Automatic Method Based on PSO-optimized Vision-Transformer for Gas Detection in 2D Seismic Images

Domingos A. Dias Junior
UFMA
Av. dos Portugueses, Campus do Bacanga
65085-580, São Luís, MA, Brazil
domingos.adj@nca.ufma.br

Luana Batista da Cruz
UFCA
Av. Tenente Raimundo Rocha, 1639, Cidade Universitária
63048-080, Juazeiro do Norte, CE, Brazil
luana.batista@ufca.edu.br

João O. Bandeira Diniz
IFMA
BR-226, SN, Campus Grajaú, Vila Nova
65940-00, Grajaú, MA, Brazil
joao.bandeira@ifma.edu.br

Aristófanes Corrêa Silva
UFMA
Av. dos Portugueses, Campus do Bacanga
65085-580, São Luís, MA, Brazil
ari@dee.ufma.br

Anselmo C. Paiva
UFMA
Av. dos Portugueses, Campus do Bacanga
65085-580, São Luís, MA, Brazil
paiva@deinf.ufma.br

Marcelo Gattass
PUC-Rio
R. São Vicente, 225, Gávea, 22453-900, Rio de Janeiro, RJ, Brazil
mgattass@tecgraf.puc-rio.br

ABSTRACT

One of the geophysical techniques most frequently utilized in the oil and gas (O&G) sector for hydrocarbon prospecting is seismic reflection. The seismic reflection technique is essential for an estimate the location and volume of gas accumulations in various onshore fields. However, this technique generates a large amount of data, and its data acquisitions are noisy. Thus it takes a while to analyze and interpret seismic data. Computational techniques based on machine learning have been proposed considering Direct Hydrocarbon Indicators (DHIs) to assist geoscientists in such activities. In this paper, we describe a method to detect gas accumulations based on the Particle Swarm Optimization (PSO) algorithm and the Vision Transformer neural network (ViT). In the best scenario, the proposed method achieved a sensitivity of 88.60%, a specificity of 99.56% and an accuracy of 99.37%. We present some tests performed on Parnaíba Basin and Netherlands F3-Block fields. Thus, it demonstrates that the proposed method is promising for assisting specialists in gas exploration tasks.

CCS Concepts

•**Computing methodologies** → *Image segmentation;*

Keywords

Deep Learning; Computer Vision; Seismic Reflection; PSO; Vision Transformer; Natural Gas Exploration

1. INTRODUCTION

Seismic reflection is one of the most used geophysical methods in the oil and gas (O&G) industry to extract information related to geological structures, lithology and rock properties [24, 9, 5]. Moreover, that method also has been used to estimate the location and volume of gas accumulations, contributing to reducing exploration risks.

Processing these type of data requires a lot of time and effort from qualified teams that choose a geological model after analyzing several potential scenarios. However, the time necessary to perform the seismic data interpretation is not compatible with the short deadlines imposed by the industry [27]. For this reason, various studies have developed methods that integrate seismic attribute extraction with machine learning algorithms [17, 31], considering supervised methods and a labeled dataset. In this scenario, the features are constituted of samples (values or derived seismic properties) that are coupled with the appropriate class name. This method was used to detect geological faults in seismic images [17] using a texture descriptor to extract seismic attributes associated with a Support Vector Machine (SVM). The seismic attributes with neural networks can be found in [31], who used 12-dimensional attributes and a Multi-Layer Perceptron (MLP) to detect fault lines.

Concerning the object of study addressed in this research, which is the detection of potential natural gas accumulations, there are few related works in the literature. Among these works stands out the work proposed by [7]. This work presents a method for automatic detection and delimitation of the natural gas region in seismic images (2D) using MLP-Mixer and U-Net. The proposed method obtained competitive results with an F1-score of 84.18%, an accuracy

of 99.6%, a sensitivity of 86.85%, and specificity of 99.79%, using the seismic imaging database named Netherlands F3-Block [30].

Furthermore, a study using a Long Short-Term Memory (LSTM) deep learning model [27] proposed a new methodology to detect DHIs in seismic data using a time series approach. In this study, they transform 3D cube seismic data from Netherleands F3-block to temporal seismic traces. Thus, each seismic trace was divided into patches that are the entrance to the LSTM network, and its outputs indicate the presence or absence of gas. The results achieved, evaluated by the indices of sensitivity, specificity, accuracy, and AUC were 97.1%, 96.83%, 97.1%, and 99.2%, respectively.

Lately, a transfer learning technique in an LSTM model [26] was proposed to expand the existing classifier and apply it to a different type of seismic surveys in the Parnaíba Basin. Moreover, in order to check that methodology based on seismic trace, other networks based on different architectures were developed using an improved encoder-decoder LSTM [4] and a Transformer neural network [10]. All those networks have been implemented in ALINE, a computational tool for the assessment of gas accumulations [25]. In the best scenario, the proposed method obtained 97.62% of accuracy, 90.28% of sensitivity, 97.67% of specificity and 98.84% of AUC.

Recently, a Deep Transformer Neural Network (Time2Vec Transformer)[11] was proposed to detect the probability of the existence of hydrocarbons in seismic images of the Park of the Hawks. In this work, seismic images 2D composed the dataset. The images were decomposed into one-dimensional time series that fed a Transformer neural network. This network had its input layer modified for a temporal embedding (Time2Vec). In addition, a fully connected layer replaced its transformer decoder output. The proposed method was able to detect DHI and help specialists in this task that demands time and human effort with less time and greater precision. The best results had an accuracy of 98.87%, a sensitivity of 96.12%, a specificity of 98.92%, and an area under the ROC curve (AUC) of 97.52%.

We present an improved method for identifying possible gas accumulations using the Vision Transformer (ViT) Neural Network in this paper. ViT is a classification neural network based on the original Transformer, which was first proposed to solve natural language processing (NLP) challenges [12]. Furthermore, we apply the Particle Swarm Optimization (PSO) algorithm to optimize the ViT model hyperparameters. [21, 19]. The results show that our proposal improves the accuracy of the Transformer neural network and increases its efficiency by reducing the spent computational time.

