of a project.
11. Testing at the end of development may slow down some development teams.

The Systems Analyst – The book “Systems Analysis and Design, 5th Edition ()” pointed out that the key person in the SDLC is the systems analyst. System analyst roles and responsibilities are to analyzes the business situation, identifies opportunities for improvements, and designs a system to implement the improvements. Many systems analysts view their profession as one of the most interesting, exciting, and challenging jobs around. As a systems analyst, you will work as a team with a variety of people, including business and technical experts. You will feel the satisfaction of seeing systems that you designed and developed make a significant positive business impact, while knowing that your unique skills helped make that happen. It is important to remember that the primary objective of the systems analyst is not to create a wonderful system. The primary goal is to create value for the organization, which for most companies means increasing profits. (Government agencies and not-for-profit organizations measure value differently.) Many failed systems were abandoned because the analysts tried to build a wonderful system without clearly understanding how the system would support the organization’s goals, improve business processes, and integrate with other information systems to provide value. An investment in an information system is like any other investment, such as a new machine tool. The goal is not to acquire the tool, because the tool is simply a means to an end; the goal is to enable the organization to perform work better so that it can earn greater profits or serve its constituents more effectively (Dennis, Wixom, Roth, 2012, 2019).

### Software Engineering

Software engineering is a branch of computer science which includes the development and building of computer systems software and applications software. Computer systems software is composed of programs that include computing utilities and operations systems. Applications software consists of user-focused programs that include web browsers, database programs, etc.

Why Software Engineering - Software Engineering, according to IEEE, is applying the principles of engineering to software development. In other words, it applies well-defined principles and methods of engineering to develop a high-quality software. On the other hand, a Software engineer is a person whose main responsibility is applying Software engineering principles to analyze, design, develop and maintain the software (Software life cycle).

Some people get confused between software engineering and computer science. One thing should be clear; they both require programming skills. Computer science focuses on “computing”, such as algorithms, programming languages, artificial intelligence and hardware design. While Software engineering focuses on all aspects of building high quality software.

Every industry, company or organization today depends on software systems, which makes the software engineering profession in demand everywhere. Software based companies, which has the most influence on the world, tend to follow software engineering principles especially since current software systems have become much more complex.

In Software development work environment, software engineers could work on different fields including requirements analysis, design, quality assurance, and programming. Software engineers are familiar with most aspects related to the software development lifecycle.

### Evaluation System

The evaluation system includes the following factors and indicators: Functionality, Content, Reliability, Availability, Maintainability, and Saleability. For Functionality, it refers to set of attributes that bear on ease of operation, provision for comfort and convenience and user-friendliness. For Content, it refers to set of attributes that bear on accuracy, updatedness and presentation of content. For Reliability, it refers to set of attributes that bear on conformance to desired results, absence of failure and accuracy in performance. For Availability, the project was evaluated according to performance to specifications, provision for security requirements and completeness of the project. For Maintainability, the project was evaluated according to ease of maintenance, provision for diagnostic tools and procedures, provision for enhancements and modifications. For Saleability, it refers to set of attributes that bear on presence of market demand, competitiveness of price and attractiveness of the design.

Software ISO Quality Standard - ISO 9126 refers to an international standard for the evaluation of software quality. The fundamental objective of this standard is to address some of the well-known human biases that can adversely affect the delivery of a software development project. These biases include changing priorities after the start of a project or not having any clear definitions of "success." The ISO tries to develop a common understanding of the project's objectives and goals.

The quality model presented the software quality standards as follows; a) Functionality, refers to set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied needs, b) Reliability, refers to set of attributes that bear on the capability of software to maintain its level of performance under stated conditions for a stated period of time, c) Usability, refers to a set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users, d) Efficiency, refers to set of attributes that bear on the relationship between the level of performance of the software and the number of resources used, under stated conditions, e) Maintainability, refers to set of attributes that bear on the effort needed to make specified modifications, and f) Portability, refers to set of attributes that bear on the ability of software to be transferred from one environment to another.

### Software Requirements

Operating System is the most important software that runs on a computer. It manages the computer's memory and processes, as well as all of its software and hardware. It also allows you to communicate with the computer without knowing how to speak the computer's language. Without an operating system, a computer is useless.

Operating systems usually come pre-loaded on any computer you buy. Most people use the operating system that comes with their computer, but it's possible to upgrade or even change operating systems. The three most common operating systems for personal computers are Microsoft Windows, macOS, and Linux.

Common desktop operating systems include Windows, OS X, and Linux. While each OS is different, most provide a graphical user interface, or GUI, that includes a desktop and the ability to manage files and folders. They also allow you to install and run programs written for the operating system. Windows and Linux can be installed on standard PC hardware, while OS X is designed to run on Apple systems. Therefore, the hardware you choose affects what operating system(s) you can run.

