UNIVERSITY OF TUNIS EL MANAR
FACULTY OF SCIENCES OF TUNIS

Department of Geology

final year thesis

Presented for obtaining:

Bachelor's degree in Geomatics Applied to Earth Sciences

Hybrid Quantum-Classical Neural Network for resources exploration

By:

Mohamed Khiari

Presented on: 21 / 06 / 2024 in front of jury composed of:

Mr Adnène AMIRI	President
Mr Tarek SBOUI	Supervisor- FST
Mr Moncef BEN SMIDA	Supervisor- company
Mr Hafedh HAMZA	moderator

Project carried out in collaboration with: C.I.P.E.M

Academic Year 2023-2024

This internship is carried out within: CIPEM

..

CIPEM is a multidisciplinary company that specializes in providing high-level consultation and training in the fields of chemical, industrial, and petroleum engineering. Among its many strengths is its expertise in maintenance engineering, which helps manage costs and equipment by finding technical solutions. It also focuses on the production, transmission, and distribution of energy in industrial zones, providing efficient energy-saving implementation strategies. Its expertise ranges from basic to detailed, including mechanical, electrical, and earth-related installations and consumer alignment and charges

Abstract:

Hydrocarbon resource exploration is a vital and intricate undertaking that is typically marked by large financial outlays, thorough planning, and drawn-out procedures. This study offers a novel strategy to transform hydrocarbon discovery by fusing classical neural networks and quantum machine learning (QML). Our goals are to improve prediction accuracy, speed up the exploration process, and broaden the application of hydrocarbon identification by utilizing the computational capacity of quantum circuits.

In order to assess geological data and anticipate the presence of hydrocarbons, our methodology blends classical and quantum machine learning models, specifically using the U-net and Quantum Neural Network (QNN) models. In image segmentation tests, the QNN model exhibited varied accuracy between 0.61 and 0.87, highlighting both its potential and the current limitations of quantum computing. In contrast, the U-net model demonstrated robust performance with an accuracy of 0.92. The research required careful data preparation, which included putting geological data into quantum circuits and preparing satellite pictures. Efficient model convergence and optimized prediction capabilities were guaranteed by the hybrid training technique. According to our findings, the hybrid QML methodology considerably lowers the time and expense involved with conventional exploration techniques, offering a more effective and sustainable framework for resource discovery in the future. In addition to advancing the field of hydrocarbon exploration, this work opens the door for more widespread uses of QML in resource management and environmental monitoring. Modern quantum computing combined with traditional machine learning techniques provides a promising way forward for more precise, economical, and ecologically friendly exploration methods. By continually refining our models and fostering interdisciplinary collaboration, we aim to further unlock the potential of quantum-classical hybrid approaches, contributing to a more resilient and sustainable energy future.

Résumé :

L’exploration des ressources en hydrocarbures est une entreprise vitale et complexe qui se caractérise généralement par des dépenses financières importantes, une planification minutieuse et des procédures longues. Cette étude propose une nouvelle stratégie pour transformer la découverte d'hydrocarbures en fusionnant les réseaux de neurones classiques et l'apprentissage automatique quantique (QML). Nos objectifs sont d'améliorer la précision des prévisions, d'accélérer le processus d'exploration et d'élargir l'application de l'identification des hydrocarbures en utilisant la capacité de calcul des circuits quantiques.

Afin d'évaluer les données géologiques et d'anticiper la présence d'hydrocarbures, notre méthodologie mélange des modèles d'apprentissage automatique classiques et quantiques, en utilisant spécifiquement les modèles U-net et Quantum Neural Network (QNN). Lors des tests de segmentation d'images, le modèle QNN a présenté une précision variable comprise entre 0,61 et 0,87, soulignant à la fois son potentiel et les limites actuelles de l'informatique quantique. En revanche, le modèle U-net a démontré des performances robustes avec une précision de 0,92. La recherche a nécessité une préparation minutieuse des données, qui comprenait la mise en place de données géologiques dans des circuits quantiques et la préparation d'images satellite. Une convergence efficace des modèles et des capacités de prédiction optimisées a été garanties par la technique de formation hybride. Selon nos résultats, la méthodologie hybride QML réduit considérablement le temps et les dépenses impliqués par les techniques d'exploration conventionnelles, offrant ainsi un cadre plus efficace et plus durable pour la découverte de ressources à l'avenir. En plus de faire progresser le domaine de l'exploration des hydrocarbures, ce travail ouvre la porte pour des utilisations plus répandues de QML dans la gestion des ressources et la surveillance de l'environnement. L'informatique quantique moderne combinée aux techniques traditionnelles d'apprentissage automatique offre une voie prometteuse pour des méthodes d'exploration plus précises, plus économiques et plus respectueuses de l'environnement. En affinant continuellement nos modèles et en favorisant la collaboration interdisciplinaire, nous visons à libérer davantage le potentiel des approches hybrides quantiques classiques, contribuer à un avenir énergétique plus résilient et durable.

ملخص

يعد استكشاف الموارد الهيدروكربونية مهمة حيوية ومعقدة تتميز عادة بنفقات مالية كبيرة وتخطيط شامل وإجراءات طويلة. تقدم هذه الدراسة استراتيجية جديدة لتحويل اكتشاف الهيدروكربون من خلال دمج الشبكات العصبية الكلاسيكية والتعلم الآلي الكمي .(QML) تتمثل أهدافنا في تحسين دقة التنبؤ، وتسريع عملية الاستكشاف، وتوسيع نطاق تطبيق تحديد الهيدروكربونات من خلال الاستفادة من القدرة الحسابية للدوائر الكمومية.

من أجل تقييم البيانات الجيولوجية وتوقع وجود الهيدروكربونات، تمزج منهجيتنا بين نماذج التعلم الآلي الكلاسيكية والكمية، وتحديدًا باستخدام نماذج U-net والشبكة العصبية الكمومية .(QNN) في اختبارات تجزئة الصور، أظهر نموذج QNN دقة تتراوح بين 0.61 و0.87، مما يسلط الضوء على إمكاناته والقيود الحالية للحوسبة الكمومية. في المقابل، أظهر نموذج U-net أداءً قويًا بدقة 0.92. وتطلب البحث إعدادًا دقيقًا للبيانات، والذي تضمن وضع البيانات الجيولوجية في الدوائر الكمومية وإعداد صور الأقمار الصناعية. تم ضمان التقارب الفعال للنموذج وقدرات التنبؤ المحسنة من خلال تقنية التدريب الهجين. وفقًا للنتائج التي توصلنا إليها، فإن منهجية QML الهجينة تقلل بشكل كبير من الوقت والنفقات المرتبطة بتقنيات الاستكشاف التقليدية، مما يوفر إطارًا أكثر فعالية واستدامة لاكتشاف الموارد في المستقبل. بالإضافة إلى تطوير مجال استكشاف الهيدروكربون، فإن هذا العمل يفتح الباب لاستخدامات أكثر انتشارًا لـ QML في إدارة الموارد والمراقبة البيئية. توفر الحوسبة الكمومية الحديثة جنبًا إلى جنب مع تقنيات التعلم الآلي التقليدية طريقة واعدة للمضي قدمًا نحو طرق استكشاف أكثر دقة واقتصادية وصديقة للبيئة. ومن خلال التحسين المستمر لنماذجنا وتعزيز التعاون متعدد التخصصات، نهدف إلى إطلاق العنان لإمكانات الأساليب الهجينة الكمومية الكلاسيكية، المساهمة في مستقبل طاقة أكثر مرونة واستدامة.

Acknowledgements:

First of all, I want to thank God because He taught me that a debt of gratitude must be repaid. Through every trial and triumph, I've come to realize that with God by my side, I'm never alone.

I'm thankful for everything—the highs and lows, the comebacks and setbacks that made me who I am today.

I eagerly anticipate the presence of my family. To my dad, who taught me how to be resilient, thank you for making countless sacrifices along the way.

To the most important person in the whole world, my mother, a wonder woman, whose boundless love and steadfast prayers have been my guiding light throughout There are 8 billion people on this planet, and you are the perfect mom for me.

I wish to offer a profound acknowledgement to Mr. Moncef Ben Smida. His generosity in offering me the internship opportunity at CIPEM has been truly appreciated. returning the favor for his guidance and mentorship I'm thanking him today for making me developing insights and skills that have significantly enriched my understanding of the industry. I am thankful for the confidence he put in me during my tenure at CIPEM.

My heartfelt gratitude extends to Mr. Tarek Sboui, my supervisor. I am deeply grateful for your enduring patience, invaluable guidance, and devoted encouragement. Your valuable lessons have not only increased my interest in the field but also fueled my desire to excel in it. Thank you for being a building block of my academic and professional journey.

Lastly, I am deeply grateful to my roommates, who have stood by me through every challenge because this honor doesn't just belong to me i wouldn't come this far If it weren't for certain significant individuals in my life, beginning with my mother, father, brother, and my other family, who I'm so glad to have

i was under the misconception that my achievements were solely my own. Nothing can be further from the truth. I've been uplifted, supported, motivated, and embraced by an extraordinary circle of companions unlike any other Rayen Dabbabi, Rayen Ben Hassen, and Mouath Haffar

being here today, brimming with gratitude, I realize that my journey is just beginning. I carry the lessons, love, and support you've all given me as precious fuel for the road ahead. As I step forward, I promise to honor your encouragement, pay it forward with kindness, and strive to make you all proud. Thank you, for being the source of empowerment.

Table of content

List of figures

List of tables

1 Chapter: Introduction

1.1 General Context :

In the evolving field of energy exploration, the search, for environmentally friendly methods of extracting hydrocarbons has brought together traditional geoscience and advanced technology.

With the demand for energy on the rise, there is a growing need to innovate hydrocarbon exploration methods. This initiative marks an effort that combines geoscience with modern artificial intelligence. Historically, discovering and describing oil and gas reservoirs has heavily relied on established models and time-consuming surveys. However, the emergence of machine learning offers an opportunity to enhance these approaches with data-driven insights and predictive analysis.

quantum machine learning (QML) for hydrocarbon exploration, with the goal of improving reservoir identification and optimizing energy extraction. The project's goal is to open up new, sustainable routes for extracting vital energy resources by integrating technology and geoscience.

1.2 problem and objectives:

1.2.1 Problem

The traditional approach to hydrocarbon exploration presents financial fortitude due to the substantial upfront investments, meticulous planning, and assembly of skilled teams needed. This process demands a significant amount of money, ranging from $8 million onshore to over $100 million offshore. To navigate this economic landscape more effectively, innovative solutions like the exploration of quantum machine learning (QML) are very promising.

By potentially optimizing drilling locations, reducing exploration time, and minimizing resource waste, a groundbreaking approach can revolutionize exploration methodologies and potentially alleviate the financial burdens associated with traditional methods.

1.2.2 Objectives :

- Revolutionize traditional hydrocarbon exploration by integrating quantum machine learning (QML) to significantly reduce the substantial financial investment and meticulous planning associated with the process.
- Leverage the unique capabilities of QML to achieve unprecedented accuracy in reservoir predictions, optimize resource extraction processes, and minimize environmental impact.

- Develop a novel, efficient, and cost-effective approach to hydrocarbon exploration that promotes environmental sustainability, thereby contributing to the advancement of the energy industry and addressing the limitations of traditional methods

Chapter 2: LITERATURE REVIEW

2 Chapter: LITERATURE REVIEW

2.1 Introduction

Finding hidden oil and natural gas resources is a major challenge. Traditional exploration methods can be very slow, taking years to get results. This process not only slows down new discoveries but also often overlooks the potential of natural gas reserves. We propose a new approach to make finding oil and gas easier by combining quantum mechanics with regular machine learning. This new hybrid strategy could change how we explore hydrocarbons by:

Accelerating the Exploration Process:

Quantum machine learning can work with difficult data quickly. It unravels complex datasets linked to oil and gas exploration. Old machine learning methods battle with intricate geological data. However, quantum algorithms like the Variational Quantum Eigen Solver (VQE) analyze this data much faster. As a result, possible locations for oil and gas reservoirs can be identified rapidly.

Expanding the Hydrocarbon Scope:

Our plan goes beyond typical ways. We look for oil and natural gas together when exploring underground. This gives us a fuller picture of what's beneath the surface.

Enhancing Predictive Accuracy:

The fusion of QML and ML holds the promise of unlocking a new level of predictive accuracy. Our enhanced technology allows for pinpointing probable mineral reservoirs more precisely.

2.2 The Power Beneath Our Feet: Geology and the Exploration of Shale Oil and Gas

Our journey begins with the fascinating world of geology, the key to unlocking the secrets hidden beneath our feet, from valuable resources to beautiful minerals. Starting with the historical significance of hydrocarbons and contrasting the properties and uses of oil and gas, covering these concepts will pave the way for examining the challenges faced by traditional methods. Next, we will delve into how AI can revolutionize this industry by overcoming long-standing limitations and unlocking new possibilities for the future.

2.2.1 geology:

Geology is the study of the Earth's structure, composition, and processes. using a variety of techniques to locate where sizable quantities of hydrocarbon exist under the surface, primarily employing seismic imaging, which can also be called an ultrasound, which involves sending sound waves into the earth and recording the echoes that bounce back. Based on these records, geologists can create a prototype of the rock layers, leading them to determine the depth and thickness of the shale formations and their size as well. Year after year, we've seen the demand for natural resources such as oil and gas progressively increase, but onshore "conventional" plays have struggled to keep up. Big firms needed a solution to continue the demand, and shifting to unconventional methods was an option to grow production volumes and secure their country's energy future. Sedimentary rocks like shale can be full of organic matter, containing remnants of plants and algal material preserved inside of them. These rocks are buried deeper beneath the earth's surface, where the presence of the best factors like pressure and temperature increases over time, backing hydrocarbons like oil and gas. After identifying the targeted areas, geologists collect rock samples during drilling. These samples provide valuable information on the composition and characteristics of the shale, like porosity and permeability. They can also run tests on the fossils in the rock to help identify the age of the shale, leading them to a better understanding of whether it's from the from the Cenozoic or any other geologic past in which the rock was formed, the environment, categorization, and organic matter composition. Based on that, they can determine the potential quantity and quality of the matter (gas and oil) that could have been generated over time. After gathering all the data, computer models simulate the movement of the oil. (desRosiers, 2023)

2.2.2 A Century of Progress:

The natural gas and oil industry, which began in the mid-1800s, has grown rapidly and significantly impacted the world. Oil was initially used for lighting and lubrication, but as the world became industrialized, it became essential for producing various products, including plastics and gasoline. The discovery of natural gas in the early 1900s increased its importance. However, environmental concerns and fluctuating prices posed challenges. Hydraulic fracturing, or "fracking," revolutionized the industry by allowing companies to extract natural gas and oil from previously inaccessible sources. Renewable energy sources like wind and solar power have also gained popularity. The industry continues to innovate and adapt to new challenges, using artificial intelligence to improve exploration and production processes and investing in carbon capture and storage technology to reduce environmental impact and ensure a sustainable future.

2.2.3 gas vs oil:

We've talked about hydrocarbons so far, which raises the question: What's the difference between oil and gas?

biologically speaking, they both come from prehistoric organisms that decomposed underground, as we talked about earlier, but the main difference in the composition and physical state is that crude oil is a complex format of hydrocarbon compound; along with hydrogen and carbon atoms, there is also a small amount of sulfur, nitrogen, and oxygen. Natural gas, on the other hand, is a simple chemical compound of methane (CH4); it may also have some propane. Also, crude oil is a liquid at standard atmospheric conditions, while natural gas is a gaseous state hydrocarbon lighter than air. Extraction methods can also be different for the oil. After drilling, additional techniques can be applied, like hydraulic fracturing and horizontal drilling, while for the gas, instead of releasing oil, it releases gas. (Yuewei Wang, 2023)

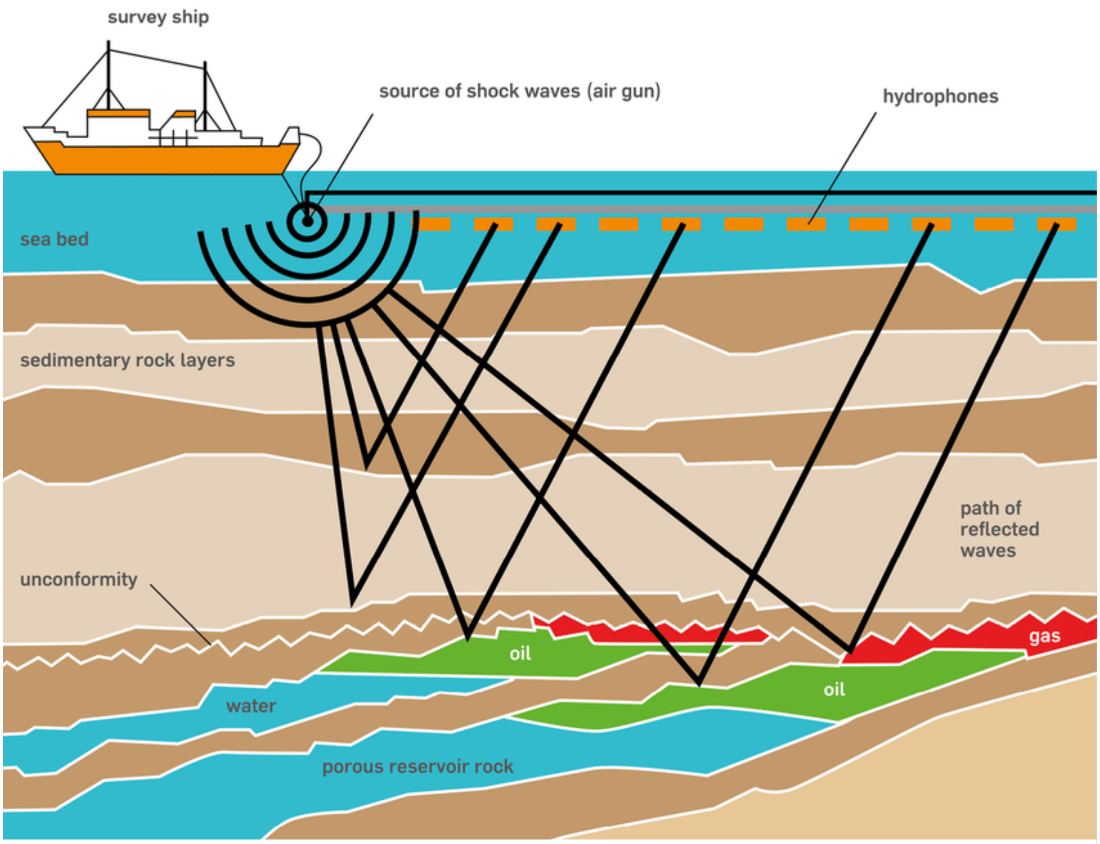


Figure 1: a simplified illustration of a marine seismic survey

2.3 Challenges and Limitations of Traditional Exploration Methods:

Traditional hydrocarbon exploration methods face challenges such as identifying potential sources, high costs, and time consumption. These methods often struggle to access resources deep underground or in environmentally sensitive areas, and data acquisition and analysis can lead to inaccurate assessments. Additionally, the exploration process can take years and requires complex reach and depth penetration capabilities. These challenges contribute to inefficiencies and increased costs in hydrocarbon exploration, paving the way for advancements like artificial intelligence to offer transformative solutions in the future.

2.4 Bridging the Gap from Remote Sensing to AI in Geomatics

2.5 Introduction :

This chapter connects the dots between technologies, such as remote sensing features, and ai within the field of geomatics.

2.6 Unveiling the Earth's Secrets: Remote Sensing:

Remote sensing is a powerful tool, providing valuable insights into Earth's surface and subsurface characteristics without physical contact. This technology utilizes various platforms to collect data from a distance, helping us see the Earth in ways traditional exploration methods cannot.