Therefore, the present work proposes a 2D method capable of detect potential gas accumulations in reflection seismic images, using a Vision Transformer Neural Network and Particle Swarm Optimization (PSO). In this way, it is expected to help specialists in the task of delimiting gas regions. Then, the key contributions of this work are summarized as follows: (a) this study develops a method capable to direct detect hydrocarbons indicators using Vision Transformer; (b) It optimizes the Vision Transformer hyperparameters using an evolutionary algorithm.

This paper is organized as follows: Section 2 presents the materials used and the proposed method. Section 3 discusses the experiments performed out to validate our research. Finally, our conclusions and future works are presented in Section 4.

2. MATERIALS AND METHOD

2.1 Datasets

Diverse samples are required to develop a robust method. In this study, we use seismic data from very different sources. The first set of seismic data comes from the F3-block of the North Sea [30]. The second is a set of reflection seismic images from Brazil's northeast region [8].

The first seismic dataset used to support the proposed method is the Netherlands Offshore F3-Block public 3D seismic survey. This dataset is available in the dGB Earth Sciences Open Seismic Repository (1987), maintained by dGB Earth Sciences. The Netherlands F3 is located in the Dutch offshore portion of the Graben Central basin, situated approximately 180 km off the Dutch coast [30] (Figure 3).

The dataset is 384 km^2 in size, with 651 inlines and 951 crosslines, a time interval of 1848ms, a sample rate of 4ms, and a bin size of 25m [30]. The research is being conducted on the Dutch coast of the North Sea. In addition to the original 3D seismic, the repository includes filtered versions of the data, acoustic impedance cubes, some already computed seismic attributes, four wells with markers and geophysical profiles (F02-1, F03-1, F03-4, F06-1), and eight corresponding seismic horizons [30, 18, 29].

To manually locate the gas pockets and their possible indicators, the specialist used information in addition to the available cube and other accessible information. Along with the original seismic data, impedance cubes and good geophysical profiles played significant roles in labeling the grid input [30, 27, 29]. Expert input is essential for proper training and validation steps of the deep learning algorithm and for evaluating the methodology [27]. However, the AI cube does not present information for the entire seismic record, and the primary concentration of gas indicators is almost restricted near the sigmoidal strata [30, 2]. In this sense, the seismic images were cropped to a region of interest, avoiding possible noisy data imputation and misleading gas labeling, thus improving the quality of neural network training [27].

The second seismic dataset used in this research is from the Paleozoic Parnaba Basin. The Basin is a typical oval-shaped intracratonic basin formed on a continental basement during the South American Platform Stabilization Stage. It is situated between the Amazonic Craton and the Borborema Province [3, 8]. It extends more than 600,000 km^2 and has a depocenter with a thickness of over 3,500 m [8].

The Poti Formation sandstones with good poroperm characteristics in the basin serve as the primary reservoirs in six of the seven existing fields in the Parnaíba Basin [8]. The major producing area is known as the 'Parque dos Gaviões' translated as 'Park of the Hawks Field' in a reference to the native Brazilian hawk species that the fields are named (Figure 2).

The available data consists of 380 seismic sections located at the Park of the Hawks Field area. The database was provided and labeled by Eneva S.A., a Brazilian energy company. These data were collected at various time intervals with varying climate, geology, acquisition procedure, and other external factors. As a result, the data are numerous and heterogeneous.

First, we employed a data preparation method, and then
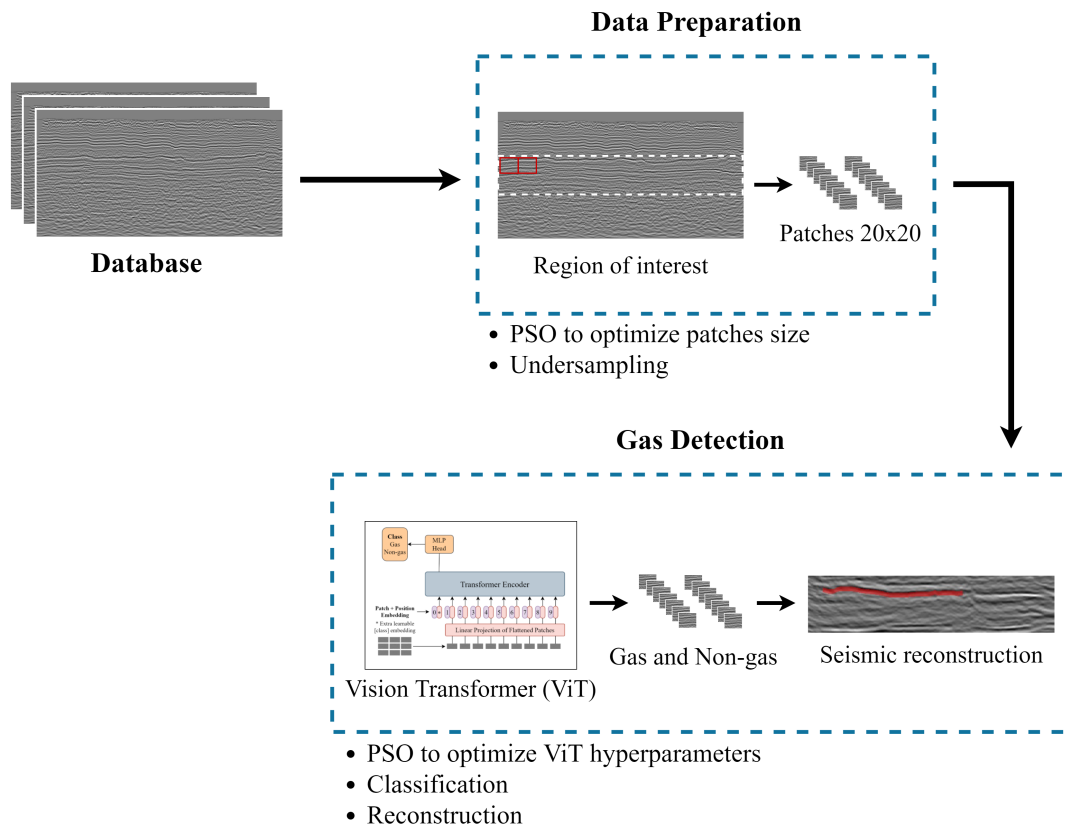
Data Preparation