Modern operating systems use a graphical user interface, or GUI (pronounced gooey). A GUI lets you use your mouse to click icons, buttons, and menus, and everything is clearly displayed on the screen using a combination of graphics and text.

Microsoft created the Windows operating system in the mid-1980s. There have been many different versions of Windows, but the most recent ones are Windows 10 (released in 2015), Windows 8 (2012), Windows 7 (2009), and Windows Vista (2007). Windows comes pre-loaded on most new PCs, which helps to make it the most popular operating system in the world.

Linux (pronounced LINN-ux) is a family of open-source operating systems, which means they can be modified and distributed by anyone around the world. This is different from proprietary software like Windows, which can only be modified by the company that owns it. The advantages of Linux are that it is free, and there are many different distributions—or versions—you can choose from. According to StatCounter Global Stats, Linux users account for less than 2% of global operating systems. However, most servers run Linux because it's relatively easy to customize.

When software developers create applications, they must write and compile them for a specific operating system. This is because each OS communicates with the hardware differently and has a specific application program interface, or API, that the programmer must use. While many popular programs are cross platform, meaning they have been developed to run on different Operating System, some are only available for a single operating system. Therefore, when choosing a computer, make sure the operating system supports the programs you want to run (Christensson, P. 2016, July 23 Operating System Definition. Retrieved 2021, May 26, from <https://techterms.com>).

### SQLite

SQLite is a C-language library that implements a small, [fast](https://www.sqlite.org/fasterthanfs.html), [self-contained](https://www.sqlite.org/selfcontained.html), [high-reliability](https://www.sqlite.org/hirely.html), [full-featured](https://www.sqlite.org/fullsql.html), SQL database engine. SQLite is the [most used](https://www.sqlite.org/mostdeployed.html) database engine in the world. SQLite is built into all mobile phones and most computers and comes bundled inside countless other applications that people use every day. [More Information...](https://www.sqlite.org/about.html)

The SQLite [file format](https://www.sqlite.org/fileformat2.html) is stable, cross-platform, and backwards compatible and the developers pledge to keep it that way [through the year 2050](https://www.sqlite.org/lts.html). SQLite database files are commonly used as containers to transfer rich content between systems and as a long-term archival format for data. There are over 1 trillion (1e12) SQLite databases in active use. ( <https://www.sqlite.org/index.html> )

### Python

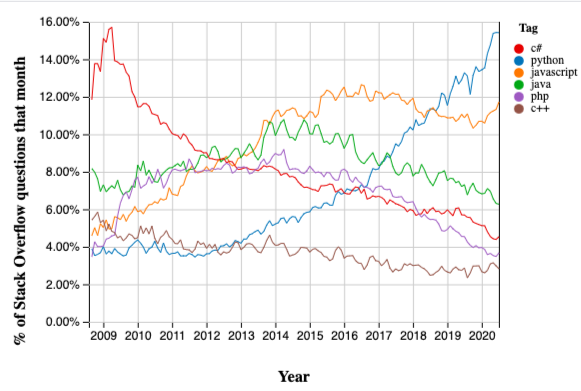
Python is a high-level programming language designed to be easy to read and simple to implement. It is open source, which means it is free to use, even for commercial applications. Python can run on Mac, Windows, and Unix systems and has also been ported to Java and .NET virtual machines.

Python is considered a scripting language, like Ruby or Perl and is often used for creating Web applications and dynamic Web content. It is also supported by a number of 2D and 3D imaging programs, enabling users to create custom plug-ins and extensions with Python. Examples of applications that support a Python API include GIMP, Inkscape, Blender, and Autodesk Maya. Scripts written in Python (.PY files) can be parsed and run immediately. They can also be saved as a compiled program (.PYC files), which are often used as programming modules that can be referenced by other Python programs (<https://techterms.com/definition/python>).

Python continues to be the most preferred language for scientific computing, data science, and machine learning, boosting both performance and productivity by enabling the use of low-level libraries and clean high-level APIs (S. Raschka, J. Patterson, and C. Nolet, 2020).

The increasing adoption of machine learning worldwide is a major factor contributing to its growing popularity. There are 69% of machine learning engineers and Python has become the favorite choice for data analytics, machine learning, and AI – all thanks to its vast library ecosystem that let’s machine learning practitioners' access, handle, transform, and process data with ease. Python wins the heart of machine learning engineers for its platform independence, less complexity, and better readability.

With over 8.2 million developers across the world using Python for coding, Python ranks first in the latest annual ranking of popular programming languages by IEEE Spectrum with a score of 100. Stack overflow programming language trends clearly show that it’s the only language on rising for the last five years.