2.6.1 Satellite Imagery: A Window to the Earth's Surface

Satellite imagery is a transformative technology that reshapes our perception and understanding of our planet. These images, captured by satellites, offer invaluable insights into Earth's features, anomalies, and patterns, as well as responses to various issues.

2.6.1.1 Satellite Imagery Definition:

Literal to its name, Satellite Imagery refers to images captured by satellites, presenting a digital visual representation of the Earth's surface through cameras or sensors mounted on satellites orbiting the Earth.

2.6.1.2 Active and Passive Satellites:

Satellites are categorized as active or passive. Active Satellites use remote sensors to detect reflected responses from objects irradiated by artificially generated energy sources. In contrast, Passive Satellites use sensors to detect reflected or emitted electromagnetic radiation from natural sources, such as the sun, magnetism, or geothermal activity.

2.6.2 Types of Satellite Imagery :

2.6.2.1 1-Visible Satellite Imagery:

Captures images using satellites to detect and record visible light wavelengths, providing a visual representation of the Earth's surface and cloud cover. Available only during the day.

2.6.2.2 2.Infrared (IR) Satellite Imagery:

Captures infrared radiation emitted or reflected by objects on the Earth's surface, highlighting temperature variations. It is valuable for weather forecasting, temperature analysis, and detecting heat signatures in different environments. Available day and night.

2.6.2.3 3.Water Vapor Satellite Imagery:

Designed to detect the concentration and movement of water vapor in the Earth's atmosphere, offering insights for weather analysis, moisture tracking, and predicting atmospheric instability. Crucial for identifying potential rainfall or thunderstorm development.

2.6.3 Resolution Matters:

Resolution is a pivotal aspect in satellite imagery, determining the quality and quantity of the captured images.

2.6.3.1 Spatial Resolution:

Refers to the level of detail captured in an image, determining the smallest distinguishable object. Higher spatial resolution allows for the identification of smaller objects, providing a more detailed representation of the Earth's surface.

2.6.3.2 Spectral Resolution:

Relates to a sensor's capacity to discern and capture specific wavelength intervals within the electromagnetic spectrum. High spectral resolution provides detailed information about the Earth's surface composition.

Multispectral and Hyperspectral Sensors:

Multispectral sensors capture data in specific spectral bands, enhancing our understanding of the environment. Hyperspectral sensors collect detailed information about materials and substances across numerous narrow spectral bands.

(Roberts, 2022)

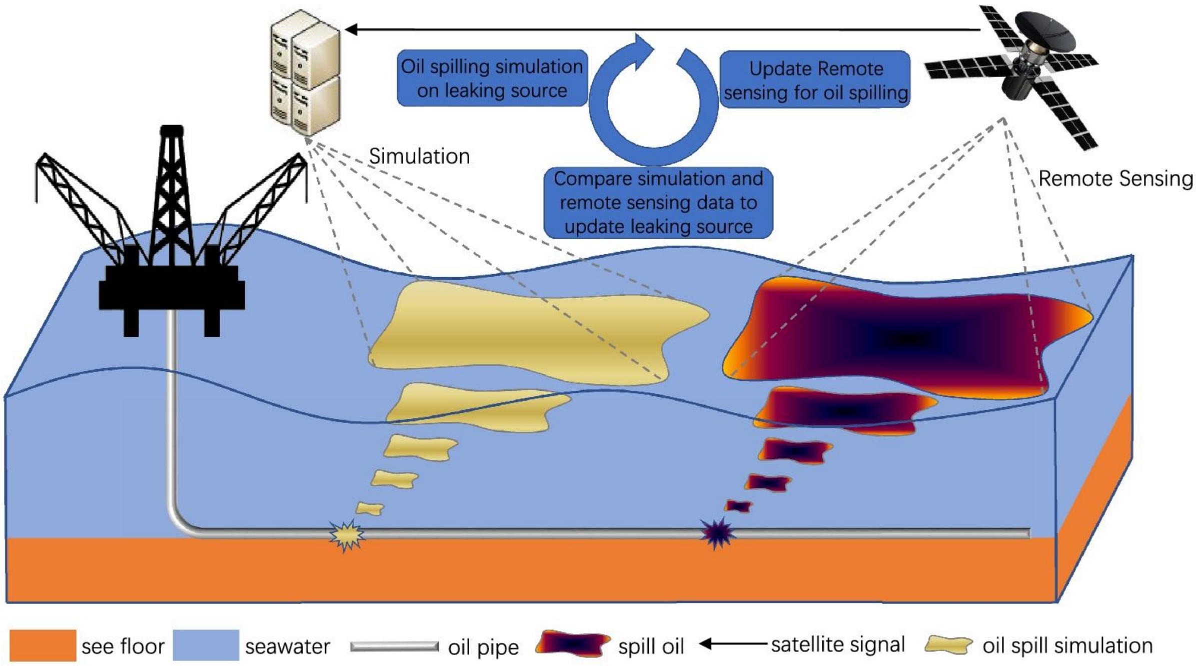


Figure 2: Cyber-physical oil spill monitoring and detection for offshore petroleum risk management service

2.7 Satellite Imagery: Landsat

The Landsat program, launched in the 1960s under Secretary Stewart Udall, has been a pioneer in remote sensing, providing valuable data and images of the Earth's surface. With eight operational satellites, the program has been instrumental in observing changes induced by human activities, such as urbanization and agricultural practices. The data, which can identify larger roads, buildings, and urban development patterns, has been used for multidecade comparisons. Despite facing challenges, such as commercialization in the 1980s, the government reclaimed the program, leading to the more accessible Landsat 7 data. Landsat 5, the longest-operating satellite, operated for 28 years, and Landsat 8 successfully took over after its decommissioning. Despite technical challenges, the Landsat program has revolutionized our understanding of the Earth's surface and continues to monitor environmental changes worldwide. The upcoming Landsat Next mission in 2030 promises further advancements, including a constellation of three observatories for enhanced temporal revisit, expanded spectral bands, and increased data collection capacity.

2.7.1 Landsat 9:

Launched on September 27, 2021, Landsat 9 is the newest satellite in the Landsat program, which has been providing valuable Earth observation data since the early 1970s. Landsat satellites take high-resolution pictures of the Earth's surface, which are utilized for a variety of applications such as agriculture, urban planning, forestry, environmental monitoring, and disaster relief. The explanation of Landsat 9 and the factors that make it superior to Landsat 8 is provided below:

Better Sensor Technology: The Thermal Infrared Sensor 2 (TIRS-2) and Operational Land Imager 2 (OLI-2) in Landsat 9 are better sensors than those on Landsat 8. These sensors' improved spectral, radiometric, and spatial capabilities result in crisper, more detailed images of the Earth's surface.

Continuity of Data: Landsat 9 ensures the continuity of the Landsat program's data record, which is crucial for monitoring long-term changes in Earth's environment. By providing consistent, high-quality imagery, Landsat 9 contributes to ongoing research and applications in areas such as land use planning, natural resource management, and climate change monitoring.

Extended Lifespan: Landsat 9 is designed to operate for at least five years, with the potential for an extended mission lifespan. This ensures a reliable and continuous stream of data for the scientific community, allowing for the monitoring of both short-term events and long-term trends.

Overall, Landsat 9 represents a significant advancement in Earth observation technology, building upon the success of previous missions like Landsat 8. With its enhanced sensors, improved spatial and temporal resolution, and continuous data, Landsat 9 will provide researchers and decision makers with valuable insight into the dynamic processes that shape the Earth. (mission, 2024)

Nine spectral bands:

- Band 1 Visible Coastal Aerosol (0.43 - 0.45 µm) 30-m
- Band 2 Visible Blue (0.450 - 0.51 µm) 30-m
- Band 3 Visible Green (0.53 - 0.59 µm) 30-m
- Band 4 Red (0.64 - 0.67 µm) 30-m
- Band 5 Near-Infrared (0.85 - 0.88 µm) 30-m
- Band 6 SWIR 1(1.57 - 1.65 µm) 30-m
- Band 7 SWIR 2 (2.11 - 2.29 µm) 30-m

Band 8 Panchromatic (PAN) (0.50 - 0.68 µm) 15-m

Band 9 Cirrus (1.36 - 1.38 µm) 30-m

Thermal Infrared Sensor 2 (TIRS-2)

Landsat 9's Thermal Infrared Sensor 2 (TIRS-2) measures thermal radiance emitted from the land surface in two thermal infrared bands using the same technology that was used for TIRS on Landsat 8, however TIRS-2 is an improved version of Landsat 8's TIRS, both with regards to instrument risk class and design to minimize stray light. TIRS-2 provides two spectral bands with a maximum ground sampling distance, both in-track and cross track, of 100 m (328 ft) for both bands. TIRS-2 provides an internal blackbody calibration source as well as space view capabilities. TIRS-2 is designed by NASA Goddard Space Flight Center in Greenbelt, Maryland. (surveey, 2022)

Two spectral bands:

- Band 10 TIRS 1 (10.6 - 11.19 µm) 100-m
- Band 11 TIRS 2 (11.5 - 12.51 µm) 100-m

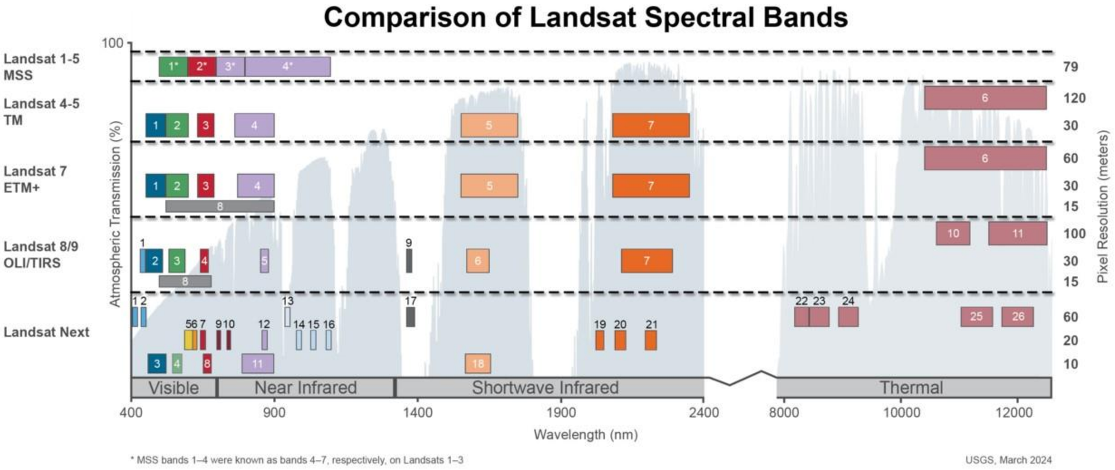


Figure 3: Landsat Spectral Bands

2.8 Region of interest (Kansas City)

The latitude of Kansas City, Missouri, USA is 39.099724 and the longitude is -94.578331. Kansas City, Missouri, United States is located in the United States in the urban areas zone with GPS coordinates 39° 5' 59.0064" N and 94° 34' 41.9916" W

Kansas City, located at the confluence of the Kansas and Missouri Rivers, is a vibrant metropolis with a rich history, cultural diversity, and economic significance. Its strategic location near the geographic center of the United States indicates its vital role as a crossroads of transportation, commerce, and cultural exchange. The city's oil and gas history dates back to the late 1800s, when miners began tapping underground reserves beneath the Midwest. The city's energy exploration landscape has evolved over the decades, with advances in technology leading to more efficient extraction methods and increased production rates. Kansas City has also embraced renewable energy sources, such as wind farms, solar farms, and biofuel plants, reflecting the growing global trend towards sustainable development and environmental protection. The city's vibrant cultural life, including a world-famous barbecue and thriving arts community, continues to attract visitors.

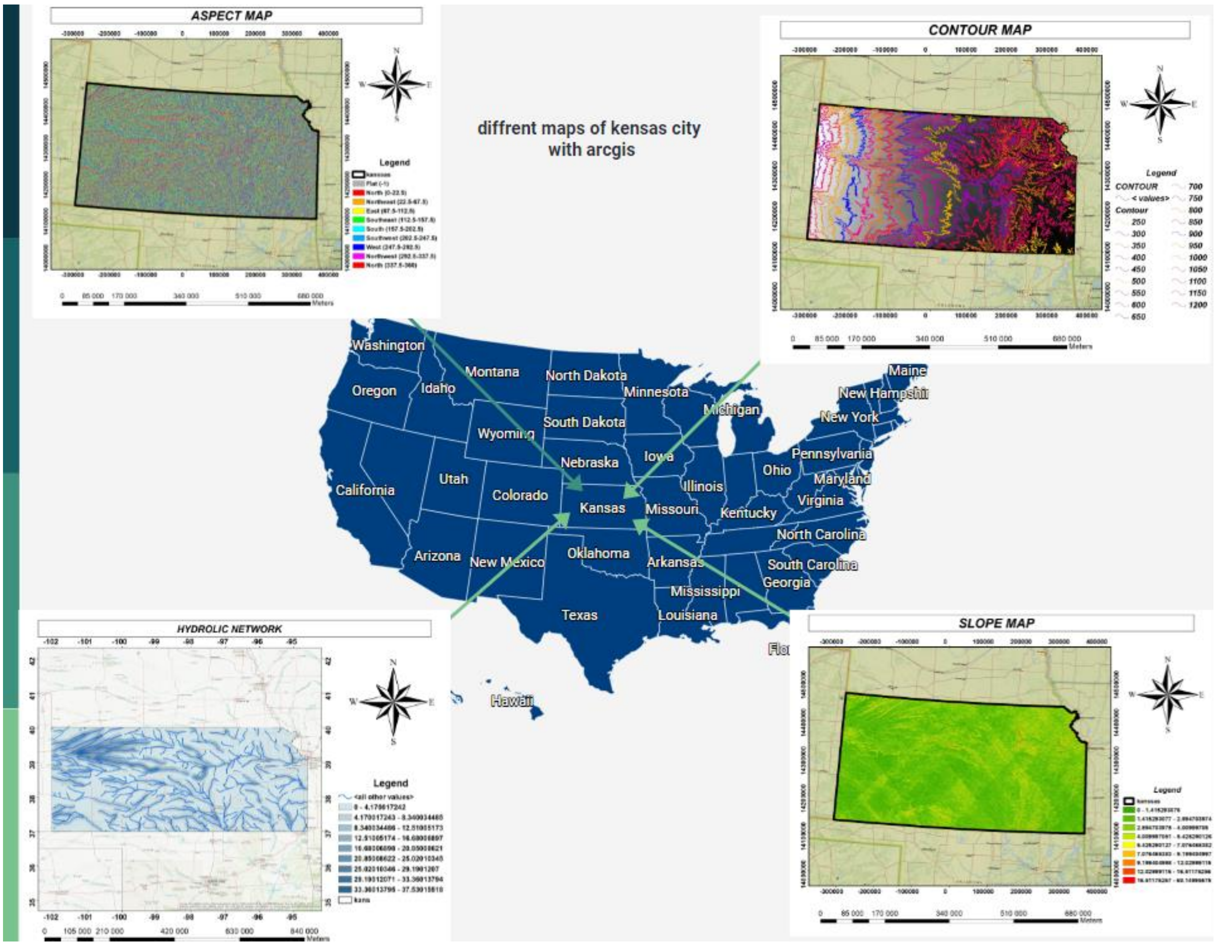


Figure 4: representation of different maps of Kansas City

2.9 limitations of remote sensing data:

In the realm of remote sensing, the orchestration of data encounters its set of limitations a symphony where spatial and temporal resolutions harmonize with atmospheric nuances, spectral paucities, and the cadence of cost and accessibility challenges. Sensory intricacies and the labyrinth of vegetation and topography stand as artistic obstacles, painting the canvas of interpretation with complexity. Processing, akin to crafting a masterpiece, unveils its own challenges. These constraints, though formidable, are integral notes in the melody of data interpretation. Users, as the virtuosos of this grand composition, find it paramount to attune themselves to these nuances, understanding that amidst limitations, the symphony of remote sensing unfolds its richness for diverse applications.

2.9.1 Bridging the Gap: Integrating GIS and Geomatics

In the never-ending quest to find new stuff we need; a wide range of tools and techniques is needed to maximize success. This section explores how geographic information systems (GIS) seamlessly integrate with geoscience and remote sensing data to create a more efficient advanced research practices

2.9.1.1 the Power of GIS:

GIS plays a pivotal role in resource exploration by acting as a central repository and analysis platform for various types of data. It acts as a digital bridge, connecting the dots between:

- **Remote sensing data:** Satellite imagery, radar data, and other remotely sensed information can be integrated into a GIS platform for analysis and visualization.
- **Geospatial data:** Existing datasets, such as geological maps, topographic data, and infrastructure information, can be incorporated into the GIS for comprehensive analysis.
- **Fieldwork data:** Data collected during field surveys, including GPS coordinates, physical samples, and observations, can be integrated for further analysis and visualization

2.9.1.2 Understanding the Role of Geomatics:

Geomatics encompasses a diverse range of techniques and technologies utilized to acquire, manage, analyze, and interpret spatial data related to the Earth. It is vital in resource exploration by providing:

- **Surveying:** Traditional and advanced surveying techniques, like GPS and LiDAR, allow for the precise measurement of the Earth's surface, facilitating the identification of potential resource locations and providing critical data for further analysis.
- **Seismic Surveys:** Subsurface exploration techniques, such as seismic surveys, offer valuable insights into the Earth's geological structure, aiding in identifying potential resource deposits hidden beneath the surface.
- **Geospatial modeling:** Creating three-dimensional models of the Earth's subsurface structure can help visualize potential resource locations and guide further exploration activities.

The true power lies in **integrating** these technologies. By combining GIS, remote sensing data, and geomatics information, we can gain a comprehensive understanding of the target exploration area:

- **Spatial analysis:** GIS allows for overlaying and analyzing various datasets, enabling the identification of spatial relationships and patterns that might be indicative of potential resource locations.
- **Decision-making support:** The combined insights from different data sources provide valuable information for informed decision-making throughout the exploration process.
- **Enhanced visualization:** Integrating data into GIS allows for creating visual representations, such as maps and 3D models, facilitating better communication and collaboration between different stakeholders involved in the exploration project.

1. Conclusion

The integration of GIS, geomatics, and remote sensing data plays a pivotal role in modern resource exploration practices by understanding the individual strengths of each technology and embracing their synergy, we can unlock the full potential of remote sensing data and pave the way for sustainable and responsible resource exploration practices

2.10 artificial intelligence: New Solutions on the Horizon:

Artificial intelligence (AI) is a game changer by offering a powerful set of tools to overcome the limitations of traditional methods. a subset of AI that enables algorithms to learn from vast datasets without explicit programming. By analyzing enormous volumes of geological, seismic, and production data, ML models can identify subtle patterns and correlations that would be impossible for humans to detect. This capability allows AI to predict the location of hydrocarbon reservoirs with greater accuracy, optimize drilling and extraction techniques, and reduce exploration costs. a 2.0 form of machine learning called quantum machine learning (QML) is showing potential for further investigation. the principles of quantum mechanics, specifically the behavior of qubits, which can exist in multiple states simultaneously, unlike the bits in traditional computers. This unique property allows QML algorithms to explore a vast number of possibilities simultaneously and calculate the outcomes three times faster than average, potentially leading to breakthroughs in data analysis and problem-solving. While still in its early stages, QML has the potential to further revolutionize exploration by tackling complex geological challenges with unprecedented power and efficiency.

2.11 Harnessing the Power of Intelligence: AI/QML in Exploration

Geo AI is a new research and application field combining spatiotemporal big data analysis and artificial intelligence technology These cutting-edge technologies bring unprecedented capabilities to analyze vast datasets, enabling the identification of intricate patterns and trends for efficient and insightful exploration practices.