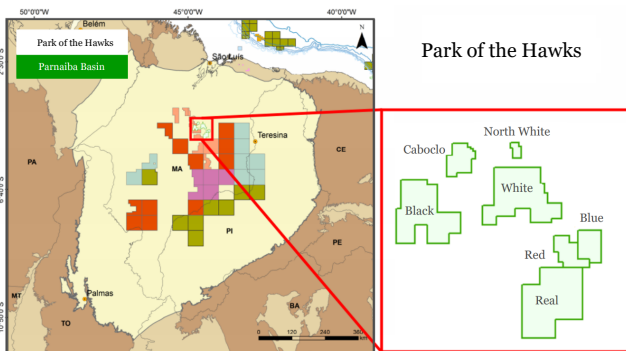

Figure 1: Steps of the proposed method
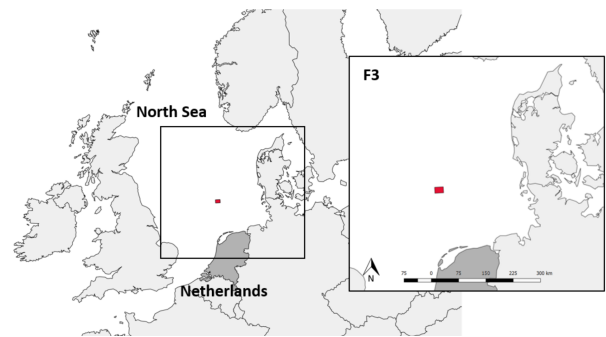

Figure 2: Park of the Hawks Field [8]

Figure 3: Location of the F3 3D survey in the North Sea, Netherlands offshore [30]

we used the Vision Transformer model to detect possible gas accumulations in seismic images. Figure 1 depicts each of these steps, with more information provided in the following sections.

2.2 Preprocessing

First, we performed a data preprocessing step in the seismic images from the two sources (Netherlands and Parnaiba Basin). First, we applied initial processing to the Netherlands F3 dataset, as it is a three-dimensional cube, to decompose it into 2D images perpendicular (crossline) and parallel (inline) to the direction in which the data were acquired

(Figure 4).

Then, ENEVA geoscientists delimited the Regions of Interest (ROI) that may contain gas accumulation based on field data, drilled exploratory wells, and inference. Besides, the ROI is individual to each image and defines a region with seismic patterns that the model may learn to detect gas or non-gas dividing structures. Subsequently, using a fixed window size, we use a sliding window technique to extract patches over ROIs. Finally, Particle Swarm Optimization (PSO) sets the window size that produces the best results. The Table 3 presents the window size hyperparameters op-
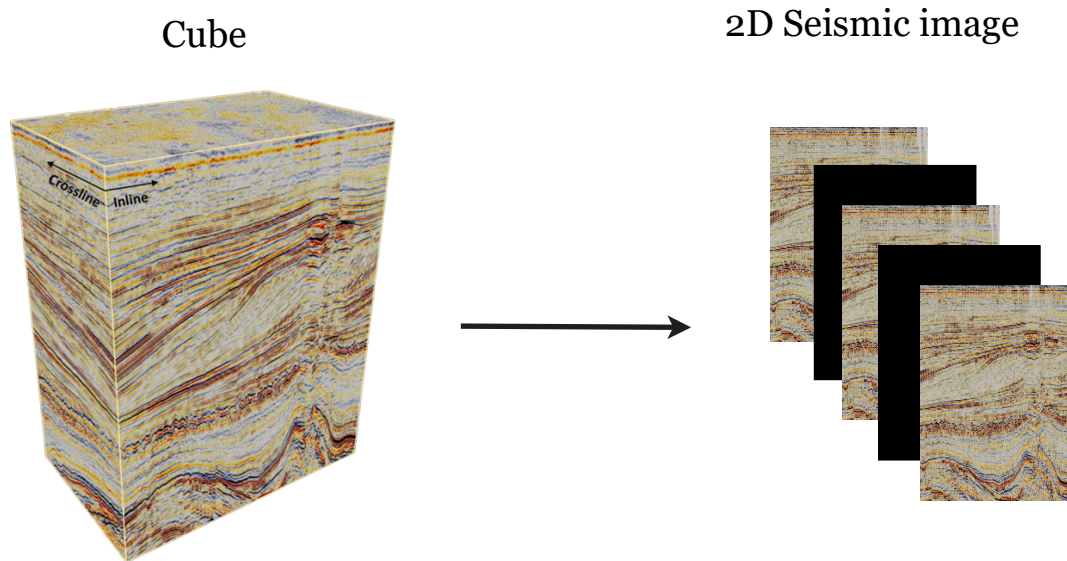
Cube   2D Seismic image

Figure 4: Cube transformation for 2D seismic imaging

timization by PSO per dataset.

After the patch generation step, there is an imbalance between gas and non-gas patches at an average ROI ratio of 1:234 to the Parnaiba Basin and a ratio of 1:531 to the Netherlands F3-block. This sample imbalance can negatively impact the model's performance in learning the correct gas patterns. For this reason, we perform the undersampling technique [14] in patches of the predominant class (non-gas) to exclude some random samples to obtain a 1:4 ratio of gas to non-gas samples in both sources. The 1:4 ratio produced the best results without compromising computational resources.

2.3 Gas detection using Vision Transformer

After data preparation, the seismic patches are classified as gas or non-gas using the ViT model. The ViT architecture employed in this research is illustrated in Figure 5. The architecture consists of an Embedding layer, a stack of Transformer blocks, and a Multilayer Perceptron (MLP). The Embedding layer transforms a 2D image into flattened token sequences, keeping its positional information, to feed the stacked Transformer blocks. A standard Transformer encoder consists of multi-head self-attention layers alternating with MLP blocks. Besides, there is a Layernorm in the block beginning and residual connections at the end of each Transformer block. Finally, an MLP layer is responsible for classifying the samples based on the stacked Transformer blocks output.