Source: Stackoverflow

Figure 5.0 Rank of Popular Programming Language

Jupyter Notebook

Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses includes data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

JupyterLab, Jupyter’s Next-Generation Notebook Interface - JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning. JupyterLab is extensible and modular: write plugins that add new components and integrate with existing ones.

Internet Browser -

A web browser, or simply "browser," is an application used to access and view websites. Common web browsers include Microsoft Internet Explorer, Google Chrome, Mozilla Firefox, and Apple Safari.

The primary function of a web browser is to render HTML, the code used to design or "mark up" webpages. Each time a browser loads a web page, it processes the HTML, which may include text, links, and references to images and other items, such as cascading style sheets and JavaScript functions. The browser processes these items, then renders them in the browser window.

Early web browsers, such as Mosaic and Netscape Navigator, were simple applications that rendered HTML, processed form input, and supported bookmarks. As websites have evolved, so have web browser requirements. Today's browsers are far more advanced, supporting multiple types of HTML (such as XHTML and HTML 5), dynamic JavaScript, and encryption used by secure websites.

The capabilities of modern web browsers allow web developers to create highly interactive websites. For example, Ajax enables a browser to dynamically update information on a webpage without the need to reload the page. Advances in CSS allow browsers to display a responsive website layout and a wide array of visual effects. Cookies allow browsers to remember your settings for specific websites.

While web browser technology has come a long way since Netscape, browser compatibility issues remain a problem. Since browsers use different rendering engines, websites may not appear the same across multiple browsers. In some cases, a website may work fine in one browser, but not function properly in another. Therefore, it is smart to install multiple browsers on your computer so you can use an alternate browser if necessary (Christensson, P. (2014, February 28). Web Browser Definition. Retrieved 2021, May 27, from [https://techterms.com](https://techterms.com/)).

Django Framework

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source.

Django Framework is a Python-based free and open-source web framework that follows the model–template–views (MTV) architectural pattern. It is maintained by the Django Software Foundation (DSF), an American independent organization established as a non-profit.

Django's primary goal is to ease the creation of complex, database-driven websites. The framework emphasizes reusability and "pluggability" of components, less code, low coupling, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings, files, and data models. Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models.

Visual Studio Code - Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages (such as C++, C#, Java, Python, PHP, Go) and runtimes (such as .NET and Unity).

Visual Studio Code is a freeware source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. Users can change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.

Microsoft has released most of Visual Studio Code's source code on the Microsoft/vscode (Code – OSS) repository of GitHub, under the permissive MIT License, while the releases by Microsoft are proprietary freeware.

In the Stack Overflow 2019 Developer Survey, Visual Studio Code was ranked the most popular developer environment tool, with 50.7% of 87,317 respondents reporting that they use it. Visual Studio Code was first announced on April 29, 2015, by Microsoft at the 2015 Build conference. A Preview build was released shortly thereafter.

On November 18, 2015, Visual Studio Code was released under the MIT License, having its source code available on GitHub. Extension support was also announced. On April 14, 2016, Visual Studio Code graduated from the public preview stage and was released to the Web.

### Hardware Requirements

Technically, a computer is a programmable machine. This means it can execute a programmed list of instructions and respond to new instructions that it is given. Today, however, the term is most often used to refer to the desktop and laptop computers that most people use. When referring to a desktop model, the term "computer" technically only refers to the computer itself -- not the monitor, keyboard, and mouse. Still, it is acceptable to refer to everything together as the computer.

Some of the major parts of a personal computer (or PC) include the motherboard, CPU, memory (or RAM), hard drive, and video card. While personal computers are by far the most common type of computers today, there are several other types of computers. For example, a "minicomputer" is a powerful computer that can support many users at once. A "mainframe" is a large, high-powered computer that can perform billions of calculations from multiple sources at one time. Finally, a "supercomputer" is a machine that can process billions of instructions a second and is used to calculate extremely complex calculations (Christensson, P. (2006). Computer Definition. Retrieved 2021, May 27, from <https://techterms.com>).

Processor - A central processing unit (CPU), also called a central processor, main processor or just processor, is the electronic circuitry that executes instructions comprising a computer program. The CPU performs basic arithmetic, logic, controlling, and input/output (I/O) operations specified by the instructions in the program.

Memory (RAM) - Random-access memory is a form of computer memory that can be read and changed in any order, typically used to store working data and machine code. A random-access memory device allows data items to be read or written in almost the same amount of time irrespective of the physical location of data inside the memory.

Storage - A computer storage device is any type of hardware that stores data. The most common type of storage device, which nearly all computers have, is a hard drive. The computer's primary hard drive stores the operating system, applications, and files and folders for users of the computer.

While the hard drive is the most ubiquitous of all storage devices, several other types are common as well. Flash memory devices, such as USB keychain drives and iPod nanos are popular ways to store data in a small, mobile format. Other types of flash memory, such as compact flash and SD cards are popular ways to store images taken by digital cameras.