Old methods often rely on human expertise for data analysis. AI/ML algorithms, however, possess the ability to:

- Process vast quantities of data: They can analyze massive datasets from remote sensing, GIS, and other sources, identifying patterns and trends that might be missed by human analysts.
- Identify subtle relationships: AI/ML can identify subtle relationships between different data points, providing valuable insights into potential resource locations and geological characteristics.
- Automate repetitive tasks: They can automate time-consuming and repetitive tasks like data cleaning and feature extraction, freeing up valuable resources for further exploration activities.

2.11.1 Embracing a Hybrid Approach: Classical vs. Quantum ML

Classic:

Classical bits: Operates on bits, which can be either 0 or 1. Think of a light switch, on or off.

Sequential processing: Analyzes data one piece at a time, similar to solving a math problem Limited by complexity: Struggles with highly complex problems with many variables.

Quantum:

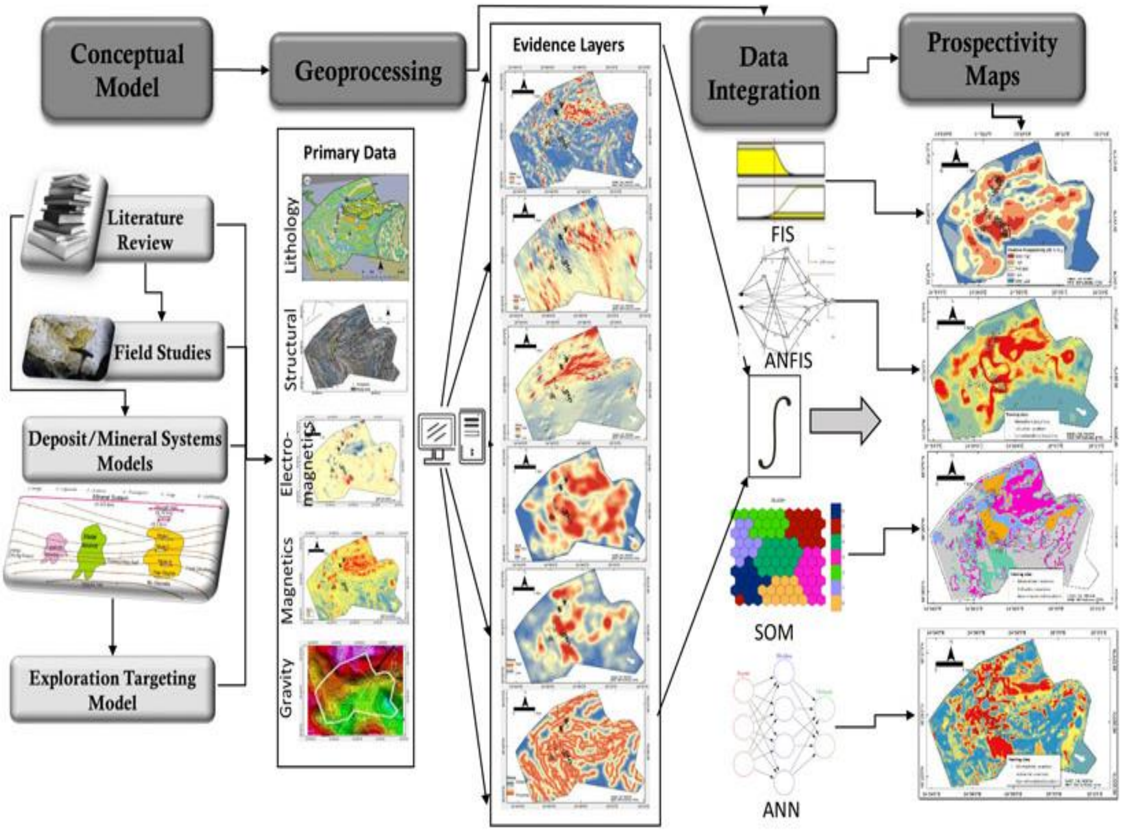
Quantum bits (qubits): Utilizes qubits, which can hold both 0 and 1 simultaneously, a concept known as superposition. Imagine a light switch being both on and off at the same time (not possible in our current understanding of physics, but an important concept in quantum mechanics).

Parallel processing: Explores multiple possibilities simultaneously, like checking all the answers in a math test at once.

Potential for tackling complex problems: Holds promise for solving problems intractable for classical ML due to their inherent complexity.

While QML holds immense potential, full-fledged quantum computers are still under development. Therefore, this project adopts a hybrid approach We combine the strengths of traditional ML, which excels at handling large datasets, with the potential benefits of QML for specific tasks like the complexity of the data and the GPU usage

(Ferrie, 2023)



(Ferrie, 2023)

Figure 6: Illustration of a Collaborative Exploration Workflow for Desert Mineral Deposit

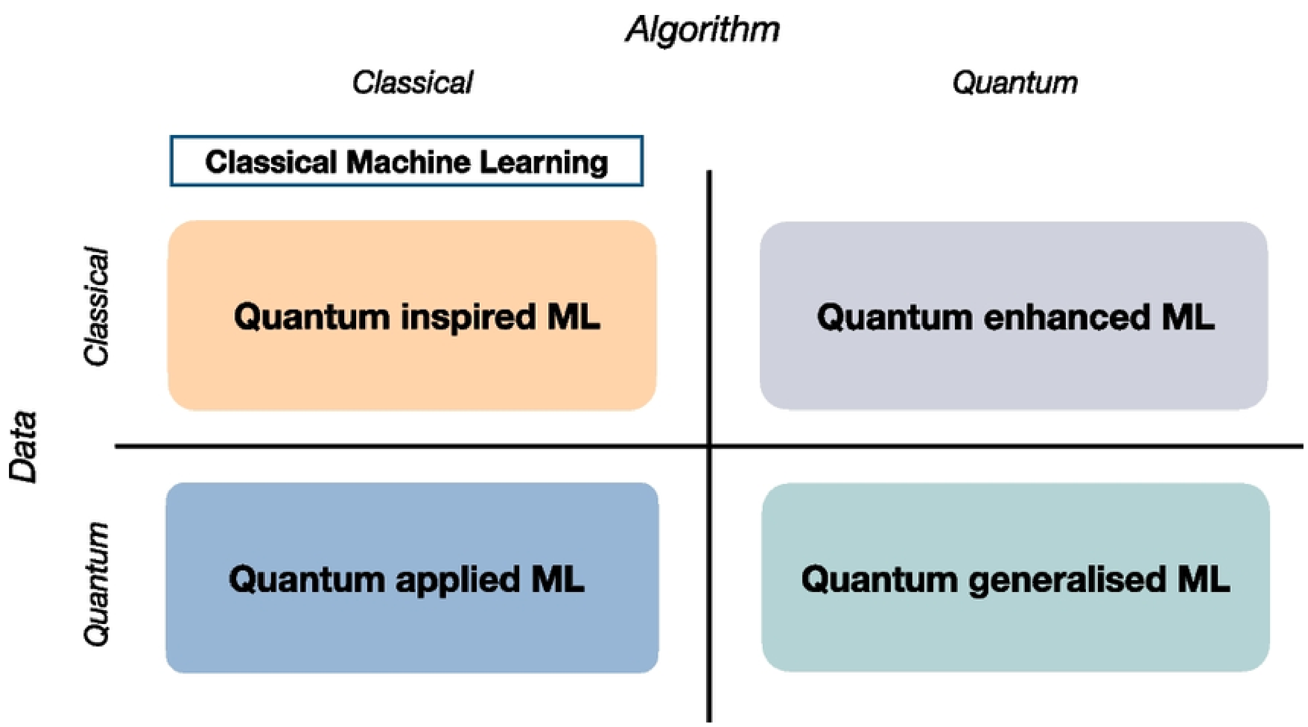


Figure 5: Qnn vs classic ml paradigm

Chapter 3 : Study conception: from pixels to predictions

3 Chapter: Study conception: from pixels to predictions

3.1 Introduction :

In this chapter, we are going to dig deeper into the roadmap of our AI/QML-driven resource exploration project. Our primary objective is to outline the primary steps of the workflow, detailing the specific software tools we will utilize and the different types of data we will employ in the data acquisition process for various sources. We aim to establish a solid foundation for building and applying our model for successful search.

3.2 Software Selection:

The building blocks of any study are the proper selection of the right software, and as GIS scientists, we know for fact that ArcGIS and other GIS software are crucial for this. With our finest programs, we begin with ArcGIS.

3.2.1 ArcMap:

<u>Strengths:</u>

Maturity and Stability: ArcMap has been around for a longer time, offering a mature and stable platform with a vast library of extensions and functionalities.

Customization: ArcMap offers a high degree of customization through add-ins and extensions, allowing users to tailor the software to their specific needs.

<u>Weaknesses:</u>

Limited Support: Esri has shifted its focus to ArcGIS Pro, resulting in diminished support for ArcMap.

Future updates and bug fixes might be less frequent.

32-bit Architecture: ArcMap operates on a 32-bit architecture, limiting its ability to handle very large datasets. Or 64-bit float architecture

Limited Python 3 Support: ArcMap's Python scripting capabilities primarily rely on Python 2.7, which is reaching end-of-life.

3.2.2 ArcGIS Pro:

<u>Strengths:</u>

Modern Architecture: ArcGIS Pro leverages a 64-bit architecture, allowing it to handle large and complex datasets more efficiently.

Advanced Functionality: ArcGIS Pro offers a wider range of built-in functionalities compared to ArcMap, including advanced spatial analysis tools and improved 3D visualization capabilities. Active Development: Esri actively develops and updates ArcGIS Pro, ensuring access to the latest features and bug fixes.

Python 3 Support: ArcGIS Pro fully supports Python 3, opening doors for leveraging a wider range of Python libraries and functionalities for geospatial analysis and AI/QML integration.

<u>Weaknesses:</u>

Learning Curve: Due to its newer interface and functionalities, ArcGIS Pro might have a steeper learning curve for users accustomed to ArcMap.

Limited Customization: While customization options exist, they are not as extensive as those offered by ArcMap.

While both ArcMap and ArcGIS Pro are valuable tools ArcGIS Pro aligns better with the contemporary approach to geospatial analysis, particularly its integration with Python 3, which is more than necessary for AI/QML development

The 64-bit architecture of ArcGIS Pro facilitates working with more complex dataset like ours, which is likely encountered in resource exploration involving satellite imagery and other geospatial information.

3.2.3 QGIS (Quantum GIS):

Open-source: Free and readily available, making it a cost-effective option for our case as a student or organizations with budget constraints.

Wide Functionality: Offers a comprehensive set of functionalities for GIS data visualization, spatial analysis, and basic image processing.

Python Integration: Similar to ArcGIS Pro, QGIS supports Python scripting, allowing for custom workflows and integration with AI/QML libraries (though potentially requiring more technical expertise compared to ArcGIS Pro).

<u>Strengths:</u>

Strong user community with extensive online resources and tutorials.

Offers plugins for specialized tasks, potentially including some related to AI/QML (although these might be less developed compared to commercial software).

<u>Weaknesses:</u>

Limited native support for advanced image processing tasks often encountered in resource exploration (compared to specialized software like ENVI).

require more technical expertise to set up and customize workflows compared to user-friendly commercial options.

3.2.4 ENVI (Environment for Visualizing Images)

Commercial Software: Paid software with a licensing fee, offering a range of functionalities tailored for remote sensing image processing and analysis.

Advanced Image Processing: ENVI excels in advanced image processing tasks like atmospheric correction, spectral band manipulation, and feature extraction, which is exactly what we need in this particular project for preparing remote sensing data for resource exploration applications.

Specialized Workflows: Offers pre-built workflows and tools designed specifically for resource exploration tasks, potentially including mineral mapping or hydrocarbon exploration.

Weaknesses:

Cost: The commercial license can be expensive, especially for individual users or smaller research projects.

Slower Learning Curve: The extensive functionalities might require more time and effort to master compared to user-friendly GIS platforms like ArcGIS or QGIS.

3.2.5 MAGMAP (Magnetic Mapping)

Commercial Software: MAGMAP is a paid software solution with a licensing fee, catering to professionals and organizations involved in geophysical exploration and research. Its pricing model typically includes commercial licenses, which can vary depending on the scale of usage and specific requirements.

Advanced Magnetic Data Processing: MAGMAP stands out for its advanced capabilities in magnetic data processing and interpretation. It offers a suite of tools and algorithms tailored to handle magnetic data effectively. This includes functionalities such as data visualization, filtering, modeling, and interpretation, essential for analyzing our magnetic anomalies and identifying potential subsurface structures. We used it to manipulate our magmatic map and the LAS file

Specialized Workflows: MAGMAP provides specialized workflows and tools explicitly designed for resource exploration tasks, particularly in the field of geophysics. These workflows are optimized for magnetic data interpretation and analysis, covering various exploration activities such as mineral mapping and hydrocarbon exploration.

By offering pre-built workflows and specialized tools, MAGMAP streamlines the exploration process, enabling efficient data analysis and decision-making.

Weaknesses:

Cost: One of the primary drawbacks of MAGMAP is its cost. The commercial license is very expensive, Organizations or research teams considering MAGMAP may need to allocate sufficient resources to cover the licensing fees.a variety of software platforms cater to different needs and project requirements. Having experience with all of them allows for a flexible approach. from data acquisition and processing to model development and analysis. It also showcases my expertise in working with all the platforms and explains how each can contribute to the overall workflow. Now that we have covered all the software we need, let's move a step forward into data analysis and types.

3.3 Data Acquisition: Gathering the Raw Materials

3.3.1 Well Logs & Geologic Information: Subsurface Insights

This subsection highlights the importance of well logs and geologic information for the project. with a breakdown of the key points.

Data Description:

(VYAS, 2020)

Well logs are digital information compiled in the course of the drilling procedure of oil and fuel line wells, supplying important insights into subsurface situations. They embody numerous parameters such as:

Depth Measurements:

Records the intensity at which every information factor is measured alongside the wellbore, facilitating unique evaluation and correlation of geological features.

Lithology:

Describes the varieties of rocks encountered in the course of drilling, helping in information the geological formations and predicting reservoir characteristics.

Porosity:

Indicates the share of void areas inside the rock formation, important for assessing the reservoir`s cap potential to save hydrocarbons.

Permeability:

Reflects the rock's cap potential to permit fluid waft thru its pore areas, influencing the convenience with which oil, fuel line, or water can circulate inside the reservoir.

Fluid Content:

Identifies the presence and distribution of various fluids (e.g., water, oil, fuel line) inside the formation, critical for estimating hydrocarbon reserves and making plans manufacturing strategies.

These information factors are pivotal for reservoir characterization, formation evaluation, and decision-making approaches in oil and fuel line exploration and manufacturing operations.

Well log’s function vital equipment for geoscientists and engineers to recognize subsurface situations and optimize drilling and manufacturing strategies

Geological information plays a pivotal role in hydrocarbon exploration, particularly in regions like Kansas City. where understanding the subsurface environment is essential.

Here's a manifestation of well logs and geological data in this context:

- Identification of Potential Reservoirs: This kind of data provides accurate information about the geological formations present in our region of interest. Analyzing the data helps us identify potential hydrocarbon reservoirs based on the presence of the right type of rock formations. These formations may include the right combination of characteristics, such as porosity, permeability, and even the location.
- Delineation of Trap Structures: Some structures, like anticlines, faults, and stratigraphic traps, can act as traps for hydrocarbons. With the right treatment of geological maps and seismic data, geologists can identify these structural features and assess their potential to trap hydrocarbons. Well logs then provide detailed information about the geometry and characteristics of these traps, which is important for their identification.
- Depth and Thickness: Well logs provide data on the depth and thickness of different geological layers, including potential mineral-bearing formations. helping in the planning and execution of drilling operations.
- Validation Using Historical Well Data: Geological interpretations and predictions can greatly benefit from the use of historical well data from earlier drilling operations in the region. Geologists can analyze and correlate geological formations in various areas using well logs from existing wells, which enhances the precision of subsurface models and predictions.

In conclusion, well logs and geological information are indispensable tools for understanding the subsurface. geology and identifying potential hydrocarbon reservoirs in regions like Kansas City. (GEOPROVIDER, 2023)

Tableau I: Classification codes for LAS formats 1.1 through 1.4

Value	**Meaning**
0	Created, never classified
1	Unassigned
2	Ground
3	Low Vegetation
4	Medium Vegetation
5	High Vegetation
6	Building
7	Low Point
8	Model Key-Point
9	Water
10	Rail
11	Road Surface
12	Reserved
13	Wire - Guard (Shield)
14	Wire - Conductor (Phase)
15	Transmission Tower
16	Wire-Structure Connector
17	Bridge Deck
18	High Noise
19	Reserved
20	Ignored Ground
21	Snow
22	Temporal Exclusion
23-63	reserved
64-255	User Definable

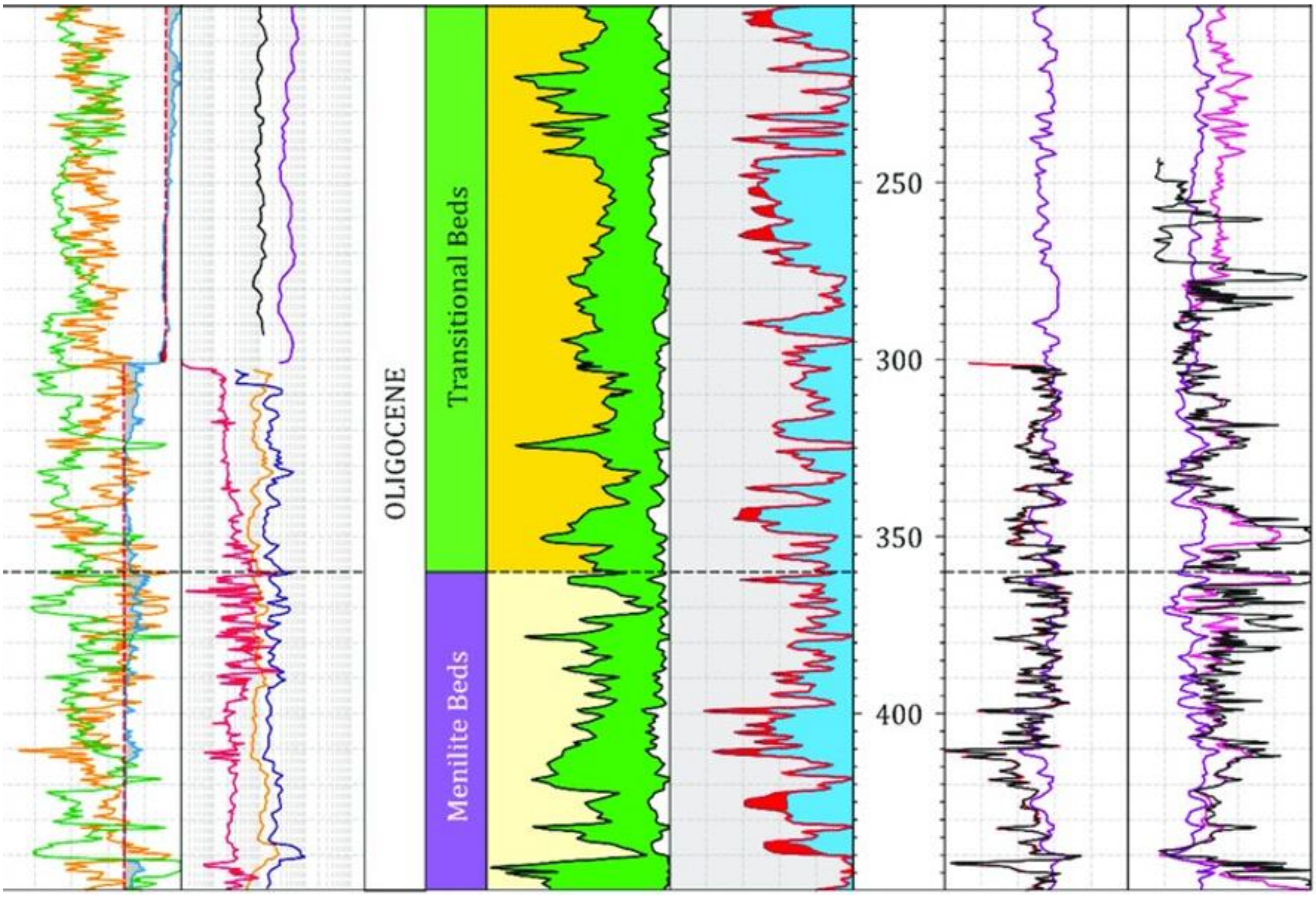


Figure 7:diagram of Well logging data in the uppermost depth section of the D-1 borehole

3.3.2 LAS File & Seismic Data: Exploring Deeper

In addition to well logs and geologic information, our exploration strategy incorporates LiDAR (Light Detection and Ranging) data stored in LAS format and seismic data. we aim to develop a comprehensive understanding of the subsurface structure within our target area (Kansas City, USA). This comprehensive approach will provide valuable input for our QML model, allowing it to analyze a broader range of features associated with hydrocarbon exploration.