ViT has a few variants (ViT-Base, Vit-Large, and ViT-Huge) [13] that differ due to specific hyperparameters: Layers, Hidden Size D, MLP Size, Dropout e Heads. The Layers hyperparameter represents the depth of the network and indicates the number of stacked Transformer encoders. Hidden Size (D) is the dimension that the two-dimensional input samples will be flattened through a linear projection. MLP Size represents the number of neurons in the hidden layer. Dropout is used in MLP to solve the problem of over-adjusting training, and Heads is the number of attention

layers present in the Transformer's encoder.

At last, we use PSO to optimize the ViT model's hyperparameters. We chose PSO because it provides a high-quality solution in less time, has more efficient agility features, and is potentially more efficient than other optimization techniques. [21, 19]. Table 4 and 5 summarize the results of the hyperparameters optimization method.

The model generates a binary classification of gas and non-gas, which we use to reconstruct the final seismic image. In the final image, we aggregate the output values associated with the same coordinate. At end, the generated image is then normalized between 0 and 1. The validation metrics listed below were used to assess the efficacy of the method adopted: accuracy (Acc), sensitivity (Sens), specificity (Spec), and area under roc curve (AUC) [15].

2.4 Parameter optimization using PSO

The PSO algorithm is one of the most well-known metaheuristics. It was first proposed by [16] as a solution methodology for continuous nonlinear problems. The PSO algorithm is an evolutionary technique inspired by the agglomeration and collaboration behavior of biological populations [16]. Since the original introduction in 1995, there have been minor adjustments and refinements to the PSO, but the fundamental principles have remained.

The PSO aims for an optimal solution by iteratively changing the velocities and positions of the particle according to the particle and flight experiences of the group, guiding them to the location of Gbest and Pbest in subsequent iterations. Gbest corresponds to the optimal value of fitness of the population achieved by any particle, while Pbest corresponds to the optimal value of fitness of the particle achieved so far [20]. Figure 6 illustrates the movement of PSO particles.

Also, PSO can generate a high-quality solution within a shorter calculation time and exhibiting more effective stable convergence characteristics than other optimization techniques. Moreover, there are fewer control parameters to adjust, and it is more efficient in maintaining the diversity of
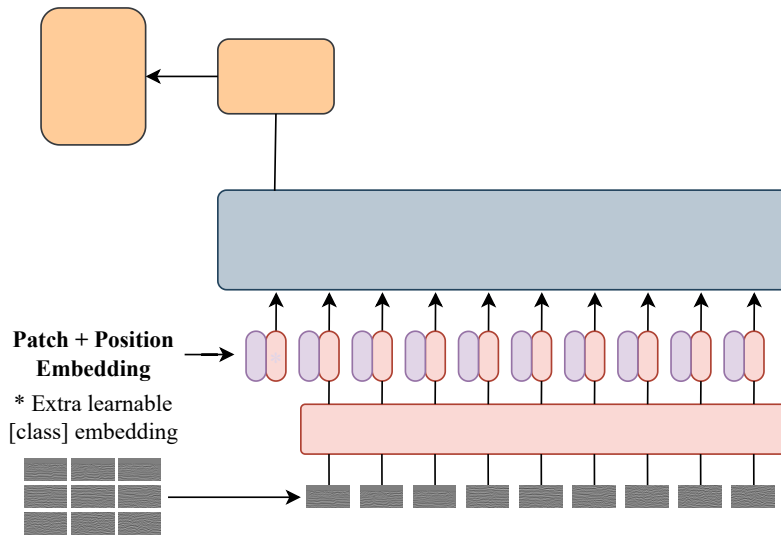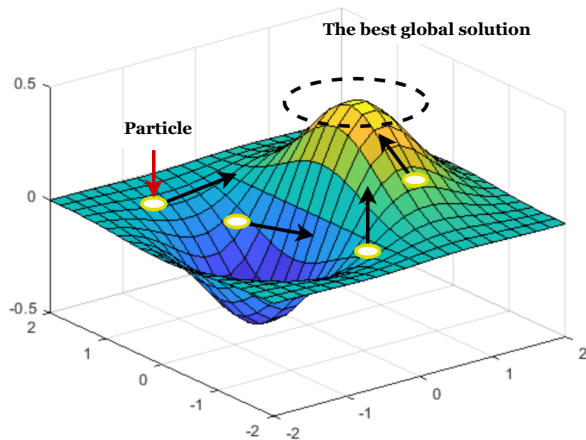
Figure 5: Vision Transformer architecture [13]

Figure 6: PSO Particle Movement

crete threshold between 128 and 1024 was set. Furthermore, to allow the PSO algorithm to evolve, a fitness function was required.

The fitness of each particle was evaluated using the results obtained by the Vision Transformer model, which operated on the validation subset according to the selected parameters. Here, we explain the weighting method used in fitness. The main purpose of weighting is to maintain a balance between sensitivity and precision; for this, we used the F-score [28] to obtain superior models in the detection of potential natural gas accumulations. This fitness is defined as Equation 1.

$$Fitness = F - Score = 2 * \frac{Pre * Rec}{Pre + Rec} \qquad (1)$$

where $Pre = \frac{TP}{TP+FP}$ and $Rec = \frac{TP}{TP+FN}$. True positive (TP) indicates the correctly detected cases. False positive (FP) denotes the negative cases mistakenly detected as positive. True negative (TN) refers to the truly detected negative cases. False negative (FN) denotes the positive cases mistakenly detected as negative.

After defining a particle, its thresholds, and its fitness function, the following steps were performed [21]:

- Step 1: A population of ten random particles was created. Then, the fitness function of each particle was computed.

- Step 2: Each particle traveled circularly through the search space at an initial speed, as established in the previous step. For each iteration, the optimal local particle was sought. The best overall result represents the current best particle. The speeds were also updated during this step.