External hard drives that connect via Firewire and USB are also common. These types of drives are often used for backing up internal hard drives, storing video or photo libraries, or for simply adding extra storage. Finally, tape drives, which use reels of tape to store data, are another type of storage device and are typically used for backing up data (Christensson, P. (2006, November 16). Storage Device Definition. Retrieved 2021, May 27, from <https://techterms.com>).

Most of the processing done on a computer is done via the computer's central processing unit, or CPU. So, in order to give the CPU a break and help it run more efficiently, a video card can be used to process the graphics portion of the processing load. Because most of today's programs are graphically oriented, the video card can help almost any program run more efficiently. However, the difference in performance is especially noticeable in image editing applications and 3D games.

Video cards, also called graphics accelerators, can speed up both 2D and 3D graphics rendering. Programs such as photo editors and Web browsers may benefit from 2D acceleration, while CAD design programs and video games will most likely benefit from the card's 3D acceleration. Some programs rely so heavily on the video card, that they will not run if a supported video card is not installed.

Most video cards support the OpenGL and DirectX libraries. These libraries include commands for manipulating graphics that programmers can include in their code. Some of these commands may include moving or rotating an object, morphing polygons, or casting light and creating shadows. By using standard OpenGL or DirectX functions, it makes it easier for developers to create graphically oriented programs. Of course, it also makes it necessary for the computer to include a supported video card in order for the program to run.

Video cards are typically installed in either the [PCI](https://techterms.com/definition/pci) or [AGP](https://techterms.com/definition/agp) slots in the back of a computer. Most computers come with a video card installed in one of these slots, which means it can be upgraded at a later time (Christensson, P. (2006). Video Card Definition. Retrieved 2021, May 27, from <https://techterms.com>).

A printer is an [output device](https://techterms.com/definition/outputdevice) that prints paper documents. This includes text documents, images, or a combination of both. The two most common types of printers are [inkjet](https://techterms.com/definition/inkjet) and [laser printers](https://techterms.com/definition/laserprinter). Inkjet printers are commonly used by consumers, while laser printers are a typical choice for businesses. Dot matrix printers, which have become increasingly rare, are still used for basic text printing.

The printed output produced by a printer is often called a hard copy, which is the physical version of an electronic document. While some printers can only print black and white hard copies, most printers today can produce color prints. In fact, many home printers can now produce high-quality photo prints that rival professionally developed photos. This is because modern printers have a high DPI (dots per inch) setting, which allows documents to printed with a very fine resolution.

In order to print a document, the electronic data must be sent from the computer to the printer. Many software programs, such as word processors and image editing programs, include a "Print..." option in the File menu. When you select "Print," you will typically be presented with a Print dialog box. This box allows you to select the print output settings before sending the document to the printer. After choosing the appropriate settings, you can hit the Print button, which will send the document to the printer.

Of course, for the document to print, the printer must be turned on and connected to the computer. Most modern printers are connected using a standard USB cable. However, some printers can be wirelessly connected to one or more computers over a Wi-Fi network. You can also use more than one printer on a single computer, as long as the correct drivers are installed (Christensson, P. (2009, January 12). Printer Definition. Retrieved 2021, May 27, from <https://techterms.com>).

## Conceptual Model of the Study

Figure 6.0 shows the conceptual model of the study. The abovementioned concept, theories, and insight of related literature were used to create the model.

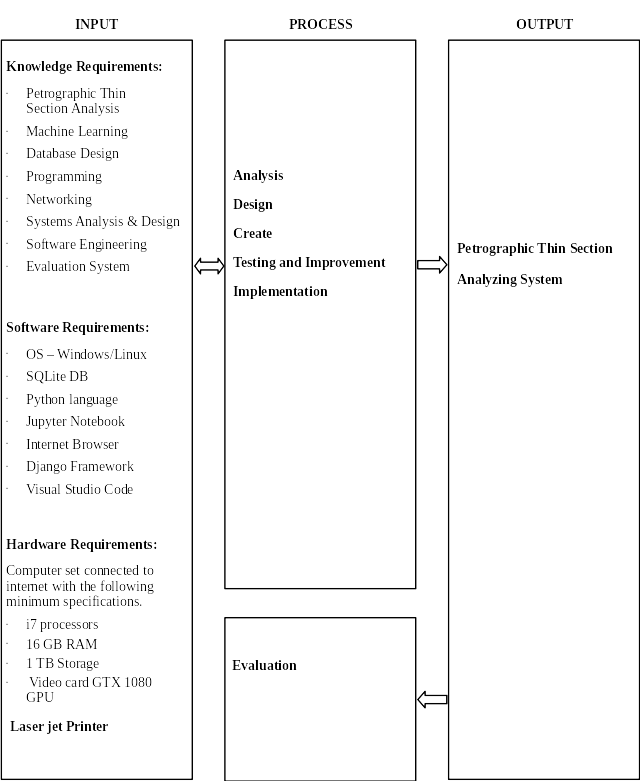


Figure 6.0 Conceptual Model of the Study

The studies follow the input, process, output (IPO) flow and ultimately the evaluation of the system.