Originally, lidar data was only delivered in ASCII format. With the massive size of lidar data collections, a binary format called LAS was soon adopted to manage and standardize the way in which lidar data was organized and disseminated. Now lidar data is commonly represented in LAS. LAS is a more acceptable file format, because LAS files contain more information and, being binary, can be read by the importer more efficiently.

LiDAR (LAS File):

LiDAR technology employs light pulses emitted from a laser sensor to accurately measure distances to the Earth's surface. These pulses generate high-resolution 3D models known as LAS files. This data is valuable for several reasons:

Identifying surface features potentially indicative of subsurface structures: LiDAR data can reveal subtle variations in terrain elevation, allowing geoscientists to identify surface expressions of underlying geological features such as faults, folds, and other structural complexities.

Generating detailed topographic maps: LiDAR captures precise elevation data, enabling the creation of highly detailed topographic maps that are essential for understanding surface morphology and landscape evolution.

Seismic Data:

Seismic data acquisition involves generating controlled sound waves (usually using specialized equipment such as seismic vibrators or explosives) and recording their reflections from subsurface rock layers. Geophysicists analyze these reflections to construct detailed images of the subsurface. Seismic data can reveal:

Faults and folds in rock formations: By analyzing the patterns of seismic reflections, geoscientists can identify faults—fractures in the Earth's crust where movement has occurred—and folds—bends or wrinkles in rock layers caused by tectonic forces.

Potential hydrocarbon reservoirs: Seismic surveys are widely used in oil and gas exploration to identify underground structures that may contain hydrocarbon reservoirs. Specific seismic reflections can indicate the presence of porous rock formations that may trap oil or gas.

3.3.3 magnetic data

Magnetic data, acquired through geophysical techniques like aeromagnetic surveys, can also be integrated into our analysis. Aeromagnetic surveys involve flying low-altitude aircraft equipped with magnetometers to measure variations in the Earth's magnetic field caused by subsurface rocks. Geologic structures like faults and ore bodies containing iron-rich minerals can cause disruptions in the magnetic field, creating measurable anomalies. By processing and analyzing this magnetic data, geophysicists can create digital aeromagnetic maps that reveal these anomalies.

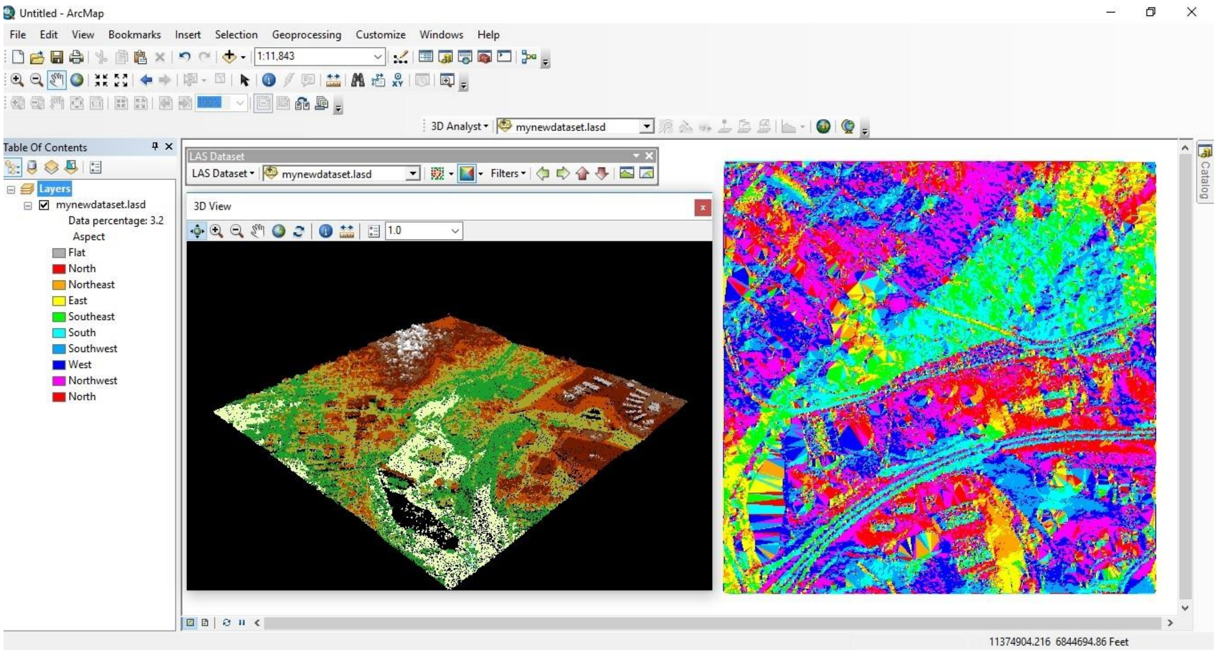


Figure 8:displaying magnetic data and lidar in ArcGIS

3.3.4 landsat9

Given its huge data archive and open access, Landsat 9 is the best option. As previously stated, the Landsat program provided by the USGS is an invaluable resource for remote sensing data collecting. We will investigate the available Landsat imagery for our target location (Kensas), focusing on spectral bands that can successfully detect hydrocarbons. This selection will take into account the trade-off between spatial resolution (30 meters) and data volume. The downloaded Landsat data, in a format compatible with our GIS platform, will be a critical component of our AI/QML model building efforts.

3.4 preparing the canvas

This section will be broken into two primary portions, with a focus on data processing and feature engineering. Because we didn't have any initial clean data to work with, we pooled all of our resources to construct one just for this project.

The collection of the various data took a long time, and for this kind of work, Google Search Engine was not enough. We needed to take the search to the next level by launching a new Web 2.0 search engine called Explorer.

https://explorer.globe.engineer

Given the easy access we have to a large amount of data, we were able to access the University of Kansas, which was the biggest source of our data, from geologic maps to well logs. Using a technique called Google Dorking, we were able to gather most of the information we needed. This technique is to use specific key words to narrow down the research to make it more specific about our interest.

3.4.1 Preprocessing Landsat Images:

now the Preprocessing Landsat 9 imagery with ArcGIS typically involves several steps to correct atmospheric, geometric, and radiometric distortions. Here's a brief description of each step:

- Enhancing the resolution:

Landsat 9 imagery can be improved from its 30-meter resolution to 15 meters using image fusion or pan-sharpening. This process combines the higher-resolution panchromatic band with lower-resolution multispectral bands, creating a single high-resolution color image. ArcGIS offers pan sharpening tools using algorithms like Bovey, Gram-Schmidt, or Principal Component Analysis, resulting in a higher-resolution image that retains both panchromatic and multispectral band details.

- Geometric Correction:

Landsat imagery often faces geometric distortions due to terrain variations and satellite sensor characteristics. Geometric correction corrects these distortions using ground control points (GCPs) and resampling techniques, aligning the image with a map projection system and registering it to a known coordinate system.

- Atmospheric Correction:

Atmospheric correction is influential for removing the effects of atmospheric scattering and absorption, which can distort the spectral characteristics of the image. This correction typically involves applying models to estimate and remove atmospheric effects, such as Rayleigh scattering, aerosol scattering, and water vapor absorption This procedure also involves the conversion of raw spectral band measurements into reflectance values, facilitating more meaningful quantitative analyses and interpretations of the observed environmental phenomena.

REFLECTANCE_MULT_BAND_1 = 2.0000E-05

REFLECTANCE_ADD_BAND_1 = -0.100000

then we multiply the reflectance mult band by the band2 and divide it by the sin of the sun elevation in radiance unit. and for this we used ArcMap for more efficiency

- Mosaicking:

Mosaicking is a technique used in Landsat imagery to create a seamless composite image by blending overlapping scenes, while subset and masking extract specific regions for analysis.

- Band compositing:

is a process that combines individual bands from satellite imagery to create a multi-band composite image. This enhances the interpretability of satellite data by highlighting specific spectral bands or features. The initial step is mosaicking, followed by subset and masking techniques to isolate specific regions of interest.

- Enhancement and visualization:

methods like contrast stretching, color compositing, and histogram equalization are used to improve clarity and highlight key features. These preprocessing steps ensure the composite image is accurate, calibrated, and ready for further analysis and interpretation within GIS platforms like ArcGIS. (Anderson, 2024)

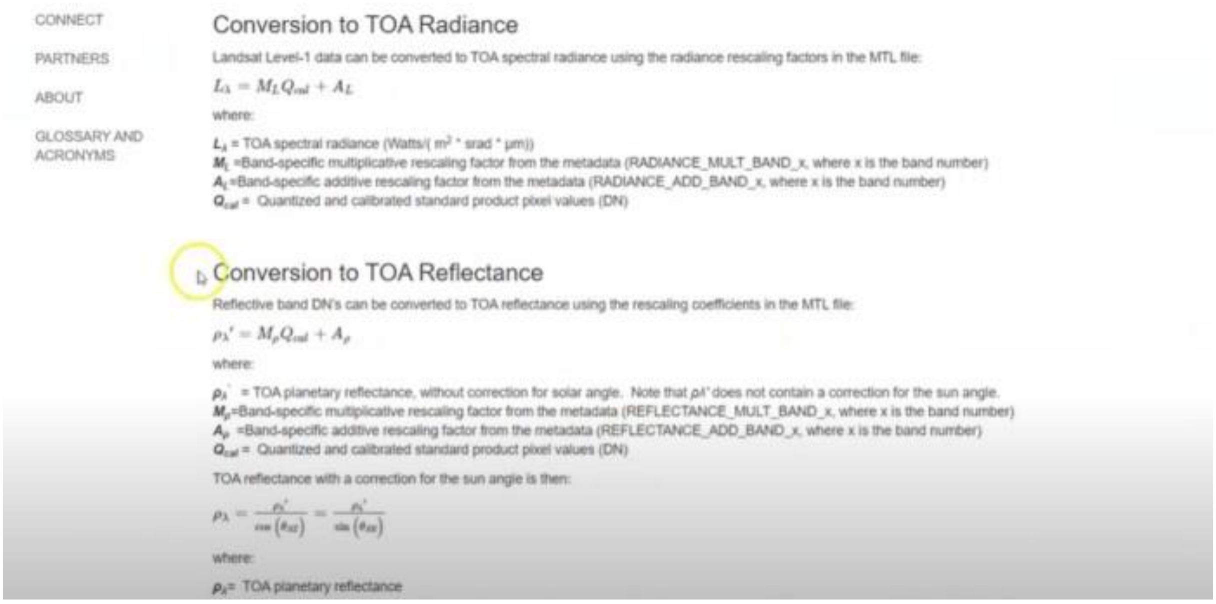


Figure 9:USGS figure of toa values and conversion

3.4.2 feature engineering: beyond preprocessing:

After atmospheric correction and preprocessing of remote sensing data, several advanced analysis techniques can be applied to derive valuable information for various applications.

Feature engineering creates new spectral indices or bands to capture specific information for study objectives, such as land cover discrimination, soil moisture estimation, or urban heat island detection.

- **Lineament Analysis:**

Lineament analysis is like giving the Earth's surface a closer look and spotting its linear features like faults, fractures, and boundaries, all with the help of remote sensing data. Think of it as using special glasses to see through the layers of the earth! We've got different tools for this job, from our keen eyes for visual interpretation to fancy algorithms that pick out edges and nifty spatial analysis tools. These tools help us create lineament maps, which are like treasure maps, revealing secrets about the Earth's makeup, its movements, and even where to dig for precious minerals or steer clear of natural hazards. We took these tools and applied them to our layer stack, diving deep into the Earth's secrets. But we didn't stop there! We also tapped into magnetic data, which can be like having a magnetic compass that points out hidden movements and faults beneath the surface. Imagine it as a sort of geological detective work! Aeromagnetic surveys play a big role here, sniffing out changes in the Earth's magnetic field caused by different rocks and structures hiding beneath the surface. Geologists love this stuff, especially when hunting for oil and gas, as it helps them pinpoint exactly where to dig. But we didn't just rely on fancy tools; sometimes, a basic Python script was all we needed to dive even deeper into the magnetic data, uncovering the area's geological faults like hidden treasure. It's all about getting to know the Earth a little better, one fault line at a time.

```
import cv2
import numpy as np
import pandas as pd
import hemi_fault_detection as hf
img = cv2.imread('ln_og.png')
from IPython.display import Image
for points in lines:
    x1,y1,x2,y2 = points[0]
    cv2.line(img, (x1,y1),(x2,y2),(0,0,0))
cv2.imwrite('faults.png', img)
edges = filters.sobel(ndvi_data)
binary_edges = edges > filters.threshold_otsu(edges[~np.isnan(edges)])
plt.imshow(binary_edges, cmap='viridis')
plt.title('Detected Edges (Structural Anomalies)')
plt.colorbar(label='Edge Strength')
plt.show()
```

The code snippet detects faults and edges in geological images.
It manipulates images and processes data using libraries like OpenCV (cv2), NumPy (np), and Pandas (pd).
The 'hemi_fault_detection',
The generated image with fault lines is stored as 'faults.png'. Edge identification is then carried out using filters from the'sobel' module to reveal structural irregularities.
The discovered edges are translated to binary representation for better visibility.
Matplotlib is used to visualize the edges, which are displayed as a heatmap with different colors representing edge strength. This code snippet helps to locate faults and structural anomalies in geological imaging, allowing for additional analysis and interpretation. As the plots below show, we obtained good results using the script that depicts the composition of the earth in the region of interest.

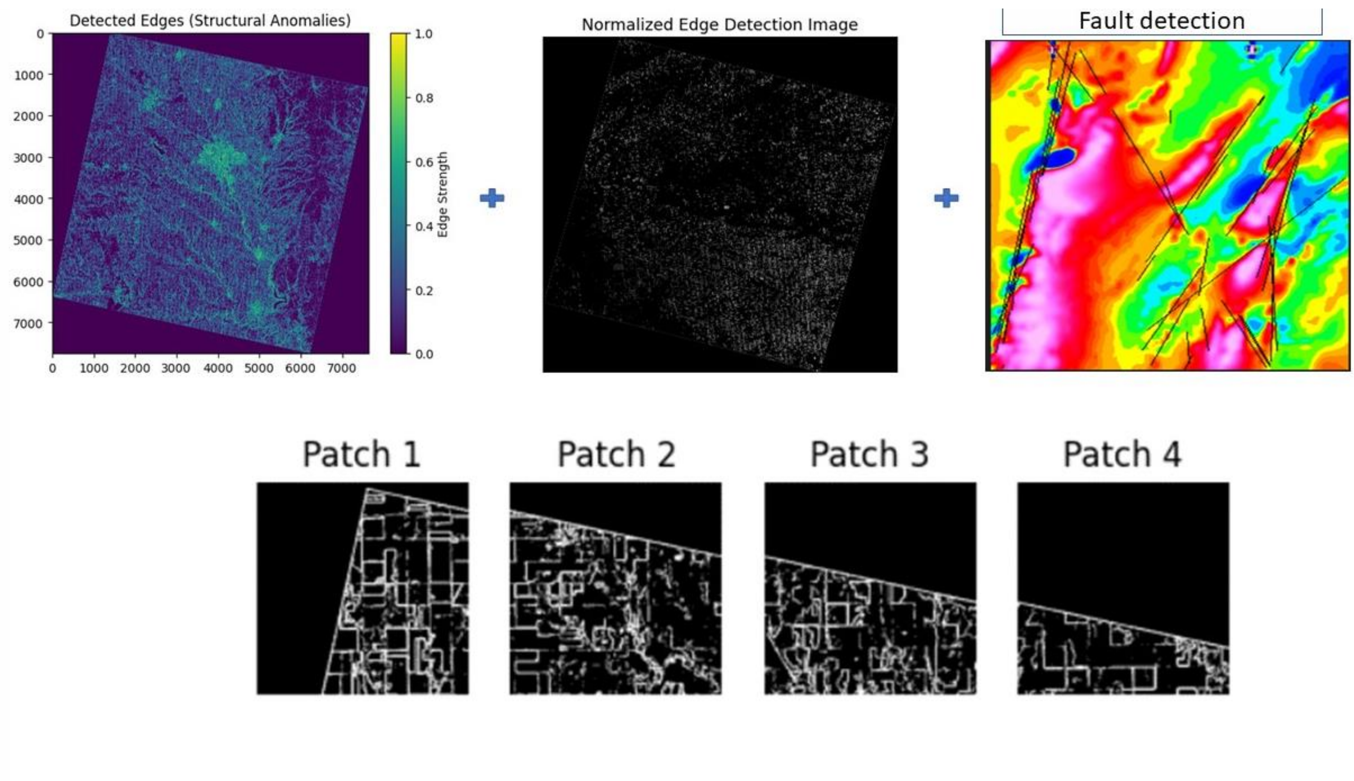


Figure 10: plot from the code represent the fault detections and geological formations

- **Calculating Indices like NDVI and Mineral Indices:**

NDVI is a widely used index calculated from near-infrared (NIR) and red bands, given by (NIR - Red) / (NIR + Red). It provides information about vegetation density, health, and coverage. Other indices include Normalized Difference Water Index (NDWI), Soil Adjusted Vegetation Index (SAVI), Enhanced Vegetation Index (EVI), and many more, each designed to capture specific vegetation or environmental characteristics.

Mineral indices are calculated to identify and map mineral composition or alterations in geological formations. These indices are based on the unique spectral signatures of minerals in certain wavelength regions.

Examples of mineral indices include the Normalized Difference Iron Oxide Index (NIR1 - SWIR1) / (NIR1 + SWIR1), which is sensitive to iron-bearing minerals, and the Clay Index, which helps in mapping clay minerals.

Now, for better accuracy and to detect any mistakes, we performed the NDVI calculation using two methods: one with ArcMap and the other with a Python script using the Rasterio library.

Both calculations were done before and after the atmospheric correction, and the result was very satisfying in the range between -1 and 1.

By applying these advanced analysis techniques, researchers can extract valuable information from remote sensing data for various applications such as land use and land cover mapping, vegetation monitoring, hydrological studies, geological mapping, environmental monitoring, and natural resource management. These analyses contribute to a better understanding of Earth's surface dynamics and facilitate informed decision-making processes.