- Step 3: After the speed was calculated and updated, the particles flew in the search space at this new speed.

- Step 4: The best location and best overall result were updated to the best position, according to the fitness function.

the swarm as all the particles use the information related to the most successful particle (Gbest particle) to enhance themselves. For this reason, we chose the PSO to optimize the Vision Transformer model hyperparameters [20].

The Vision Transformer hyperparameters used for optimization are those described in Subsection 2.3: Hidden Size D, Layers, Heads, MLP Size, Dropout. A five-position vector represents a PSO particle, which each vector position represents one of the hyperparameters mentioned. Each parameter requires a search space, which is simply denoted by the limits of the maximum and minimum values that can be assumed. For the Hidden Size D, a set of values were defined as 128, 256, 512, and 1024. The Layer depth were set as integer values between one and six. Then for the head parameters, discrete thresholds were defined between eight and thirty and six. Afterward, the dropout, a continuous threshold was defined between zero and one. Finally, a dis-

- Step 5: The search stop condition was checked. If the fitness function of the particle was optimal, the search was interrupted. Otherwise, we returned to Step 2.

At the end of this optimization, we obtained the best Vision Transformer parameters of the model validation step. Finally, a test database was applied, and validation metrics were extracted to calculate the method's robustness.

3. RESULTS AND DISCUSSION

In this section, we show the training environment, the result of each step, and the performance of the method in the case study. The proposed method was implemented by using the Python language. We mainly used the Keras deep learning library [6] with tensorflow-gpu [1] as the back-end. Also, we use a python library pyswarm [23] to perform the Particle Swarm Optimization. The computer used in the experiments consists of an Intel Core i7-9700K 4.20 GHz CPU, 24 GB of RAM, and Nvidia GeForce RTX 2070 super graphics card, running on the Windows 11 operating system. The split of the seismic images 2D dataset for the experiments is described in Table 1 and 2.

Table 1: Park of the Hawks seismic images division

Dataset	Train	Validation	Test
Black	15	2	4
Real	14	2	5
White	24	4	7
Red	7	1	2
All	60	9	18

Table 2: F3-block seismic images division

Dataset	Train	Validation	Test
Inline	391	130	130
Crossline	571	190	190
All	962	320	320

We can see that the datasets differ in terms of the quantity of images that can be used for training, validation, and testing. As a result of the large number of training samples, the results may vary depending on the number of representative individuals in each dataset.

Then, the experiments conducted to validate the suggested method are then shown. First, the results are provided independently for each dataset, then according to the use of PSO, and finally, the results are compared to other approaches.

Table 3: PSO Window size optimization hyperparameters

Dataset	With	Height	Step size
Park of the Hawks	20	20	1
F3-Block	64	64	3

3.1 Results per datasets

Following base splitting, the next step is to extract patches from each ROI for each image in each dataset. For this, Fixed window size patches were extracted from each image based on the size indicated by the PSO (Table 3). It is then under sampled to maintain a 1:4 ratio for each gas and non-gas patch (Section 2.2).

After the patches have been extracted, the ViT model is trained to classify them as gas or non-gas. As highlighted in Section 2.3, ViT hyperparameters have been optimized by PSO (Table 4 and 5). Table 6 and 7 describe the results produced by applying the method to the described datasets.

Table 4: Optimized hyperparameters of the ViT model using PSO (Park Of The Hawks)

Hidden Size D	Layers	Heads	MLP Size	Dropout
256	3	20	512	0.125

Table 5: Optimized hyperparameters of the ViT model using PSO (Netherlands F3-Block)

Hidden Size D	Layers	Heads	MLP Size	Dropout
512	4	24	1024	0.25

Table 6: Results per dataset (Park of The Hawks)

Dataset	Sen (%)	Spec (%)	Acc (%)	AUC (%)
Black	37.63	89.29	88.29	63.46
Real	50.13	94.29	93.29	72.21
White	58.02	88.77	88.29	73.39
Red	67.49	96.85	96.20	82.17
All	75.14	96.14	95.60	85.64

Table 7: Results per dataset (Netherlands F3-Block)

Dataset	Sen (%)	Spec (%)	Acc (%)	AUC (%)
Inline	75.03	98.88	98.11	86.95
Crossline	77.44	98.65	97.90	88.05
All	88.60	99.56	99.37	94.08

We noticed that specificity and accuracy scores more than 88% were produced in "all" dataset. The Gavião Black dataset, on the other hand, produced a poor sensitivity in comparison to the others. This conclusion can be attributable to the low and varied quality of seismic data. [27].

However, we observed that when we train "all" of the dataset together in the Park of the Hawks, the results yield the best metrics. This is justified by an increase in data variability, which allows the ViT to better learn the patterns of differentiation across classes, increasing the network's generalization capacity and yielding metrics of 75.14% sensitivity, 96.14% specificity, 95.60% accuracy, and 85.64% AUC. Furthermore, the same phenomenon appears in the F3-Block, where, when joining all datasets, the metrics present higher outcomes, such as 88.60% , 99.37%, 99.56%, and 94.08%, sensitivity, accuracy, specificity, and AUC, respectively.

3.2 Results: Vit with PSO and without

It is worth mentioning that the use of PSO for hyperparameter optimization is helpful because the search for these parameters is time-consuming and subject to large variation given the search space; nevertheless, by utilizing PSO, we can identify them automatically and increase the ViT performance even further. Thus, in order to validate the

PSO efficiency, we give Table 8, which shows the outcomes obtained by ViT with its default hyperparameters and optimized by the PSO across all datasets.

Table 8: Results with and without PSO

Dataset	Sen (%)	Spec (%)	Acc (%)	AUC (%)
Vit	63.81	90.20	89.44	77.01
Vit + PSO	**75.14**	**96.14**	**95.60**	**85.64**

We can see that the use of PSO provided a significant improvement in validation metrics. Where sensitivity has improved by more than 11%, this means that more gas regions are being found. Furthermore, the specificity improves by almost 6%, which shows that with the use of PSO the method produced fewer false positives. Thus, we emphasize that the use of PSO for ViT optimization was essential to produce promising results.