The three inputs for the system are knowledge requirements, software requirements, and hardware requirements. To develop the system, the developer must have the knowledge on Petrographic thin section, machine learning, computer programming mostly Python during the building of analytical model, Database Design, Computer Networking, Systems Analysis and Design, Software Engineering, and Evaluation System. The developer must have at least the following computer software; MS Windows or Linux Operating System, SQLlite database during the development phase, Jupyter Notebook, and Internet browser like IE, Edge or even Firefox latest version to be able to progress in making the system.

The process block is composed of system design, system development, system testing and improvement. The system design includes the context diagram, architectural diagram, and database design. The system development comprises the schedule of activities, system flow diagram, transaction flow diagram, reports design and specification, and program coding. Finally, the system testing, and improvement includes the unit testing, integration testing, and user’s acceptance testing.

This study aimed to utilize and maximize the use of organization IT infrastructure, network and computing devices/services. All software requirements are available and in working condition. The computers are all connected to the organization’s local area network or “network ready”.

## Operational Definition of Terms

To facilitate understanding of this project study, the following terms are operationally defined.

**Thin Section** – also called petrographic thin section, is a laboratory preparation of a mechanically ground optically flat fragment or slice of rocks to a thickness of 0.03 millimeters (30 µm) placed between glass plate.

**Petrographic Thin Section** - is a laboratory preparation of a mechanically ground optically flat fragment or slice of rocks to a thickness of 0.03 millimeters (30 µm) .

**Petrographic microscope** – a laboratory instrument used in petrology and optical mineralogy to identify rocks and minerals in thin sections.

**Petrographer** – A person who studies petrography.

**Petrography** – is a subdiscipline of Petrology that focus on the analysis of rocks in thin section by means of a petrographic microscope.

**Petrology** - as a scientific study of rocks that deals with their composition, texture, and structure; their occurrence and distribution; and their origin in relation to physicochemical conditions and geologic processes.

**Photomicrograph** – is a photograph of rock from thin section.

**Micromorphology -** is the microscopic examination of the composition and structure of sediments.

**Machine Learning** - a branch of artificial intelligence that focuses on the study of computer algorithms that use data to improve a system through experience.

**Data Set** – is the complete data, composed of training, validation and test data, that is used as an input in machine learning algorithms that produces a model as an output.

**Machine Learning Model** – also called analytical model in this study.

**Deep Learning** - is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

**Convolutional Neural Network** – a method of deep learning that is used in computer vision tasks including object localization, detection, and image classification.

# **Chapter 3**

# **Methodology**

This chapter presents the project design, the project development, and the operation, testing and evaluation procedure.

## Project Design

The project is a development of a system that will analyze petrographic thin section images using machine learning methodologies and techniques for use in an organization in the oil and gas industry.

Figure 7.0 show the basic workflow for the petrographic thin section analysis. The process starts by collecting rock samples. The rock samples are taken to a laboratory to take or cut a piece of thin slice which will undergo a process to create the petrographic thin section. Using a polarizing microscope with digital camera attached to it, and connected to a computer, digital images of petrographic thin section is produced. With the thin section and its digital images, geoscientists can now perform analysis and based on their skills and knowledge, a report is generated or written about the analysis.

Still in figure 7.0, a workflow is added, represented in green color broken line and rectangular objects, shows the workflow portion where and how a system can be developed and incorporated in the process to automate the analysis of petrographic thin section images using machine learning techniques.