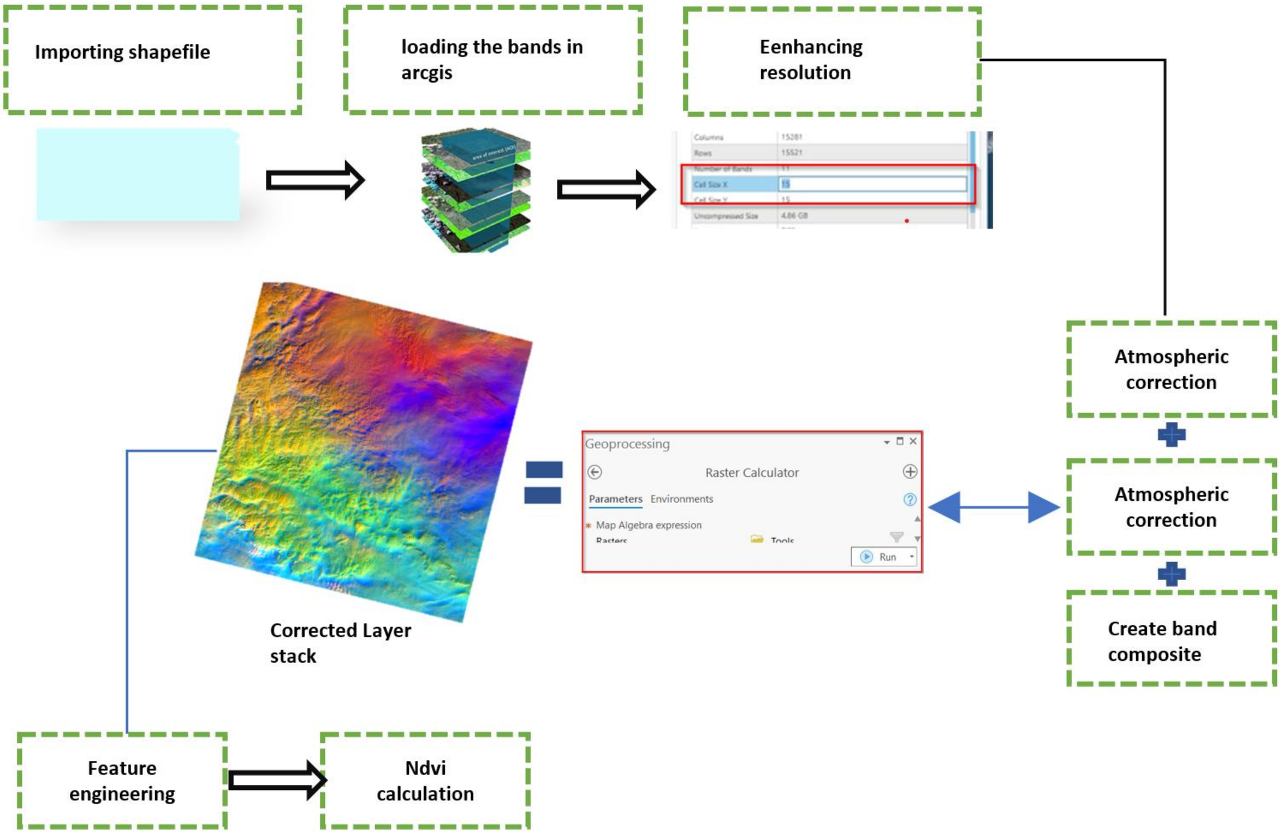


Figure 11: diagram 1 explaining the workflow

A snip of the code

```
red_band_path = '/content/drive/MyDrive/f12/bd12radiom4.tif'
nir_band_path = '/content/drive/MyDrive/f12/bd12radiom5.tif'
red_band = load_raster_from_drive(red_band_path)
nir_band = load_raster_from_drive(nir_band_path)
red_data = red_band.read(1).astype(float)
nir_data = nir_band.read(1).astype(float)
ndvi_data = (nir_data - red_data) / (nir_data + red_data)
print("NDVI Range:")
print("Min:", np.min(ndvi_data))
print("Max:", np.max(ndvi_data))
ndvi_data_rescaled = (ndvi_data - np.min(ndvi_data)) / (np.max(ndvi_data) - np.min(ndvi_data))
ndvi_data_rescaled[np.isnan(ndvi_data_rescaled)] = 0
plt.figure(figsize=(8, 6))
plt.imshow(ndvi_data_rescaled, cmap='viridis')
plt.colorbar(label='NDVI')
plt.title('Rescaled NDVI Image')
plt.show()
bands = [load_raster_from_drive(path) for path in band_paths]
spectral_signatures = extract_spectral_signatures(bands, ndvi_data, representative_areas)

bands_labels = ['Band 1', 'Band 2', 'Band 3', 'Band 4', 'Band 5', 'Band 6',
                'Band 7', 'Band 8', 'Band 9', 'Band 10', 'Band 11', 'NDVI']
plt.ylim(0, 1)
plt.xlabel('Landsat Bands / NDVI')
plt.ylabel('Normalized Reflectance')
plt.title('Normalized Spectral Signatures of the Entire Region')
plt.legend()
plt.grid(True)
plt.show()
```

the code performs spectral analysis on raster data, including NDVI calculation and extraction of spectral signatures, providing valuable insights into the characteristics of the studied area

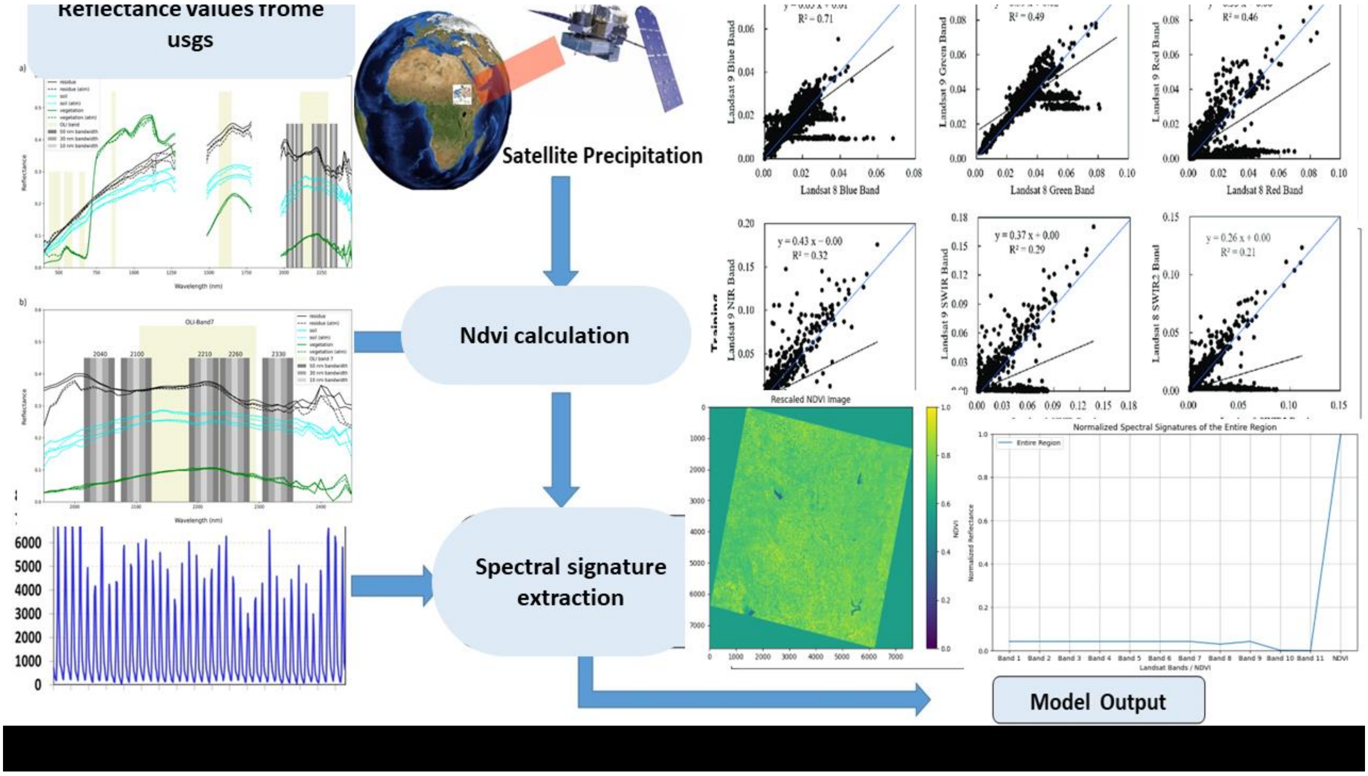


Figure 12: graphical abstract 3

- **well logs and LAS file:**

Navigating through the project, we encountered a pivotal piece of data: well logs and LAS files. These components require a deep understanding of petroleum engineering and geophysics, making their acquisition and handling both challenging and exciting. Our journey led us to the University of Kansas website, where we secured the well log and LAS files. However, they arrived in raw format, necessitating thorough cleaning and preparation before integrating them into our algorithm. Gratefully, Python libraries came to our rescue, streamlining the process with just a few lines of code, yielding tidy results.

Well logs serve as treasure troves of information about subsurface formations encountered during drilling, demanding expertise in petroleum engineering and geophysics for interpretation. Meanwhile, LAS files, born from LiDAR surveys, furnish detailed 3D representations of surface topography. Our data comprised three main elements. The first, the LAS file in netcdf format, adheres to industry-standard file protocols prevalent in oil-and-gas and water well sectors. Each LAS file encapsulates data for a single well, with the capacity for numerous datasets, or "curves." Accessing this file type requires specialized software, such as Magmap, while leveraging Armap's capabilities facilitates data conversion into suitable formats, including CSV.

Next in our arsenal was the well log file, a rich repository of subsurface formation data. These logs, culled during the drilling process, encompass a gamut of measurements and observations. From lithology to porosity, permeability to fluid saturation, well logs offer comprehensive insights into penetrated formations.

Gamma Ray Log: Measures natural radioactivity, aiding in lithology identification and correlation.

Spontaneous Potential (SP) Log: Records natural electrical potential differences, offering insights into fluid content and formation boundaries.

Resistivity Logs: Provide data on formations' electrical resistivity, aiding in fluid saturation assessment and hydrocarbon zone identification.

Density Log: Measures bulk density for porosity determination and lithology identification.

Neutron Porosity Log: Gauges hydrogen content for porosity determination and fluid identification.

Sonic Log: Records sound wave travel times, facilitating porosity determination, rock mechanical properties assessment, and depth correlation.

Caliper Log: Measures wellbore diameter, assisting in hole condition evaluation and well integrity assessment.

Finally, we simplified yet potent steps. We harnessed the geologic map of Kansas City, extracting invaluable data from its legend and translating it into a CSV table. This step was instrumental in augmenting our dataset with geographic insights crucial for our analyses.

First, we needed to prioritize transforming all the files into a format that Python could understand, namely a CSV file. Then, we combined and merged the files into one. Once our file was ready, we delved deep into the data, understanding the concepts within to discern what was essential for our project.

When it was time for the real work to begin, breaking it down into small tasks was our strategy. In this paragraph, we'll outline these steps, explaining how we arrived at the dataset used later in the project.

3.4.3 Breakdown of the Data Preprocessing Code for Well Log Data:

Importing Libraries and Loading Data: We imported necessary libraries for data manipulation (Pandas), visualization (Matplotlib), and scaling (Scikit-learn) (NumPy, Matplotlib.pyplot, Pandas, Scikit-learn, Os, IO).

Data Cleaning: Handling Missing Values: We first removed leading/trailing spaces from column names and checked for missing values, printing columns with significant data absence. We identified numerical and categorical columns, separating them based on data types (Integer, Float, or Object).

Targeting Missing Values in Different Data Types: We addressed missing values specifically in numerical and categorical columns using list comprehensions. For numerical columns, we filled missing entries with the mean value of each column, excluding missing values, while categorical columns were filled with the most frequent value (mode) within each column.

Handling Duplicate Rows: We created a new Data Frame as a copy of the updated data, identified and removed duplicate rows, and saved the data without duplicates as a new CSV file.

Data Normalization: Another copy of the data was created for normalization. We specified columns for normalization, created a Min Max Scaler object, applied it to the specified columns, and saved the normalized data as a new CSV file.

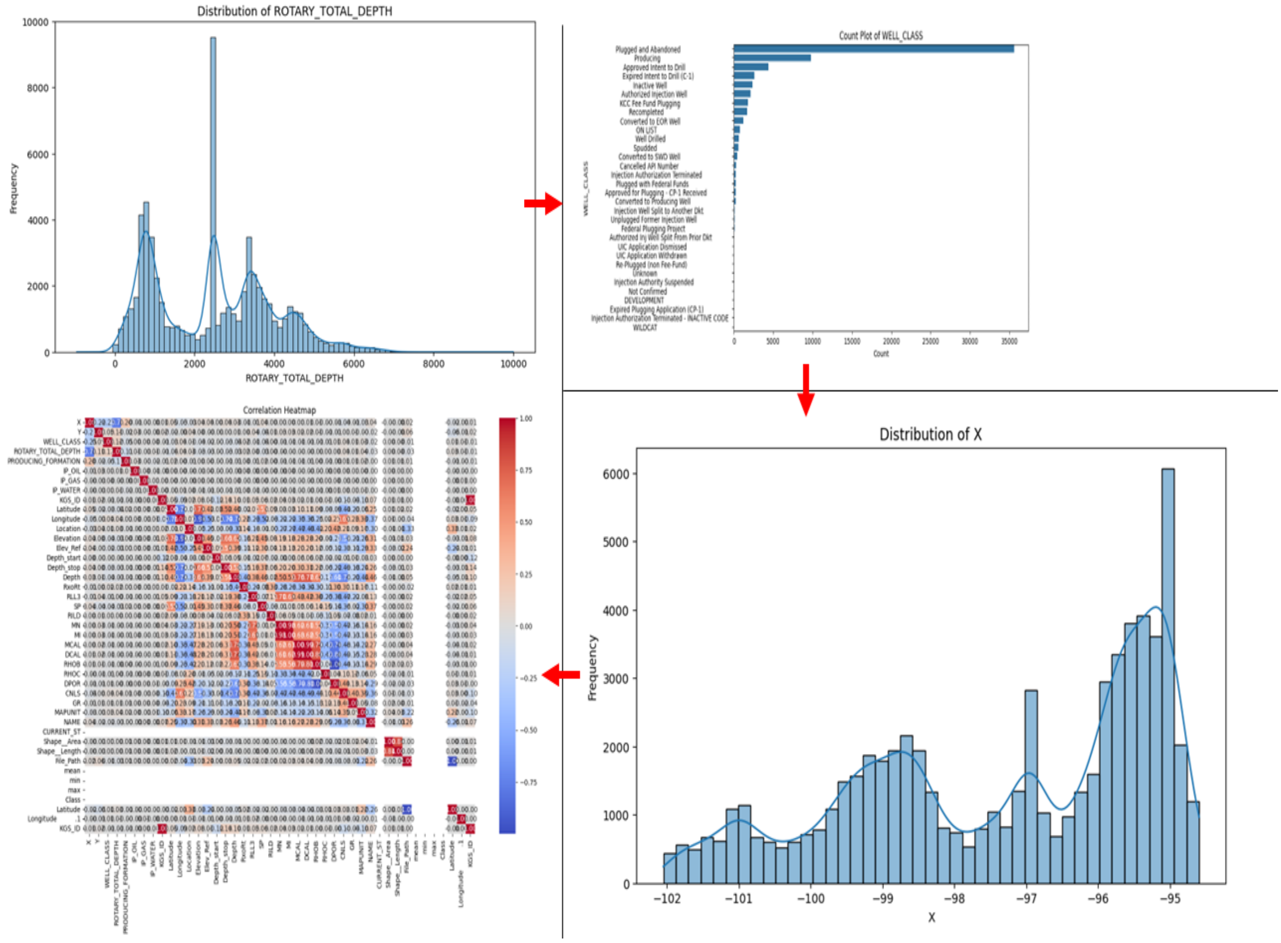


Figure 13: diagram 2 plot of the data distribution

Conclusion

Our data preprocessing workflow effectively handled missing values, identified data types, removed duplicates, and normalized specific columns. This preparation was crucial for the machine learning model. We integrated well logs, LAS files, remote sensing data, and geologic information into a single dataset, addressing inconsistencies and performing feature engineering to extract relevant geological features.

Chapter 4 : Conducting the AI/QML Symphony in Resource Exploration

4 Chapter: Conducting the AI/QML Symphony in Resource Exploration

4.1 prelude:

Quantum Machine Learning (QML) is a transformational strategy that combines quantum computing and machine learning approaches. QML algorithms, which use quantum mechanics concepts, have the ability to tackle difficult problems more efficiently than classical methods. In this chapter, we look at how QML can be used to analyze and understand geological data, namely well logs, LAS files, and remote sensing data.

4.2 concept

In the current age of the Fourth Industrial Revolution, the digital world has a wealth of data, such as Internet of Things data, cybersecurity data, health data, etc. To intelligently analyze these data and develop the

corresponding *smart and automated* applications, knowledge of artificial intelligence (AI), particularly *machine learning (ML),* is the key. Various types of machine learning algorithms, such as supervised, unsupervised, semi supervised, and reinforcement learning, exist in the area. Besides, *deep learning*, which is part of a broader family of machine learning methods, can intelligently analyze the data on a large scale. In this chapter, we present a comprehensive view of these algorithms that can be applied to enhance the intelligence and capabilities of an application. Thus, this section's key contribution is explaining the principles of different techniques and their applicability in various real-world application domains. We also highlight the challenges and potential research directions based on our study. Overall, these nexus lines aim to serve as a reference point. in the context of data analysis and computing that typically allow the applications to function in an intelligent manner. ML usually provides systems with the ability to learn and enhance from experience automatically without being specifically programmed and is generally referred to as the most popular and latest technology in the fourth industrial revolution. The learning algorithms can be categorized into four major types, such as supervised, unsupervised, semi-supervised, and reinforcement learning.

4.2.1 Types of Real-World Data

Structured Data: It has a well-defined structure, conforms to a data model, and is highly organized, making it easily accessed and used by entities or computer programs.

Unstructured Data: This type lacks a pre-defined format or organization, making it more difficult to capture, process, and analyze. It mostly contains text and multimedia material, such as sensor data.

Semi-structured Data: Not stored in a relational database but possesses some organizational properties that make it easier to analyze, such as HTML, XML, JSON documents, and NoSQL databases.

Metadata: Data about data, providing additional information to give more significance to data users. For example, a document's metadata might include an MTL file with the Landsat bands containing the sun's elevation and reflectance values.

4.2.2 Machine Learning Techniques

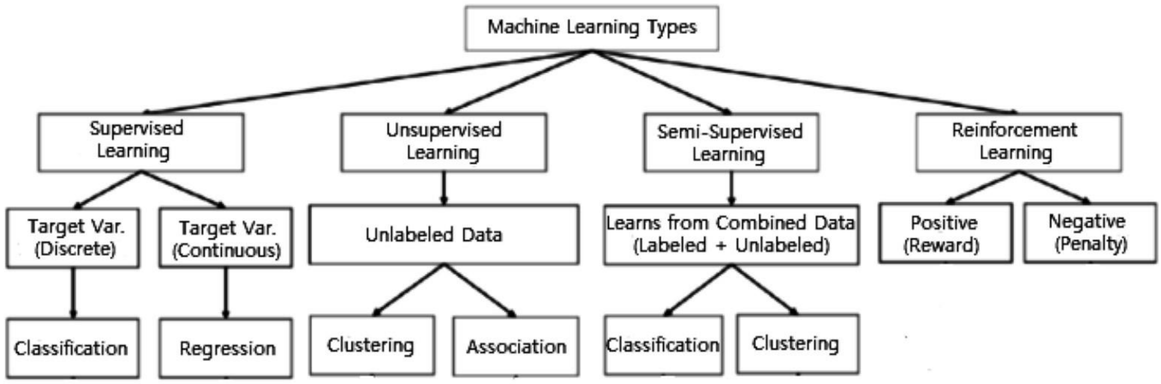


Figure 14: Various types of machine learning techniques

Supervised learning: involves machine learning tasks where a function is learned to map inputs to outputs based on labeled input-output pairs. It utilizes labeled training data and aims to infer a function using a collection of training examples. Supervised learning is task-driven, with common tasks including classification to separate data and regression to fit the data. An example of supervised learning is text classification, where the goal is to predict the class label or sentiment of a piece of text, such as a tweet or a product review.