3.3 Comparison with other approaches

In this section, we present a comparison of the results achieved in the 'All dataset' with the work proposed by [27] that uses LSTM and we also trained and tested a LeNet-5 [22] network to validate the effectiveness of ViT in relation to a conventional CNN. The Table 9 and 10 displays the results.

Table 9: Results with other approaches (Park of the Hawks)

Approach	Sen (%)	Spec (%)	Acc (%)	AUC (%)
LSTM	52.99	96.69	93.97	74.84
LeNet-5	30.54	93.61	90.57	62.08
Vit + PSO	**75.14**	**96.14**	**95.60**	**85.64**

Table 10: Results with other approaches (Netherlands F3-Block)

Approach	Sen (%)	Spec (%)	Acc (%)	AUC (%)
LSTM	62.51	98.39	97.42	80.45
LeNet-5	49.54	98.84	96.69	74.37
Vit + PSO	**88.60**	**99.56**	**99.37**	**94.08**

We observed that the proposed method surpasses the other comparatives in relation to sensitivity. We highlight once again the sensitivity metric as being crucial, given the importance of gas detection. Compared to LeNet-5, our method outperforms all validation metrics, showing its generalization power compared to a conventional CNN. On the other hand, the work of [27] presents metric of specificity slightly higher than the proposed method. However, it is worth noting that our method produces greater sensibility and accuracy in the gas class, which demonstrates greater robustness.

3.4 Case study

To evaluate the results achieved in the proposed method, we define four case studies. In the first and second case, the model can detect the gas reservoir effectively. In the third and fourth case, the model presents some deficiencies in gas reservoir detection results.

In Figure 7 and 8, we can see six cases that had good results in detecting gas reservoirs. However, some false positives are generated (in red), the method can distinguish the

aimed region (in blue), which can facilitate the analysis of the data by an expert. Thus, these cases demonstrate that the proposed method is promising for both quantitative and qualitative results. It is worth mentioning that several similar results were found across all datasets.
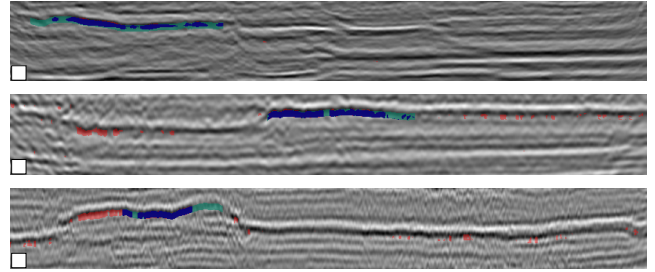

Figure 7: Case Study 1: (a), (b) and (c) represent three different seismic images. In red, it represents false positives. In blue, the true positives. In green, false negatives. Park of the Hawks dataset
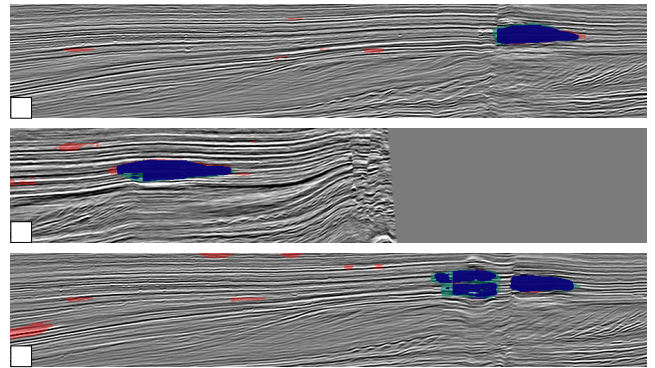

Figure 8: Case Study 2: (a), (b) and (c) represent three different seismic images. In red, it represents false positives. In blue, the true positives. In green, false negatives. Netherlands F3-Block dataset

The third and fourth case study are illustrated in Figure 9 and 10. Then, in these cases, we can assume that the proposed method cannot detect some potential gas reservoirs. Although the model identified some of the targeted regions, the majority of the gas reservoir regions were not detected (in green). Furthermore, ViT confuses the gas prediction with similar locations, resulting in false-positive predictions.

It is important to note that data analysis is not an easy task. As a result, it necessitates professional knowledge and is time-consuming. As a result, we believe that the proposal, when combined with the expert's data analysis knowledge, can be a better technique of discovering potential gas reservoirs.

3.5 Comparison with related works

After the presentation of the proposed method results, a comparative analysis of the these results achieved with the results of the related works was carried out. For a more rigorous comparison, the results obtained in the proposed method were compared with related works that seek the same objective, the detection of potential gas accumulations. In Table 11, we present information on the tech-
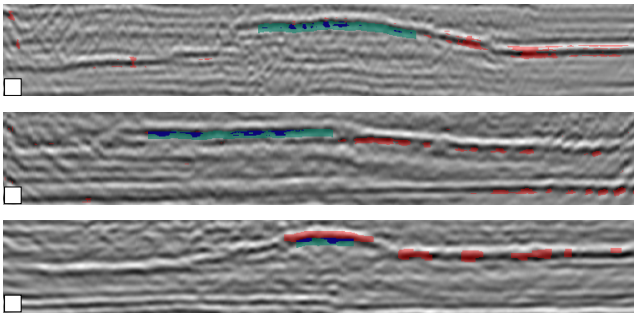
Figure 9: Case Study 3: (a), (b) and (c) represent three different seismic images. In red, it represents false positives. In blue, the true positives. In green, false negatives. Park of the Hawks dataset
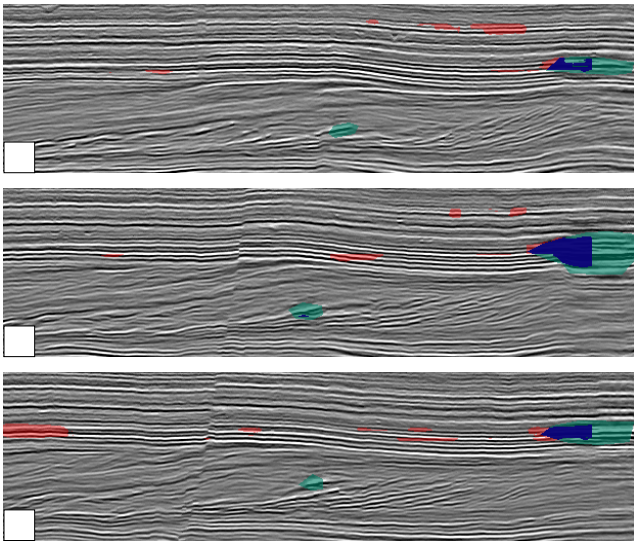