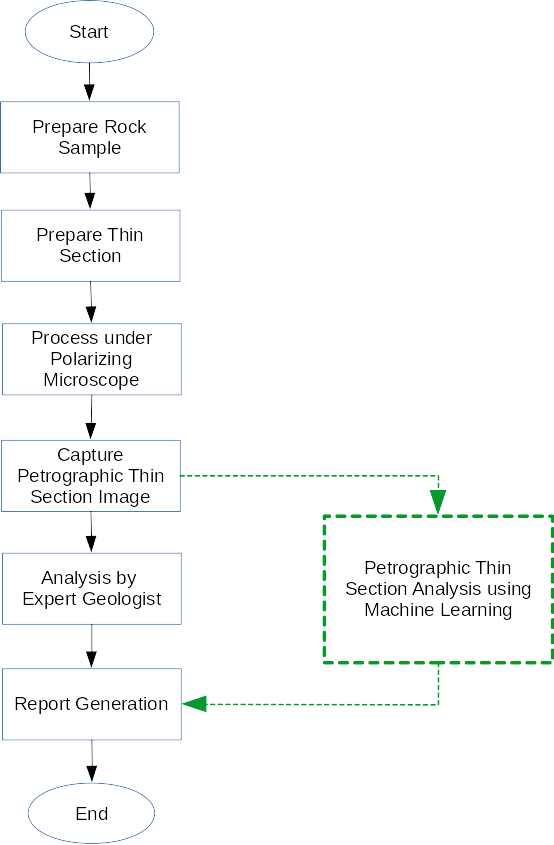


Figure 7.0 - Petrographic Thin Section Analysis Workflow

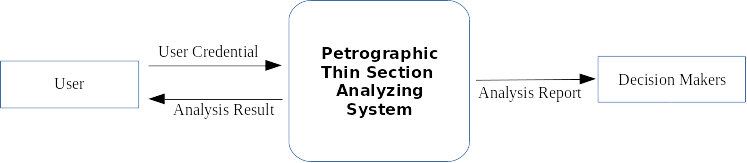


Figure 8.0 - Context Diagram of Petrographic Thin Section Analyzing System

Figure 8.0 shows the context diagram of the Petrographic Thin Section Analyzing System. The system at its at-a-glance-view shows single high-level process interacting with the external entities. The Petrographic Thin Section Analyzing System mainly serves the petrographers as the primary user who performs the thin section analysis. The Decision Maker here are the divisions heads, team leads and other positioned geoscientist in the same division who can view or print analysis report.

Figure 9.0 shows the decomposed context diagram into multiple bubbles/processes. This level shows 3 sub processes, the acquire image process, analyze image process and generate report process. Also, two data file sources were shown, the image folder and the analytical model file. Image folder contains thin section images that will be used as an input data, which will be accessed by the acquire images module whose output will be the input for the analyze image process. The analytical model file will be accessed by the analyze image module for the required analysis of the input image. Data output from the analyze image process will be fed to the generate report process whose output will be the result receive by external entity, the organization division’s head or team lead.

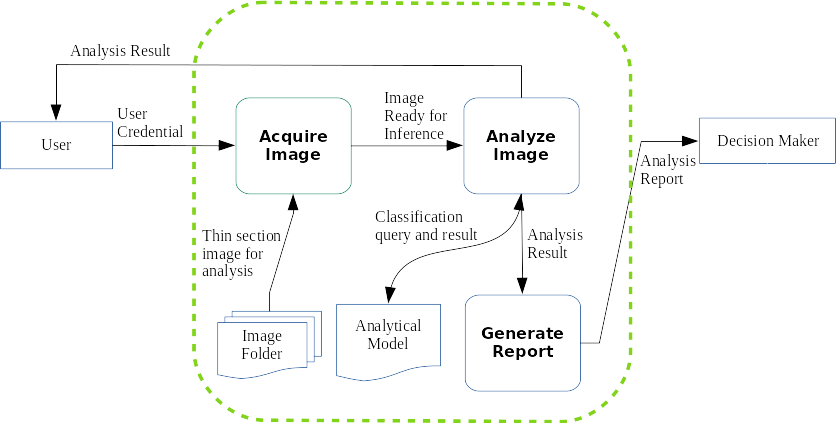


Figure 9.0 - Petrographic Thin Section Analyzing System with its Basic Subprocesses

Figure 10.0 shows deeper processes from the previous diagram (Figure 8.0) of the Petrographic Thin Section Analyzing System. Starting from the Petrographer or non-expert users with valid user credential accesses the system will use the get image process. Get image process will access the image folder to get the image to be analyzed. Image data will be passed to the view image process to be viewed and then send data to preprocess image process. Thin section image will be prepared and then send to the analyze image process. Analyze image process will use the updated analytical data to analyze the thin section image. After the analysis, resulting data will go to two processes, the generate report process and the TBD process. The generate report process will use the data to produce an analysis report to be used by the organization division head or team lead, the other external entity of the system. The TBD process will handle the output data from the analyze image process and save it to PTS database.