Unsupervised learning: involves analyzing unlabeled datasets without human interference. It is used to extract generative features, identify trends and structures, group results, and for exploratory purposes. Common tasks in unsupervised learning include clustering, density estimation, feature learning, dimensionality reduction, finding association rules, and anomaly detection.

Semi-supervised: learning combines supervised and unsupervised methods by operating on both labeled and unlabeled data. It is useful in contexts where labeled data is rare and unlabeled data are abundant. The goal of semi-supervised learning is to improve prediction outcomes compared to using labeled data alone.

Applications of semi-supervised learning include machine translation, fraud detection, labeling data, and text classification.

Reinforcement learning: is a type of machine learning algorithm that enables software agents and machines to automatically evaluate optimal behavior in a specific context or environment to improve efficiency. This learning method is environment-driven and relies on rewards or penalties. Its goal is to use insights from the environment to maximize rewards or minimize risks. Reinforcement learning is useful for training AI models to increase automation or optimize the operational efficiency of complex systems such as robotics, autonomous driving, manufacturing, and supply chain logistics. However, it is not preferred for solving basic or straightforward problems.

Many classification algorithms have been proposed in the machine learning and data science literature. In the following, we summarize the most common and popular methods that are widely used in various application areas.

4.2.3 Machine Learning algorithms

Linear Discriminant Analysis (LDA)

LDA is a linear decision boundary classifier that fits class-conditional densities to data and applies Bayes' rule. It reduces dataset dimensionality to minimize model complexity and computational costs, assuming Gaussian density for each class and a shared covariance matrix. (Sarker, 2021)

Logistic Regression (LR)

LR is a probabilistic statistical model for classification, using a logistic function to estimate probabilities. Regularization techniques like L1 and L2 can prevent overfitting, though the assumption of linearity between dependent and independent variables can be a drawback.

K-Nearest Neighbors (KNN)

KNN is an instance-based learning algorithm that stores all training data instances in n-dimensional space and classifies new data points based on similarity measures like the Euclidean distance function.

Decision Tree (DT)

DT is a non-parametric supervised learning method used for both classification and regression tasks. It classifies instances by sorting them down the tree from root to leaf nodes, with common splitting criteria being "gini" for Gini impurity and "entropy" for information gain.

4.3 Deep Learning in Resource Exploration

While machine learning focuses on algorithms and models to parse data, learn from it, and make informed decisions, deep learning takes this a step further by mimicking the human brain to process data and create patterns for use in decision making. This approach, inspired by the structure and function of the brain's neural networks, has proven to be particularly effective in tasks such as image and speech recognition, natural language processing, and more. The shift to deep learning opens up a world of possibilities for solving complex problems and unlocking new frontiers in artificial intelligence.

This new technology can actually be categorized into two main sections: artificial neural networks (ANN) and convolutional neural networks (CNN).

4.3.1 Artificial Neural Networks (ANN)

Well, we can find a hundred definitions for neural networks on Google, but by a simple definition, I would like to call it a simulation of the human brain mechanism. Our brains identify things through certain patterns. And according to these patterns, the brain activates certain neurons to trigger specific areas of the brain, like releasing cortisol and adrenaline, which alerts the nervous system. stimulated by the amygdala. the same principles applied to computers: by feeding it a certain input, neural networks process data by recognizing patterns and adjusting the connections between artificial neurons (nodes) targeted neurons get activated, doing the calculations of the weight and the bios to get the wanted output.

4.3.2 Architecture

The neural network architecture consists of interconnected layers of neurons, categorized into input, hidden, and output layers. Data is received by the input layer and processed through hidden layers via weighted connections and activation functions. The output layer generates the network's final predictions or classifications. Each connection between neurons has an associated weight, which is modified during training to enhance the network's predictive abilities. The network learns by iteratively adjusting these weights based on prediction errors, facilitated through backpropagation. This brain-inspired architecture is used in various tasks, including image and speech recognition and natural language processing. The flow of information occurs in two ways: forward networks, which have an input layer and a single output layer with zero or multiple hidden layers, and feedback networks, which use their internal state to process input sequences.

4.3.3 Components

A neural network is a system where neurons, or activation functions, receive input from external sources and establish connections with nodes in subsequent layers. The input layer's node values are combined to create a hidden layer with a pre-defined activation function. This layer processes inputs to generate results forwarded to the output layer for user accessibility. For example, an image of an apple with an output value ranging from 0 to 1 can be identified as an orange if the value falls below 0.5, and an apple if it surpasses 0.5. The network then discerns patterns based on shape rather than color, a process called forward propagation.

4.3.4 algorithms

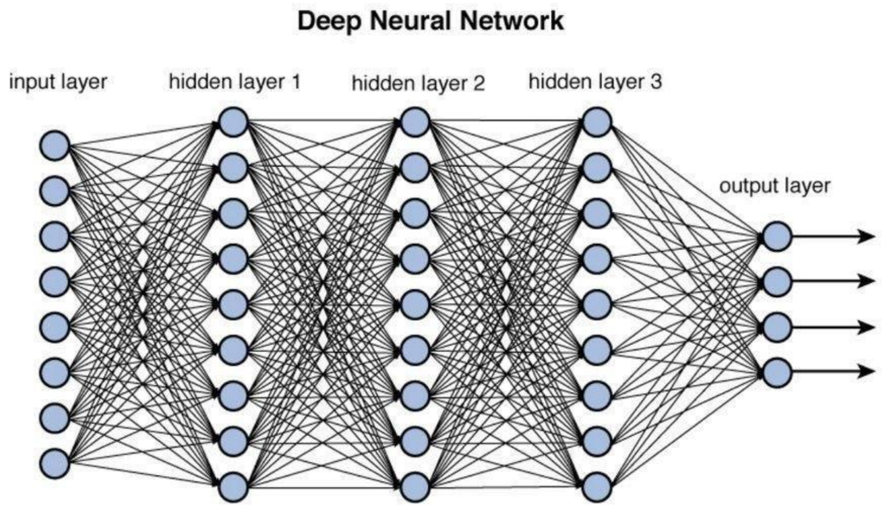


Figure 15: deep neural network architecture

The learning process in machine learning involves dividing data into three sets: training, validation, and testing. Optimization algorithms are applied to minimize errors and adjust network parameters. The learning problem is minimizing a loss function (f) that evaluates performance on a dataset, consisting of an error term and a regularization term. The loss function depends on adaptive parameters, represented as a single n-dimensional weight vector.

The learning problem involves a loss function that depends on multiple parameters, and five main algorithms are discussed.

One-dimensional optimization is used to minimize the loss in the training direction.

The Golden Section Method is used to find the minimum or maximum of a single-variable function.

Brent's Method is a root-finding algorithm that combines root bracketing, bisection, secant, and inverse quadratic interpolation.

Multidimensional optimization involves searching through the parameter space and adjusting the parameters at each step.

Three examples of multidimensional optimization algorithms include gradient descent, Newton's method, and gradient descent.

One-dimensional optimization uses a training direction and a training rate to minimize the loss.

The gradient descent algorithm starts at w(0) and progresses iteratively until a specified criterion is met.

It updates the weight w(i) by moving in the opposite direction of the gradient vector multiplied by a learning rate η(i).

Newton's method is a second-order algorithm that incorporates second derivatives of the loss function. Iterating as $w(i+1) = w(i)−H(i)^{−1}·g(i)$ is a common method for training. (Gurucharan, 2022)

4.3.5 neural network applications in oil and gas

The study uses a neural network to estimate ultimate recovery in shale reservoirs and develop PVT (pressure, volume, temperature) correlations for crude oils. The calculations are based on phase behavior and PVT data, including in-place estimation, pressure loss, fluid flow, rate transient analysis, and well testing. Lab testing is the most common method, but complex, timely, and expensive. Statistical regression techniques are used to derive empirical correlations for different PVT properties, with computers excelling in these calculations.

4.4 Convolutional neural network (CNN)

CNNs were created to recognize handwritten digits. CNNs decompose images into three scales, from low to high order. Low-order details or local aspects of the neural network's early layers can extract image-like lines, edges, and curves. Deeper layers would then integrate the features before reconstructing the entire image. CNN training for image classification involves multiple processes, including convolution, activation function, pooling, and fully linked layers.

The CNN architecture is mainly 2 parts

A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.

A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.

- Convolutional Layer:

Separates the image into pixels, with each pixel valued based on form and color. Applies convolution between the input image and an MxM filter to recognize local characteristics. Uses filters or kernel functions to focus on specific features of the image, detecting patterns by returning high positive numbers when a pattern is found.

- Activation Function:

Most common in CNNs is the rectified linear unit (ReLU), which sets negative values to zero, applied after the convolution process to each feature element.

- Pooling Layer:

Added after a convolutional layer to reduce the size of the convolved feature map and computational expenses.

- Types of pooling:

Max Pooling: Selects the largest element from the feature map.

Average Pooling: Computes the average of the elements in a specified area.

Sum Pooling: Calculates the sum of the elements in a specified area.

- Fully Connected Layer:

Flattens the pooled feature maps into a one-dimensional matrix for the input layer of a deep neural network. Each input is multiplied by its own weight and combined with a bias, typically placed before the output layer and comprising the final few layers of a CNN architecture.

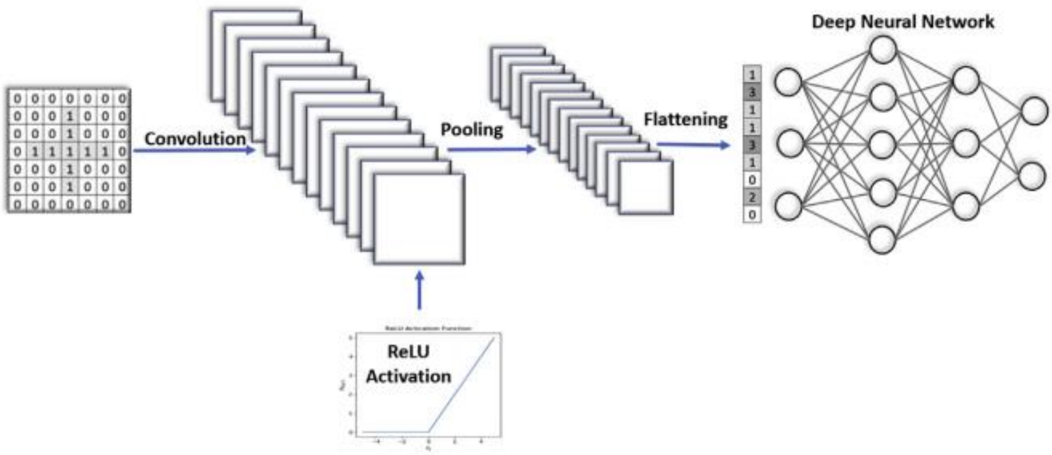


Figure 16: the whole workflow of CNN convolution for image classification

To summarize, functions are input-output machines.

They take an input set of numbers and output a corresponding set of numbers, and the functions define the relationship between these two the particular problem that neural networks solve is when we don't know the definition of the function that we are trying to approximate. Instead, we have a simple data point from which we must approximate a function. that fits the data point and allows us to accurately predict outputs given inputs that are not in the dataset. This process is called curve fitting.

4.5 Quantum circuit architecture and algorithm

Quantum Neural Networks (QNNs) are advanced technologies that combine quantum computing and artificial intelligence. They use quantum physics concepts to conduct complex computations efficiently and scalable. QNNs manipulate quantum bits, enabling them to investigate large solution spaces and solve optimization issues. This article delves into the core ideas of QNNs, their representation as quantum circuits, and their distinct advantages over traditional neural networks. It also looks into their possible uses in optimization, machine learning,

4.5.1 Qubits and quantum circuits

QNNs are wrapped inside the beautiful framework of quantum circuits, which coordinate computational tasks through the delicate dance of quantum bits (qubits) and quantum gates. In this section, we explain the essence of QNNs by depicting them as quantum circuits, giving light on the fundamental features of qubits and quantum gates, which serve as the networks' building blocks. At the heart of QNNs are qubits, the quantum equivalents of classical bits.

Qubits, unlike classical bits, can exist in both 0 and 1. qubits possess the property of superposition, allowing them to exist in a simultaneous blend of both states. This quantum superposition gives qubits a powerful ability to encode and interpret information in a way that is fundamentally different from classical bits.

Quantum gates, which act similarly to conventional logic gates but in the quantum environment, are critical to the functionality of quantum circuits. These gates act as the fundamental operations that manipulate qubits, organizing the complex dance of quantum states to complete computational tasks. QNN circuits use a diversified set of quantum gates to shape the flow of quantum information and perform transformative operations on qubits. Among the fundamental quantum gates used in QNN circuits are:

Hadamard Gate (H): This gate causes superposition, allowing qubits to exist in a probabilistic combination of 0 and 1 states.

The CNOT Gate (Controlled-NOT) acts on two qubits to entangle their states, enhancing quantum entanglement and allowing the formation of complex quantum states.

Pauli Gates (X, Y, and Z): These gates represent rotations within the Bloch sphere, which affect the phase and amplitude of qubit quantum states.

Together, these quantum gates form the foundation of QNN circuits, managing the complicated quantum operations that enable the processing ability of quantum.

4.5.2 Quantum encoding

Quantum data encoding is a technique used to convert classical data into quantum states.

Two commonly used encoding methods are amplitude encoding and basis encoding.

- Amplitude encoding

Basis encoding is a straightforward technique where classical data is represented in a manner similar to digital computing. For instance, if we have the number 3, it would be expressed in binary as 11, where each digit corresponds to a bit: 0 or 1. In quantum terms, this binary representation is converted into a quantum state, denoted as |11>, utilizing two qubits. Each qubit is then set to represent one of the binary digits, with both qubits initialized to 1. For larger datasets or vectors, the same process is repeated, with each component encoded using qubits, and the resulting quantum states concatenated. While basis encoding offers simplicity, as most computers start with qubits initialized to 0, it comes with the drawback of inefficient qubit utilization. Additionally, this method is versatile and can handle non-numerical data, expanding its applicability across various types of datasets. Each vector encoded using basis encoding requires a substantial number of qubits, for example, four qubits for a single vector, leading to potential wastage of quantum resources.

- Amplitude encoding:

Amplitude encoding involves representing classical data by leveraging the amplitudes of quantum states. Consider a normalized vector where each component corresponds to a value. In this encoding method, each value in the vector is multiplied by its original coefficient and concatenated together. The advantage of amplitude encoding lies in its efficiency in qubit usage. Unlike basis encoding, which requires a significant number of qubits, amplitude encoding can represent data with fewer qubits.

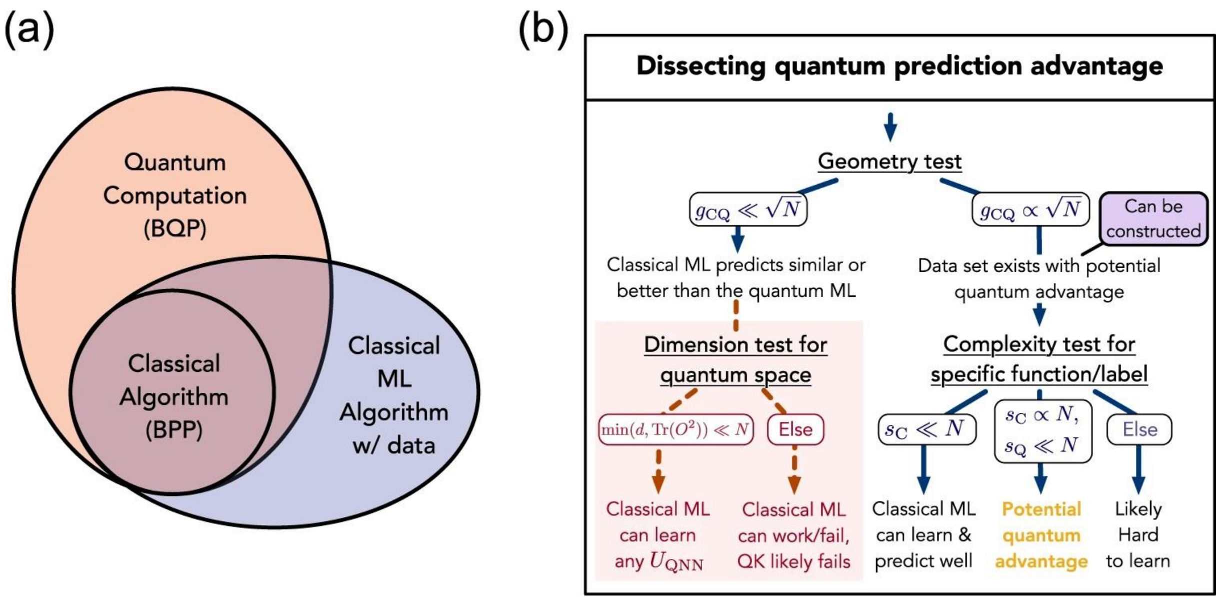


Figure 17: Illustration of the relation between complexity classes and a flowchart for understanding and prescreening potential quantum advantage.

(Hsin-Yuan Huang, 2021)

4.5.3 Conclusion

This chapter explores Quantum Machine Learning (QML) and its implications for resource exploration. It highlights the potential of AI and quantum technologies for unlocking new insights from geological data, leading to smarter, more efficient resource exploration. The symbiotic relationship between AI and QML is poised to revolutionize resource management.

Chapter 5 : Manipulation and model training

5 Chapter : Manipulation and model training

5.1 Prologue

This chapter highlights the complete workflow of approach step by step, connecting the dots of the previous chapters' fusion of numerous models and peace of code to make it. This chapter takes center stage by giving a complete, step-by-step approach that merges theoretical foundations and practical application.

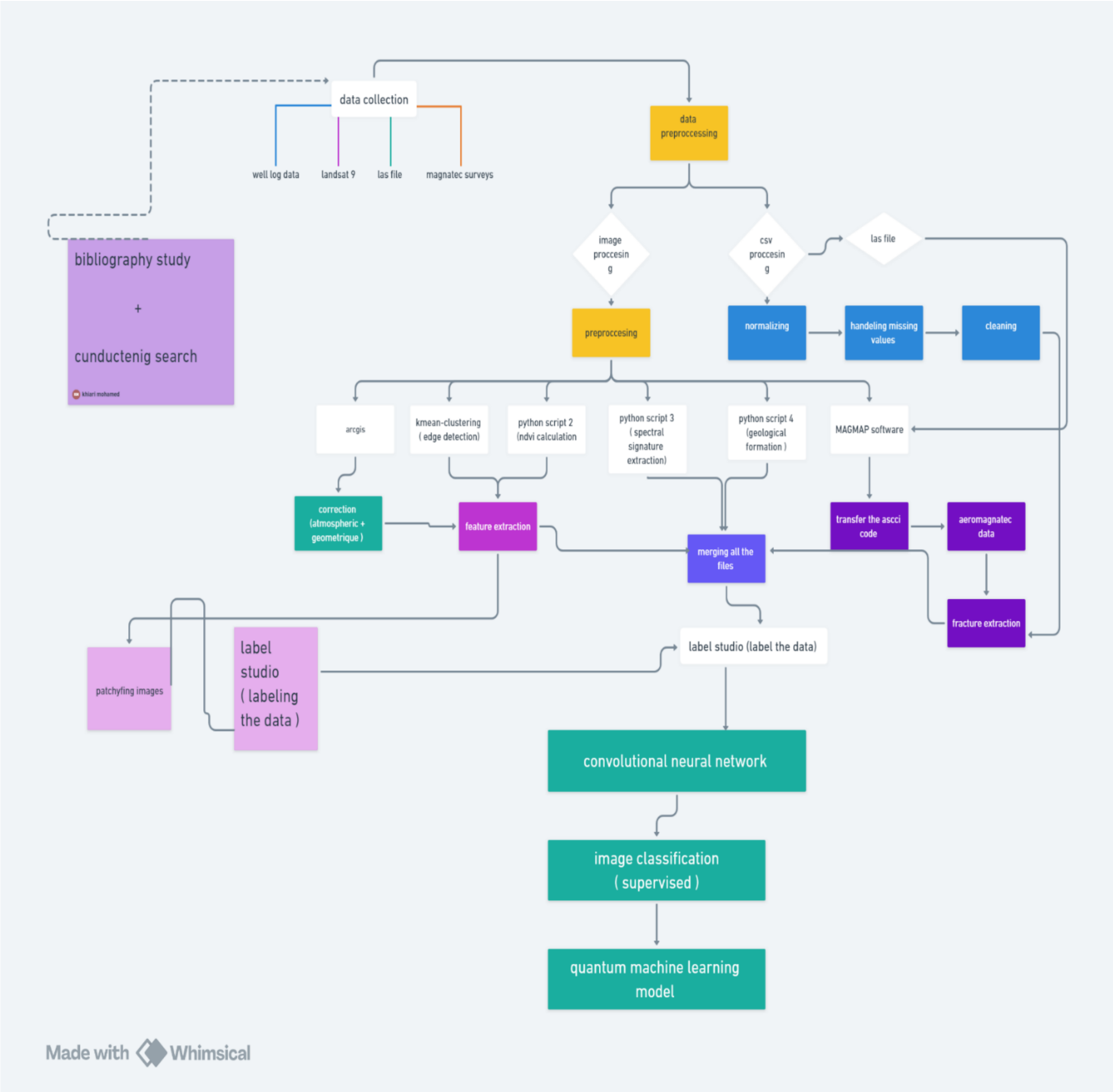


Figure 18: the whole workflow

5.2 Technological Arsenal: Tools Empowering Innovation

This table summarizes the varied spectrum of cutting-edge technology used throughout the project, each playing a critical role in promoting innovation and advancing the discipline.