Figure 10: Case Study 4: (a), (b) and (c) represent three different seismic images. In red, it represents false positives. In blue, the true positives. In green, false negatives. Netherlands F3-Block dataset

niques, datasets, and a summary of the results found in related works and the proposed method.

3.6 Advances and limitations of the method

Because it proposes a new and automated method to the detection of potential natural gas accumulations, our method offers a series of merits and advances, of which we highlight the main ones as follows:

1. It offers an automated method, developed using two different sources, public databases (Subsection 2.1). The diversity of the databases simulates the real specialist context, and their publicity makes the method amenable to comparison.

2. Because the reflection sesmic images databases differ and are not standardized in either the data-acquisition or examination processes, the images present numerous patterns. Thus, our proposed method implements a crucial stage of image preprocessing; despite its simplicity, this is necessary for the successful execution of the entire method.

3. To the best of our knowledge, this is the first method to use Vision Transformer combined with a PSO optimization algorithm to detect potential natural gas accumulations.

4. Automated methods have been studied by many researchers; however, these methods always encounter a parameterization barrier. This is no different for the Vision Transformer model, which features a range of parameters. Thus, we proposed a PSO to automatically optimize these parameters and bypass the parameter selection step.

5. All of these steps increase the method's utility. The present study obtained results comparable to those found in the literature and proposes an innovative new method to detect potential natural gas accumulations in reflection seismic images.

6. The proposed method, using Vision Transformer and PSO, achieved a maximum a sensitivity of 88.60%, a specificity of 99.56% and an accuracy of 99.37%. These results indicate the method's potential.

However, as with any computational method, it has some limitations. We highlight these as follows:

1. Our method did not propose a new deep architecture. We used the existing architectures and demonstrated their effectiveness for capturing deep features. We believed that developing a new architecture will further improve results.

2. The present method indicates the region that presents gas or no t in a reflection seismic image. However, the proposed method need a specialist assistence to define the Region of Interest. Adding a segmentation step to identify the Region of Interest could further help gas exploration professionals.

4. CONCLUSION

In this work a method for detecting gas using seismic data has been proposed. A ViT Transformer network optimized by PSO was presented for this purpose. The proposed method used a two-dimensional approach with an architecture based on attention processes, which was originally developed to solve difficulties in natural language processing but has since been adapted to other domains such as image processing. The proposed method comprises of improving and adjusting the settings of the ViT Transformer using an evolutionary algorithm to better adapt it to pattern detection in seismic images.

The proposed method produces promising outcomes. The approach was effective for gas detection, with sensitivity comparable to other models investigated in the literature. The use of PSO to optimize ViT hyperparameters was also a significant step, resulting in an improvement in all validation measures. As a result, it is thought that the proposed method, when combined with professional practice, can be significant for gas detection.

As future work, we recommend testing the method on more datasets, as the model may not have reached its full generality given the limitations and variability of the data.

Table 11: Comparison with related works

Method	Seismic Database	Sen (%)	Spec (%)	Acc (%)	AUC (%)
MLP-Mixer and U-Net [7]	Netherlands F3-Block	86.85	99.79	99.6	93.27
Time2Vec Transformer [11]	Park of the Hawks	96.12	98.92	98.87	97.52
LSTM [26]	Park of the Hawks	90.28	97.67	97.62	96.32
LSTM [27]	Netherlands F3-Block	97.1	96.83	97.1	97.67
Proposed method (Vit + PSO)	**Netherlands F3-Block**	**88.60**	**99.56**	**99.37**	**94.08**
Proposed method (Vit + PSO)	**Park of the Hawks**	**75.14**	**96.14**	**95.60**	**85.64**

Besides, we suggest creating a method for determining the ROI to transform the method into a fully automated approach. Also, an adaptation of ViT so that it works with a semantic segmentation network to improve the performance of the results already achieved by ViT-Seismic. Finally, another possible improvement would be to combine the 1D information with the 2D information achieved by ViT.

5. ACKNOWLEDGMENTS

6. ADDITIONAL AUTHORS

Additional authors: Roberto Quispe (PUC - Rio de Janeiro, Rio de Janeiro, RJ, Brazil, email: `rquevedo@tecgraf.puc-rio.br`); Roberto Ribeiro (ENEVA S.A., Rio de Janeiro, RJ, Brazil, email: `roberto.ribeiro@eneva.com.br`) and Vinicius Riguete (vinicius.riguete@eneva.com.br, email: `ENEVA S.A., Rio de Janeiro, RJ, Brazil`).

7. REFERENCES

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org*, 2015.

[2] Y. Alaudah, P. Michałowicz, M. Alfarraj, and G. AlRegib. A machine-learning benchmark for facies classification. *Interpretation*, 7(3):SE175–SE187, 2019.

[3] F. d. Almeida, C. D. R. Carneiro, et al. Inundações marinhas fanerozóicas no brasil e recursos minerais associados. *Mantesso Neto, V.; Bartorelli, A.; Carneiro, CDR*, pages 43–60, 2004.

[4] F. Andrade, L. Fernando Santos, M. Gattass, R. Quevedo, D. Michelon, C. Siedschlag, and R. Ribeiro. Gas reservoir segmentation in 2d onshore seismics using lstm-autoencoder. In *First International Meeting for Applied Geoscience & Energy*, pages 1651–1655. Society of Exploration Geophysicists, 2021.