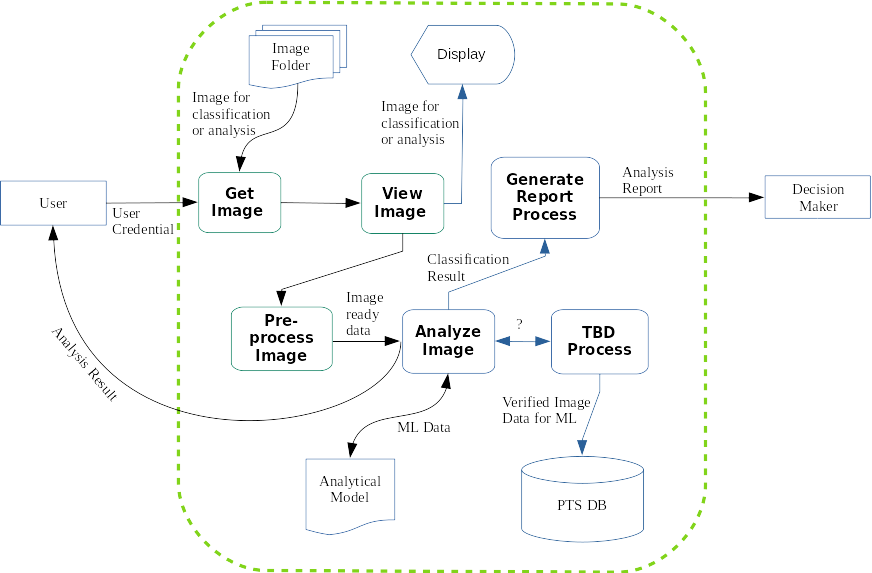


Figure 10.0 Petrographic Thin Section Analyzing System

The main component of a machine learning analytical model is data. Figure 11.0 shows the high-level representation of how to build and evaluate a machine learning analytical model to be used in making classification in the case of this project. The quality, quantity and how the data are prepared, called “pre-processing” in machine learning, is the most important part of model preparation.

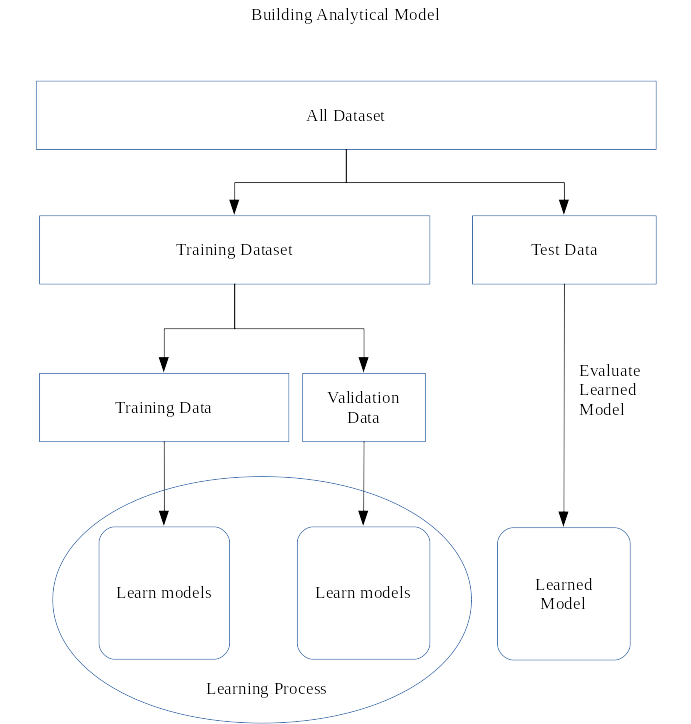


Figure 11.0 - Flow Chart of Building Machine Learning Analytical Model

## Project Development

Figure 12.0 show the top-level view of the phases for the development of the Petrographic Thin Section Analyzing System project.

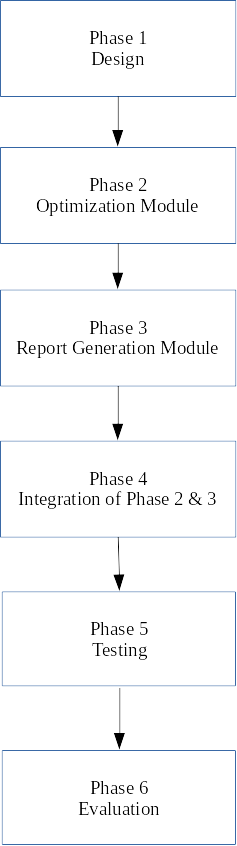


Figure 12.0 - Project Development Phases

### Phase 1

The design of the system will refer to the main and specific objectives that were formulated thru the identification of the problems that were discussed in the introduction of the study in Chapter 1. The development of the system assumes that the training data, hardware and software requirements in the conceptual framework are all at hand.

### Phase 2

This phase focuses on the development of an analytical model that will be used for the prediction. Development of the analytical model will use labeled petrographic thin section digital images. Machine learning methods and techniques will be used to build the analytical model. Jupyter Notebook will be used to import necessary python modules that will be used to process the data to build and validate model. Confusion matrix, a machine learning model performance measuring tool will be used to check the accuracy of the analytical model.

### Phase 3

Machine learning analytical model created in the previous phase will be tested using test data. This part of the development will check the readiness of the analytical model by introducing new data, a labeled petrographic data that it hasn't see or encountered yet. This phase will tell the readiness of the model to be deployed in the system.