Tableau II: technologie used

Technologies	Version	Utility	Logo
Qiskit pennylane	0.37.0-dev	Open-source framework for quantum computing	
TensorFlow	V2.16.1	Open-source machine learning framework by Google	
Google colab	3.10	Cloud-based platform for writing and executing code	
Python	V3.12	Programming language for executing codes	
Numpy	V1.26.4	library for numerical computing in Python	
matplotlib	3.16.1	Library for creating static, animated, and interactive visualizations in Python	
Pandas	V3.12	Library for data manipulation and analysis in Python	

Visual studio (vs code)	V1.89	Free source-code editor developed by Microsoft	
Kali Linux	V2024.1	Debian-based Linux distribution for digital forensics and penetration testing	

5.3 Model implementation

This section will outline the key steps involved in the process of creating the model from scratch. 2 models were implemented in this project segmentation tasks with UNET and the final QNN for the prediction of the complex data

5.3.1 Unet architecture:

Since 2015, it has been the go-to architecture for many machine-learning tasks due to its incredible performance in image generation, for example, with a set of input images and hand-annotated masks.

We can train an unet model by passing the image to a model, and we can produce an initial guess of the ground truth mask. Initially, our guess will be very good, but we can still use it to compare to our ground truth labels. This comparison gives us an error we can use as bios to adjust our model parameters, meaning the next time we get a slightly better prediction.

The unet model consists of an encoder floated by a decoder. The encoder is responsible for extracting features from the image, while the decoder is responsible for upsimling intermediate features and producing the final output. These two parts are symetrical.

Let's have a closer look at the encoder and decoder and the connection between them.

Features are passed through an encoder consisting of repeated convolution followed by pooling layers that extract intermediate features; these last get upsimpled by a corresponding decoder.

With saved copies of the encoder's features concatenated on the decoder's features by connected patterns, which finally produce the output, we can then simply calculate the loss and back propagate the gradients through the network to improve the model prediction.

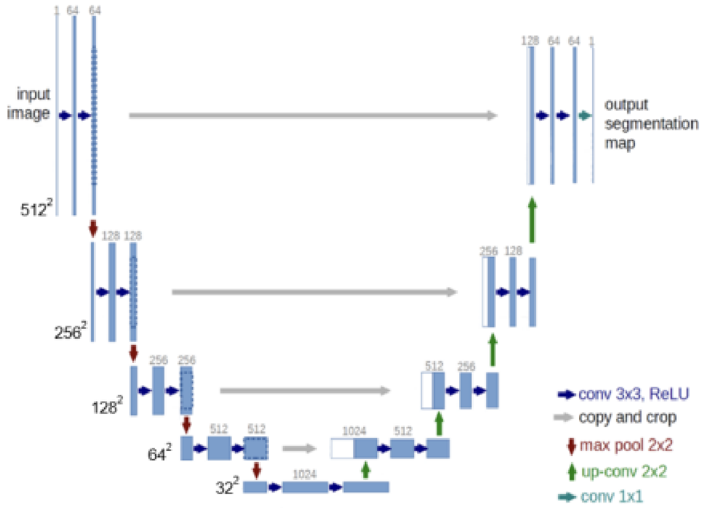


Figure 19:unet architecture

(vooban-a, 2017)

the encoder:

the encoder is made up of a series of repulsive convolutions. 3*3 convolutional layers, and after each one of them, the relu function is applied element-wise to each of the features. Between these stages, a *2 max pooling operation down simplifies the features, reducing the spatial dimensions of the features, and to compensate for this, the channels are doubled after each down simple operation.

the decoder:

In many ways, the reverse of the encoder is also a 3*3 convolution layer and a 2*2 pooling layer, but this time it up simplifies the data to restore the spatial resolution of features that were lost during the encoding phase.

Unet is a good choice for semantic segmentation tasks, such as the segmentation of Landsat patches, due to its design, skip connections, pooling techniques, data efficiency, adaptability, and performance. Because objects of interest in satellite images can vary greatly in size and appearance, its capacity to record both local features and global context is very useful

To train a U-net model, we need a large amount of data, which we obtained after processing Landsat 9 images. These images were converted into numpy arrays with dimensions of

[9775 * 7669], which are too large for direct computation. To address this, we used the Patchify library to divide these images into smaller patches of 256*256 pixels, each containing three dimensions (RGB colors). This resulted in approximately 6547 patches.

Creating a neural network model requires labeled data for ground truth. Manually labeling this many patches is time-consuming, so we labeled a small subset of patches. We then trained the U-net model to perform two tasks:

Labeling Unlabeled Patches: The model was trained to label the remaining unlabeled patches.

Semantic Segmentation: The model classified each pixel in the patches into one of four vegetation classes, ranging from no vegetation to high vegetation, based on different NDVI thresholds.

This approach allowed us to compare different areas or the same area over different time periods to detect anomalies.

Our U-net model architecture consists of:

20 convolutional layers

4 pooling layers

2 dropout layers

4 up sampling layers

4 concatenate layers

We defined data augmentation strategies and used the Adam optimizer with callbacks and checkpoints. The model was trained for a total of 100 epochs.

5.4 Quantum Neural Network (QNN) Model

After completing the U-net model, we moved on to implementing a Quantum Neural Network (QNN) model to further enhance our data analysis capabilities. The QNN model was designed to handle the complexity of predicting hydrocarbon presence by leveraging the power of quantum computing.

Implementation Details

Data Preparation: The data from the U-net model's segmentation output was preprocessed and encoded into quantum circuits. Additionally, we incorporated other data, including well logs and geological formations, to create a comprehensive dataset. The non-numeric columns were encoded using one-hot encoding, while the numeric columns were encoded using amplitude encoding.

Model Training: The QNN model was trained using a hybrid approach, where classical data was processed by quantum circuits before being passed to classical neural network layers.

Optimization: The Adam optimizer was used with callbacks and checkpoints over 180 epochs to ensure efficient training and model convergence.

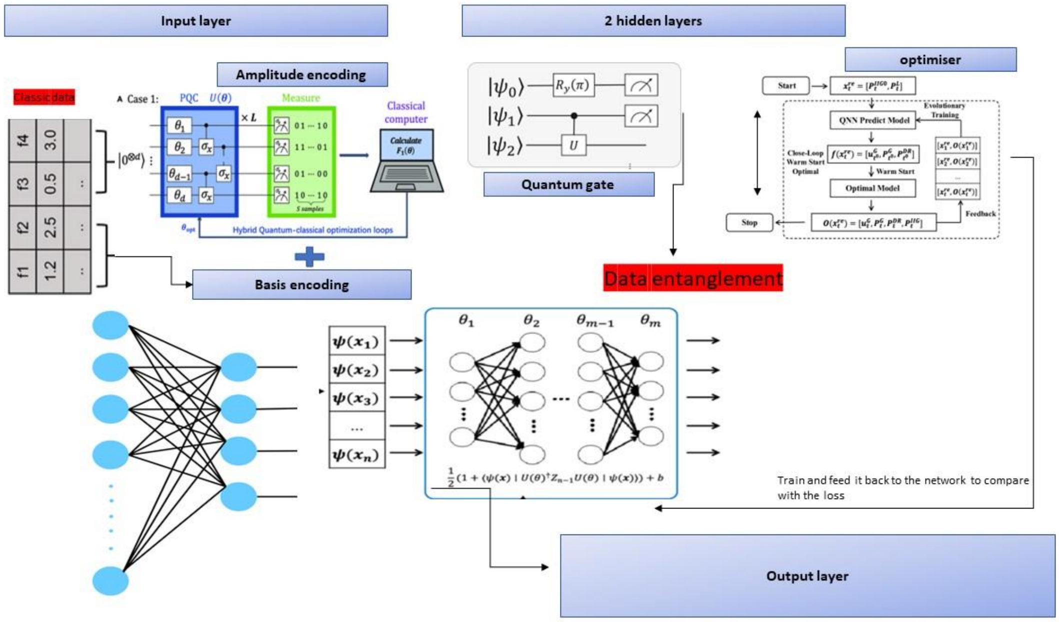


Figure 20: illustration of the quantum circuits used in this project

5.5 Model evaluation

- U-net Model

The U-net model produced impressive results, achieving an accuracy of 0.87. This indicates the model's robustness in performing the required segmentation and classification tasks. The ability of the U-net model to accurately classify each pixel in the image patches into the four defined vegetation classes demonstrated its effectiveness in processing and analyzing large-scale satellite imagery. This capability is crucial for identifying anomalies and changes in vegetation over time, which are essential for monitoring environmental changes and predicting hydrocarbon presence

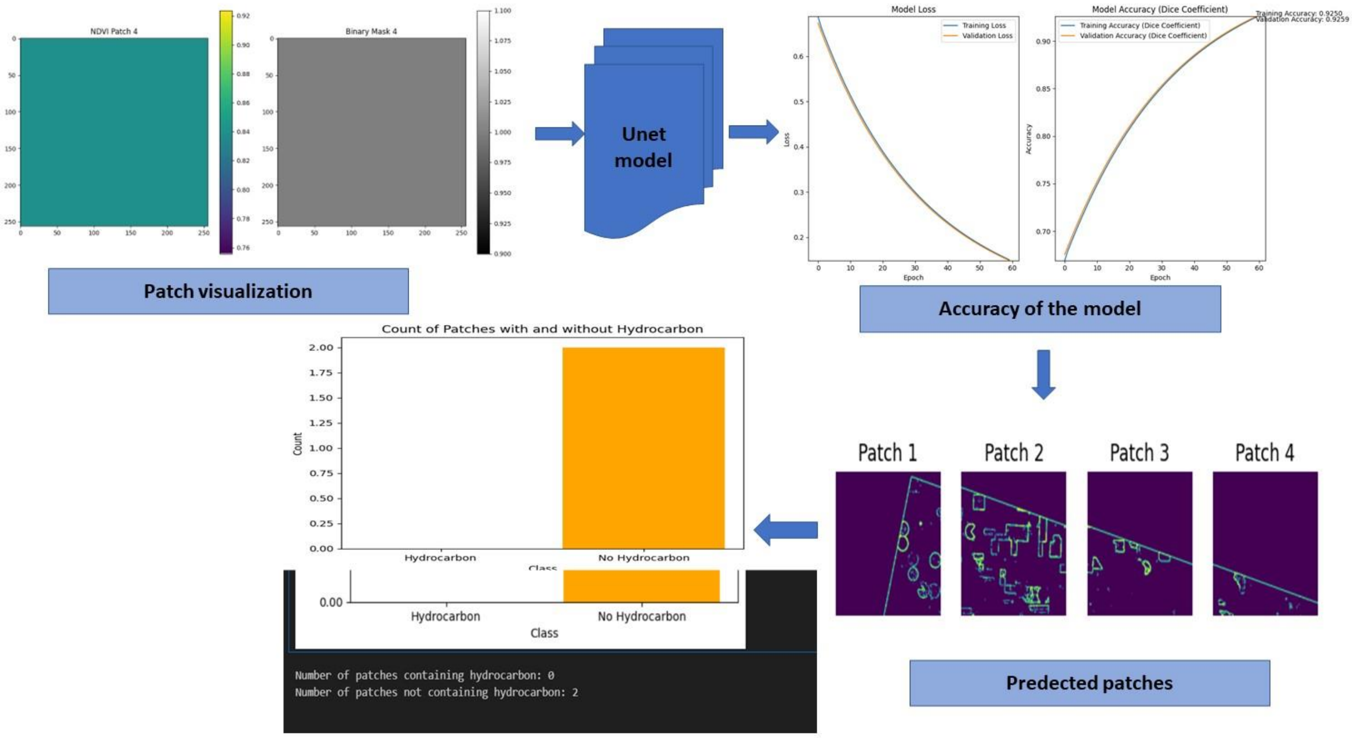


Figure 21: evaluation and plotting the result of unet

- Qnn

The QNN model demonstrated a significant improvement in prediction accuracy, achieving an accuracy ranging between 0.61 and 0.87. This variability highlights both the potential and the current limitations of quantum computing in handling complex data prediction tasks. By leveraging quantum circuits, the QNN model effectively processed and analyzed the combined dataset of segmented imagery and additional geological data, leading to enhanced predictive capabilities. The integration of quantum computing with classical machine learning techniques showcased the effectiveness of this hybrid approach in managing and interpreting complex datasets.

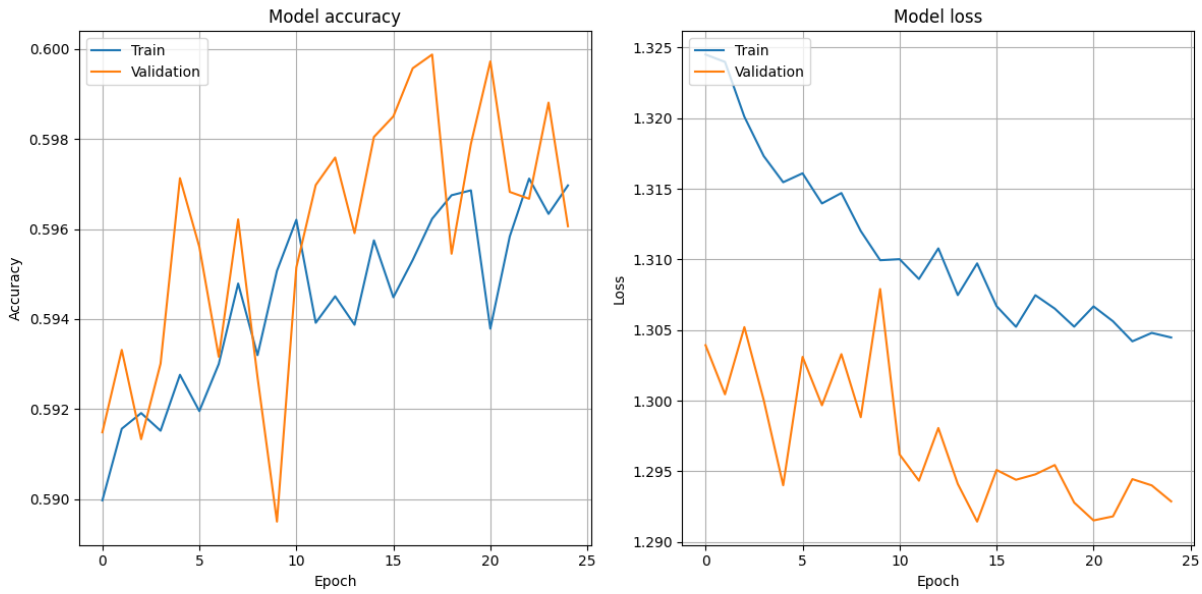


Figure 22: QNN result after first training

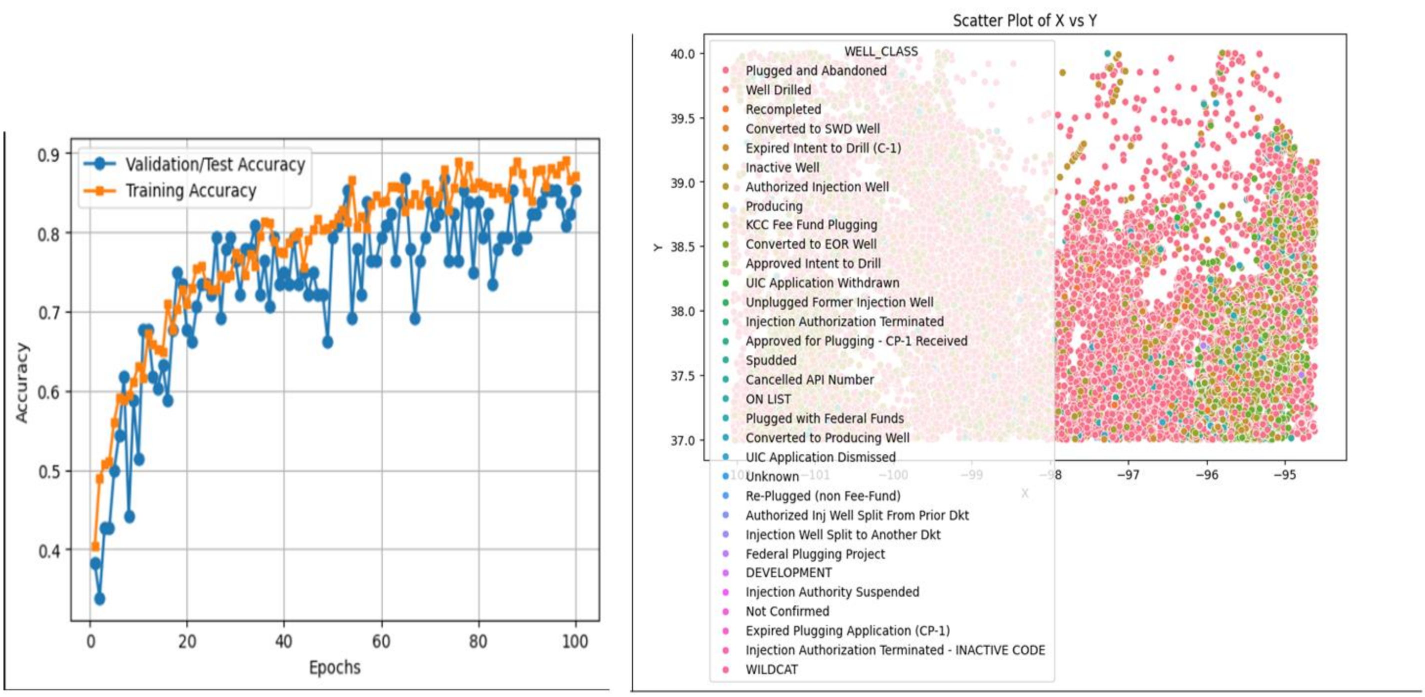


Figure 23: QNN result after regularization

5.5.1 Impact of Regularization

After applying regularization techniques, the QNN model demonstrated enhanced stability and improved accuracy. Regularization helped mitigate overfitting, ensuring that the model generalized better to unseen data. This process refined the model's predictive capabilities, leading to more consistent and reliable outcomes as depicted in Figure 23

Final result of the workflow :

This Streamlit app provides an interactive demonstration of our hybrid classical-quantum model for hydrocarbon exploration. By leveraging the computational power of both classical machine learning techniques and quantum computing models, this app showcases the results of our innovative approach. Users can visualize the processed satellite imagery, examine the accuracy of the image segmentation and hydrocarbon prediction models, and explore the advantages of this cutting-edge method in a user-friendly interface

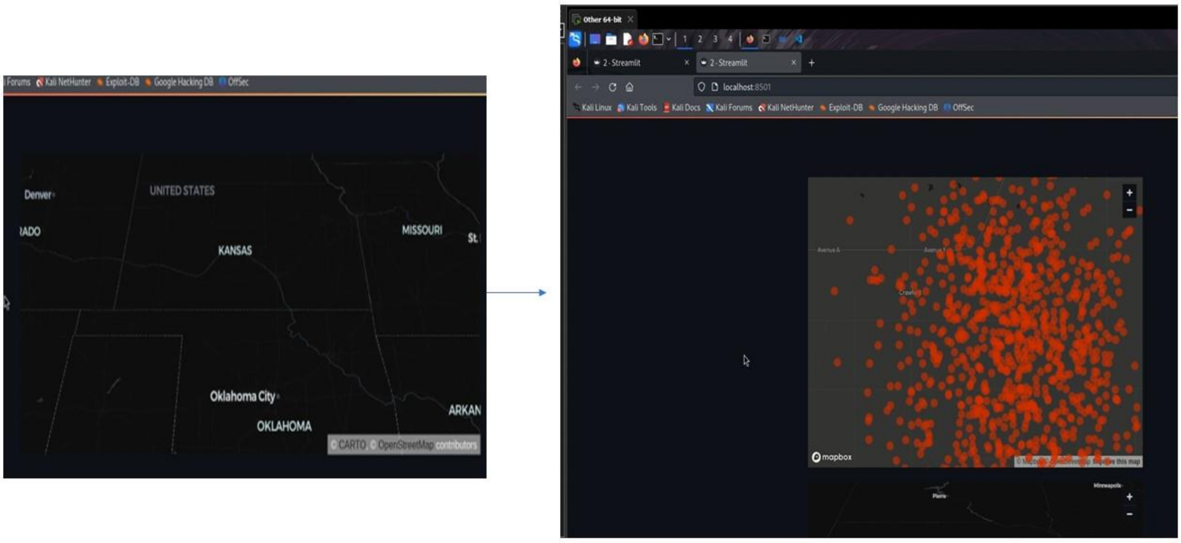


Figure 24: mapping the result

In comparison, the U-net model attained a high accuracy of 0.92 when segmenting and classifying satellite imagery into vegetation classes. This high accuracy demonstrates the U-net model's ability to handle large-scale picture data, which is vital for environmental monitoring and hydrocarbon prediction. The performance of the U-net model established a standard for evaluating the QNN model's upgrades and the integration of quantum computing techniques.