[5] D. S. Chevitarese, D. Szwarcman, R. G. e Silva, and E. V. Brazil. Deep learning applied to seismic facies classification: A methodology for training. In *Saint Petersburg 2018*, volume 2018, pages 1–5. European Association of Geoscientists & Engineers, 2018.

[6] F. Chollet et al. Keras. https://keras.io, 2015.

[7] C. L. S. Cipriano, D. A. D. Júnior, P. S. Diniz, L. F. Marin, A. C. de Paiva, J. O. B. Diniz, and A. C. Silva. Detection and delimitation of natural gas in seismic images using mlp-mixer and u-net. In J. Filipe, M. Smialek, A. Brodsky, and S. Hammoudi, editors, *Proceedings of the 24th International Conference on Enterprise Information Systems, ICEIS 2022, Online Streaming, April 25-27, 2022, Volume 1*, pages 578–585. SCITEPRESS, 2022.

[8] F. S. de Miranda, A. L. Vettorazzi, P. R. da Cruz Cunha, F. B. Aragão, D. Michelon, J. L. Caldeira, E. Porsche, C. Martins, R. B. Ribeiro, A. F. Vilela, et al. Atypical igneous-sedimentary petroleum systems of the parnaíba basin, brazil: seismic, well logs and cores. *Geological Society, London, Special Publications*, 472(1):341–360, 2018.

[9] H. Di, Z. Wang, and G. AlRegib. Seismic fault detection from post-stack amplitude by convolutional neural networks. In *80th EAGE Conference and Exhibition 2018*, volume 2018, pages 1–5. European Association of Geoscientists & Engineers, 2018.

[10] D. Dias, P. Diniz, L. Marin, C. Cipriano, M. Gattass, L. Santos, R. Quevedo, D. Michelon, C. Siedschlag, and R. Ribeiro. Automatic gas detection in land seismic using transformer neural network. In *17th International Congress of the Brazilian Geophysical Society*. Brazilian Geophysical Society, 2021.

[11] P. Diniz, D. A. D. Junior, J. O. Diniz, A. C. de Paiva, A. C. d. Silva, M. Gattass, R. Quevedo, D. Michelon, C. Siedschlag, and R. Ribeiro. Time2vec transformer: a time series approach for gas detection in seismic data. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pages 66–72, 2022.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020.

[14] C. Drumnond and R. Holte. Class imbalance and cost sensitivity: Why undersampling beats oversampling.

In *ICML-KDD 2003 Workshop: Learning from Imbalanced Datasets*, volume 3, 2003.

[15] R. Duda. *Pattern classification and scene analysis.* Wiley-Interscience Publication, 512, 1973.

[16] R. Eberhart and J. Kennedy. Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, volume 4, pages 1942–1948. Citeseer, 1995.

[17] A. Guitton, H. Wang, and W. Trainor-Guitton. Statistical imaging of faults in 3d seismic volumes using a machine learning approach. In *SEG Technical Program Expanded Abstracts 2017*, pages 2045–2049. Society of Exploration Geophysicists, 2017.

[18] E. Illidge, J. Camargo, and J. Pinto. Turbidites characterization from seismic stratigraphy analysis: Application to the netherlands offshore f3 block. In *Proceedings of the AAPG/SEG 2016 International Conference & Exhibition, Cancun, Mexico*, pages 6–9, 2016.

[19] D. A. D. Júnior, L. B. da Cruz, J. O. B. Diniz, G. L. F. da Silva, G. B. Junior, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass. Automatic method for classifying covid-19 patients based on chest x-ray images, using deep features and pso-optimized xgboost. *Expert Systems with Applications*, 183:115452, 2021.

[20] J. Kennedy. Particle swarm optimization. *Encyclopedia of machine learning*, pages 760–766, 2010.

[21] L. T. Le, H. Nguyen, J. Zhou, J. Dou, H. Moayedi, et al. Estimating the heating load of buildings for smart city planning using a novel artificial intelligence technique pso-xgboost. *Applied Sciences*, 9(13):2714, 2019.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[23] L. J. V. Miranda. Pyswarms, a research-toolkit for particle swarm optimization in python. 2017.

[24] A. Pochet, P. H. Diniz, H. Lopes, and M. Gattass. Seismic fault detection using convolutional neural networks trained on synthetic poststacked amplitude maps. *IEEE Geoscience and Remote Sensing Letters*, 16(3):352–356, 2018.

[25] L. Santos, F. Jordao, M. Gattass, R. Quevedo, M. J. Lima, D. Michelon, C. Siedschlag, R. Ribeiro, and S. Pereira. Natural gas detection in onshore data using transfer learning from a lstm pre-trained with offshore data. Society of Exploration Geophysicists, 2021.

[26] L. F. Santos, M. Gattass, A. Silva, F. Miranda, C. Siedschlag, and R. Ribeiro. Natural gas detection in onshore data using transfer learning from a lstm pre-trained with offshore data. In *SEG Technical Program Expanded Abstracts 2020*, pages 1190–1195. Society of Exploration Geophysicists, 2020.

[27] L. F. T. Santos. Detector de assinaturas de gás em levantamentos sísmicos utilizando lstm. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro, 2019.

[28] Y. Sasaki and R. Fellow. The truth of the f-measure, manchester: Mib-school of computer science. *University of Manchester*, 2007.

[29] L. Septian and A. Maulana. Geological model of reservoir based on seismic attributes and ant tracking case study-f3 block, offshore netherlands. In *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, pages cp–348. European Association of Geoscientists & Engineers, 2013.

[30] R. M. Silva, L. Baroni, R. S. Ferreira, D. Civitarese, D. Szwarcman, and E. V. Brazil. Netherlands dataset: A new public dataset for machine learning in seismic interpretation. *arXiv preprint arXiv:1904.00770*, 2019.

[31] K. M. Tingdahl and M. De Rooij. Semi-automatic detection of faults in 3d seismic data. *Geophysical prospecting*, 53(4):533–542, 2005.