### Phase 4

This phase will focus on the development of an application interface. The GUI for the application will run on the web browser. A python framework will be used to create the interface where the admin and user can interact with the system to perform their task. The model that was built and tested in the previous phase will now be used or deployed in this web browser-based application.

### Phase 5

The integration of phase 2 and 3 in the phase 4 of the development will now be tested as a system. An admin creates and grant privilege to user to use the system. User will upload and submit petrographic thin section images to be analyzed to test the application based on the design functionality in the project objective.

### Phase 6

The developed system will be evaluated using the ISO 25010 Evaluation of Software/System criteria.

## Operation and Testing Procedures

### Operation Procedure

This section will present the procedures on how to operate the system.

* 1. Users access the system with valid credentials.
  2. Upload-Image button is clicked to browse for the image for analysis.
  3. Image for analysis is shown in the computer screen for visual verification.
  4. User click Submit button to proceed for image analysis.
  5. Analysis result is displayed on the screen.
  6. Users click on the generate report button to generate report on analysis result.

### Testing Procedure

This section will present the procedures on how to test the system.

1. System with optimized processing time.

**Table 1.** *Test Procedure for Optimizing Processing Time.*

|  |  |
| --- | --- |
| Procedure | Expected Outcome |
| 1. Selecting image for analysis. | Able to browse and select image the image for analysis. |
| 1. Uploading Image for analysis. | Image can be viewed properly in the computer monitor. |
| 1. Image analysis | Display image along with the predicted label. |
| 1. Generate report. | Generate a report in a user preferred format and with required data. |

Table 1 shows the test procedure for optimizing processing time.

1. Reliable and accurate output

The reliability of Machine Learning systems is primarily related to the fault tolerance and recoverability of the system in production. In addition, the reliability of ML systems is related to how reliable is the training process of the machine learning model.

**Table 2.** *Test Procedure for Reliable and Accurate Output.*

|  |  |
| --- | --- |
| Procedure | Expected Outcome |
| 1. ML model training process | Acceptable percentage |
| 1. ML model in production | Acceptable percentage |

Table 2 shows the reliable and accurate output.

## Evaluation Procedure

The Technological University of the Philippines (Manila) - Software Evaluation Tool will be used to evaluate the system. The factors and indicators are as follows: Functionality, Content, Reliability, Availability, Maintainability, and Saleability.

Functionality refers to a set of attributes that bear on ease of operation, provision for comfort, and convenience and user-friendliness of the system. Content refers to a set of attributes that bear on accuracy, updatedness and presentation of content. Reliability refers to a set of attributes that bear on conformance to desired results, absence of failure and accuracy in performance.

For Availability, the project will be evaluated according to performance to specifications, provision for security requirements and completeness of the project. For Maintainability, the project was evaluated according to ease of maintenance, provision for diagnostic tools and procedures, provision for enhancements and modifications.

Saleability refers to a set of attributes that bear on presence of market demand, competitiveness of price and attractiveness of the design. These are all the attributes of the factors and indicators to evaluate the developed system.

The procedure to evaluate the system are as follows:

1. Thirteen evaluators from different departments (i.e., Asset Department, UR-IT Project and Data Management, and Other Geoscientists). The respondents will be as follows: Data Controller Analysts (3), Data Analysts (3), Business Analysts (2), Chief Position Holder Asset (2), Systems Analyst (1), Programmer (2).
2. Each evaluator will be provided with a survey questionnaire.
3. Each evaluator will be given a chance to use the system and evaluate its performance.
4. The evaluator will use a survey questionnaire to carefully evaluate the system based on TUP-Manila Software Evaluation Tool.

**Table 3.** Five-Point Likert Scale

|  |  |
| --- | --- |
| **Scale Range** | **Descriptive Rating** |
| 5 | Excellent |
| 4 | Very Good |
| 3 | Good |
| 2 | Fair |
| 1 | Poor |

1. This study will utilize a 5-point Likert-type scale as shown in Table 3 and a scale range as shown with the corresponding descriptive rating. The results of the survey will be tabulated, and the mean per item and the overall mean will compute to determine the average rating of the respondents and the performance of the system. The software criteria include the following: functionality, content, reliability, availability, maintainability, saleability. Each criterion will be rated in a scale from 1 to 5, with five (5) being the highest, and one (1) being the lowest.

**Table 4. Scale Range and Its Descriptive Rating**

|  |  |
| --- | --- |
| **Scale Range** | **Descriptive Rating** |
| 4.51 – 5.00 | Excellent |
| 3.51 – 4.50 | Very Good |
| 2.51 – 3.50 | Good |
| 1.51 – 2.50 | Fair |
| 1.00 – 1.50 | Poor |