5.5.2 Advantages of the Hybrid Approach

In comparison to traditional methods, which often take over a year, the hybrid approach of classical and quantum computing technologies provides improved precision, time and cost efficiency, and predictive capabilities, particularly in complex and high-dimensional data fields. These advancements not only pave the way for more sustainable and efficient resource discovery practices but also highlight the potential for broader applications in environmental monitoring and resource management

5.5.3 Conclusion

Looking ahead, the continuing refining of these combinations has enormous potential for major improvements in understanding and managing natural resources. Encouraging interdisciplinary collaboration and leveraging new innovations will be key in moving toward a more sustainable and resilient future.

Chapter 6 : conclusion

6 Chapter: conclusion

Finally, this study represents a ground-breaking approach to geographic data analysis and prediction, providing a major improvement over existing methods of hydrocarbon exploration. By seamlessly merging conventional machine learning approaches with cutting-edge quantum computing models, we have paved the way for more accurate, efficient, and sustainable resource discovery practices.

Throughout the project, which spanned three months from inception to completion, we meticulously navigated through the complexities of preprocessing vast amounts of satellite imagery, training sophisticated U-net and QNN models, and evaluating the results to extract meaningful insights. The culmination of these efforts has yielded promising outcomes, with our models demonstrating high accuracy in tasks such as image segmentation and prediction of hydrocarbon presence.

Importantly, the novel combination of classical and quantum computing technology has created new opportunities for environmental monitoring and resource development. By leveraging quantum circuits' tremendous computational capacity, we have overcome the constraints of traditional approaches, allowing us to examine complex datasets with unparalleled precision and efficiency.

Moreover, our approach offers notable advantages in terms of time and cost efficiency. While traditional convolutional methods often require extensive computational resources and lengthy processing times, our hybrid approach significantly reduces both the time and resources needed for analysis. In comparison to the traditional methods that can take more than a year to yield results, our methodology completed the analysis in a fraction of the time, saving valuable resources and minimizing costs.

Looking ahead, we see enormous potential for further refining and growth of our approaches. As we continue to investigate the synergies between classical and quantum computing, we will be able to open up new possibilities for better understanding and management of our natural resources. Furthermore, by encouraging interdisciplinary collaboration and harnessing emerging technology, we may speed progress toward a more sustainable and resilient future.

In essence, this project demonstrates the transforming power of new thought and technical innovation in addressing pressing environmental issues. By pioneering new techniques to geospatial data analysis and prediction, we are not only modernizing resource discovery, but also laying the groundwork for a more wealthy and sustainable future.

7 Références

Anderson, G. A. (2024). *Using the USGS Landsat Level-1 Data Product*. Récupéré sur usgs: https://www.usgs.gov/landsat-missions/using-usgs-landsat-level-1-data-product

desRosiers, B. (2023, june 8). *Using Geology to Find Natural Gas & Oil.* Retrieved from wellsaidcoterra.com: https://wellsaidcoterra.com/using-geology-to-find-natural-gas-oil/#:~:text=Once%20the%20geologists%20have%20gathered,techniques%2C%20such%20as%20hydraulic%20fracturing.

Ferrie, C. (2023, june). *available from: Quantum Machine Intelligence*. Récupéré sur researchgate: https://www.researchgate.net/figure/The-four-quantum-machine-learning-development-paradigms-compared-against-data-and_fig5_371604978

GEOPROVIDER. (2023). *GEOPROVIDER*. Récupéré sur GEOPROVIDER: https://geoprovider.no/wells/

Gurucharan, M. (2022, juillet 27). *Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network.* Retrieved from upGard: https://www.upgrad.com/blog/basic-cnn-architecture/

Hsin-Yuan Huang, M. B. (2021, may 11). *Power of data in quantum machine learning.* Récupéré sur Nature Communications: https://www.nature.com/articles/s41467-021-22539-9#Fig1

Kalbas, K. A. (2008, may 21). *Kansas Geological Survey*. Récupéré sur university of kensas : https://kgs.ku.edu/

Manuela Weigold, J. B. (2021, december 10). *Encoding patterns for quantum algorithms*. Retrieved from the institution of engeenring technology : https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/qtc2.12032

mission, l. (2024, march 19). *Spectral Bandpasses for all Landsat Sensors*. Retrieved from usgs: https://www.usgs.gov/media/images/spectral-bandpasses-all-landsat-sensors

quantum-ml. (2019, december 28). *Quantum Machine Learning - 24 - Encoding Classical Information*. Retrieved from youtube: https://www.youtube.com/watch?v=LHzWAyT5rMI

Roberts, K. (2022, november 22). *spyro: a Firedrake-based wave propagation and full-waveform-inversion finite-element solver*. Récupéré sur researchgate: https://www.researchgate.net/publication/365877070_spyro_a_Firedrake-based_wave_propagation_and_full-waveform-inversion_finite-element_solver

Sarker, I. H. (2021, March 22). *Machine Learning: Algorithms, Real-World Applications and Research Directions*. Récupéré sur ncbi.nlm.nih.gov: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7983091/

surveey, u. g. (2022, febraery). *Landsat 9*. Récupéré sur usgs: https://www.usgs.gov/landsat-missions/landsat-9

TensorFlow. (2024). *pennylane.ai*. Récupéré sur pennylane: https://pennylane.ai/

underground, s. (2021, april 20). *Tutorial: Big data lithology prediction with machine learning*. Récupéré sur youtube: https://www.youtube.com/live/TBa791rkIDE

vooban-a. (2017, augest 24). *Satellite Image Segmentation: a Workflow with U-Net*. Retrieved from medium: https://medium.com/vooban-ai/satellite-image-segmentation-a-workflow-with-u-net-7ff992b2a56e

VYAS, P. (2020). *Well log data*. Retrieved from Kaggle: https://www.kaggle.com/datasets/prateekvyas/well-log-data/data

Yuewei Wang, X. C. (2023, march 20). *Cyber-physical oil spill monitoring and detection for offshore petroleum risk management service*. Récupéré sur Scientific Reports: https://www.nature.com/articles/s41598-023-30311-w

Annexes

8 Annexes of google colab

#Installing libraries and mounting the derive

```
!nvidia-smi
!pip install pandas scikit-learn tensorflow pennylane
!pip install rasterio
import rasterio
import numpy as np
import tensorflow as tf
from tensorflow.keras.layers import Input, Conv2D, MaxPooling2D, concatenate, UpSampling2D
from tensorflow.keras.models import Model
from tensorflow.keras.callbacks import ModelCheckpoint, EarlyStopping, ReduceLROnPlateau
import matplotlib.pyplot as plt
from google.colab import drive
import os
from google.colab import drive
drive.mount('/content/drive')
```

#Load the data

```
def load_ndvi_patches(directory, patch_size=(256, 256)):
    ndvi_patches = []
    for filename in os.listdir(directory):
        if filename.endswith(".tif"):

#  load binary masks as ground truth
def load_ground_truth(directory, patch_size=(256, 256)):
    ground_truth_masks = []
    for filename in os.listdir(directory):
        if filename.startswith("binary_mask"):
data_path = '/content/drive/MyDrive/QNNdataset/feature_engineered_data.csv'
data = pd.read_csv(data_path)
```

#Building the architecture of unet

```
def unet_2(pretrained_weights=None, input_size=(256, 256, 3)): #input size 256*256 ,3 = 3chanels = rgb colors
    inputs = Input(input_size)
    conv1 = Conv2D(64, 3, activation='relu', padding='same', kernel_initializer='he_normal')(inputs)
    conv1 = Conv2D(64, 3, activation='relu', padding='same', kernel_initializer='he_normal')(conv1)
    pool1 = MaxPooling2D(pool_size=(2, 2))(conv1)
```

#Create data augmentation

```
train_datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.1,
    height_shift_range=0.1,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    vertical_flip=True,
    fill_mode='nearest'
)
val_datagen = ImageDataGenerator()
```

#Train the model with dice coefficient

```
def dice_coefficient(y_true, y_pred):
    y_true_f = K.flatten(y_true)
    y_pred_f = K.flatten(y_pred)
    intersection = K.sum(y_true_f * y_pred_f)
    return (2. * intersection + 1.) / (K.sum(y_true_f) + K.sum(y_pred_f) + 1.)

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=[dice_coefficient])

#train the model
history = model.fit(train_patch_generator,
                    steps_per_epoch=len(train_patches) / 16,
                    epochs=60,
                    validation_data=val_patch_generator,
                    validation_steps=len(val_patches) / 16,
                    callbacks=[checkpoint_callback, early_stopping, reduce_lr])
```

#Encode the data into a quantum state

```
#  binary encoding
def binary_encode(x):
    binary_string = '{:08b}'.format(int(abs(x) * 255))  # 8-bit encoding
    return [int(bit) for bit in binary_string]

# numeric columns
numeric_columns = X_train.select_dtypes(include=np.number).columns.tolist()
scaler = MinMaxScaler()
X_train_normalized = scaler.fit_transform(X_train[numeric_columns])
min_value = np.min(X_train_normalized)
X_train_normalized += abs(min_value)
X_train_rescaled = X_train_normalized / (np.max(X_train_normalized) + 1e-10)
X_train_binary = np.array([[binary_encode(value) for value in row] for row in X_train_rescaled])
X_val_normalized = scaler.transform(X_val[numeric_columns]) + abs(min_value)
```

#Build the QNN architecture

```
# Define the model
model = Sequential([
    Dense(512, activation='relu', input_shape=(363 * 8,)),
    BatchNormalization(),
    Dropout(0.5),
    Dense(256, activation='relu'),
    BatchNormalization(),
    Dropout(0.5),
    Dense(128, activation='relu'),
    BatchNormalization(),
    Dropout(0.5),
    Dense(64, activation='relu'),
    BatchNormalization(),
    Dropout(0.5),
    Dense(len(y_train.unique()), activation='softmax')
])
```

train the model

```
# Flatten
X_train_binary_flat = X_train_binary.reshape(X_train_binary.shape[0], -1)
X_val_binary_flat = X_val_binary.reshape(X_val_binary.shape[0], -1)
```

```
X_test_binary_flat = X_test_binary.reshape(X_test_binary.shape[0], -1)
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])

# Train
history = model.fit(
    x=X_train_binary_flat,
    y=y_train_encoded,
    validation_data=(X_val_binary_flat, y_val_encoded),
    epochs=60,
    callbacks=callbacks_list
)
```

#predect and evaluate

```
# Evaluate
test_loss, test_accuracy = model.evaluate(X_test_binary_flat, y_test_encoded)
print("Test Accuracy:", test_accuracy)
test_predictions = model.predict(test_patches)
for i in range(5):
    plt.imshow(test_patches[i][:,:,0], cmap='gray')
    plt.show()
    plt.imshow(test_predictions[i][:,:,0], cmap='viridis')
    plt.show()
```

Résumée :

L'exploration des ressources en hydrocarbures est une entreprise vitale et complexe qui se caractérise généralement par des dépenses financières importantes, une planification minutieuse et des procédures longues. Cette étude propose une nouvelle stratégie pour transformer la découverte d'hydrocarbures en fusionnant les réseaux de neurones classiques et l'apprentissage automatique quantique (QML). Nos objectifs sont d'améliorer la précision des prévisions, d'accélérer le processus d'exploration et d'élargir l'application de l'identification des hydrocarbures en utilisant la capacité de calcul des circuits quantiques.

Afin d'évaluer les données géologiques et d'anticiper la présence d'hydrocarbures, notre méthodologie mélange des modèles d'apprentissage automatique classiques et quantiques, en utilisant spécifiquement les modèles U-net et Quantum Neural Network (QNN). Lors des tests de segmentation d'images, le modèle QNN a présenté une précision variable comprise entre 0,61 et 0,87, soulignant à la fois son potentiel et les limites actuelles de l'informatique quantique. En revanche, le modèle U-net a démontré des performances robustes avec une précision de 0,92. La recherche a nécessité une préparation minutieuse des données, qui comprenait la mise en place de données géologiques dans des circuits quantiques et la préparation d'images satellite. Une convergence efficace des modèles et des capacités de prédiction optimisées a été garanties par la technique de formation hybride. Selon nos résultats, la méthodologie hybride QML réduit considérablement le temps et les dépenses impliqués par les techniques d'exploration conventionnelles, offrant ainsi un cadre plus efficace et plus durable pour la découverte de ressources à l'avenir. En plus de faire progresser le domaine de l'exploration des hydrocarbures, ce travail ouvre la porte pour des utilisations plus répandues de QML dans la gestion des ressources et la surveillance de l'environnement. L'informatique quantique moderne combinée aux techniques traditionnelles d'apprentissage automatique offre une voie prometteuse pour des méthodes d'exploration plus précises, plus économiques et plus respectueuses de l'environnement. En affinant continuellement nos modèles et en favorisant la collaboration interdisciplinaire, nous visons à libérer davantage le potentiel des approches hybrides quantiques classiques, contribuer à un avenir énergétique plus résilient et durable.

Abstract:

Hydrocarbon resource exploration is a vital and intricate undertaking that is typically marked by large financial outlays, thorough planning, and drawn-out procedures. This study offers a novel strategy to transform hydrocarbon discovery by fusing classical neural networks and quantum machine learning (QML). Our goals are to improve prediction accuracy, speed up the exploration process, and broaden the application of hydrocarbon identification by utilizing the computational capacity of quantum circuits.

In order to assess geological data and anticipate the presence of hydrocarbons, our methodology blends classical and quantum machine learning models, specifically using the U-net and Quantum Neural Network (QNN) models. In image segmentation tests, the QNN model exhibited varied accuracy between 0.61 and 0.87, highlighting both its potential and the current limitations of quantum computing. In contrast, the U-net model demonstrated robust performance with an accuracy of 0.92. The research required careful data preparation, which included putting geological data into quantum circuits and preparing satellite pictures. Efficient model convergence and optimized prediction capabilities were guaranteed by the hybrid training technique. According to our findings, the hybrid QML methodology considerably lowers the time and expense involved with conventional exploration techniques, offering a more effective and sustainable framework for resource discovery in the future. In addition to advancing the field of hydrocarbon exploration, this work opens the door for more widespread uses of QML in resource management and environmental monitoring. Modern quantum computing combined with traditional machine learning techniques provides a promising way forward for more precise, economical, and ecologically friendly exploration methods. By continually refining our models and fostering interdisciplinary collaboration, we aim to further unlock the potential of quantum-classical hybrid approaches, contributing to a more resilient and sustainable energy future.

ملخص

يعد استكشاف الموارد الهيدروكربونية مهمة حيوية ومعقدة تتميز عادة بنفقات مالية كبيرة وتخطيط شامل وإجراءات طويلة. تقدم هذه الدراسة استراتيجية جديدة لتحويل اكتشاف الهيدروكربون من خلال دمج الشبكات العصبية الكلاسيكية والتعلم الآلي الكمي (QML). تتمثل أهدافنا في تحسين دقة التنبؤ، وتسريع عملية الاستكشاف، وتوسيع نطاق تطبيق تحديد الهيدروكربونات من خلال الاستفادة من القدرة الحسابية للدوائر الكمومية.

من أجل تقييم البيانات الجيولوجية وتوقع وجود الهيدروكربونات، تمزج منهجيتنا بين نماذج التعلم الآلي الكلاسيكية والكمية، وتحديدًا باستخدام نماذج U-net والشبكة العصبية الكمومية (QNN). في اختبارات تجزئة الصور، أظهر نموذج QNN دقة تتراوح بين 0.61 و0.87، مما يسلط الضوء على إمكاناته والقيود الحالية للحوسبة الكمومية. في المقابل، أظهر نموذج U-net أداءً قويًا بدقة 0.92. وتطلب البحث إعدادًا دقيقًا للبيانات، والذي تضمن وضع البيانات الجيولوجية في الدوائر الكمومية وإعداد صور الأقمار الصناعية. تم ضمان التقارب الفعال للنموذج وقدرات التنبؤ المحسنة من خلال تقنية التدريب الهجين. وفقًا للنتائج التي توصلنا إليها، فإن منهجية QML الهجينة تقلل بشكل كبير من الوقت والنفقات المرتبطة بتقنيات الاستكشاف التقليدية، مما يوفر إطارًا أكثر فعالية واستدامة لاكتشاف الموارد في المستقبل. بالإضافة إلى تطوير مجال استكشاف الهيدروكربون، فإن هذا العمل يفتح الباب لاستخدامات أكثر انتشارًا لـ QML في إدارة الموارد والمراقبة البيئية. توفر الحوسبة الكمومية الحديثة جنبًا إلى جنب مع تقنيات التعلم الآلي التقليدية طريقة واعدة للمضي قدمًا نحو طرق استكشاف أكثر دقة واقتصادية وصديقة للبيئة. ومن خلال التحسين المستمر لنماذجنا وتعزيز التعاون متعدد التخصصات، نهدف إلى إطلاق العنان لإمكانات الأساليب

الهجينة الكمومية الكلاسيكية، المساهمة في مستقبل طاقة أكثر مرونة واستدامة.