

APLICAÇÃO DE ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS PARA CLASSIFICAÇÃO MINERALÓGICA UTILIZANDO ESPECTROS DE RAMAN*

Caio Trindade de Almeida¹
Adilson Ribeiro Prado²
Cassius Zanetti Resende³
Daniel Cruz Cavalieri⁴

Resumo

Métodos de aprendizado de máquina têm encontrado muitas aplicações na espectroscopia Raman, especialmente para a identificação de espécies químicas. Este trabalho apresenta uma abordagem de reconhecimento de padrões para a classificação de minerais usando espectros Raman, focando nos seis principais componentes do minério de ferro: dolomita, goethita, hematita, magnetita, quartzo e siderita. Especificamente, foram utilizados os classificadores baseados em SVM (Support Vector Machines), LDA (Linear Discriminant Analysis), kNongs mais próximos - kNN) e Autoencodificadores de Redes Neurais. Além disso, o tratamento do banco de dados foi realizado, principalmente, na correção da linha de base do espectro e no balanceamento das amostras através da geração de dados com ruídos, simulando os dados coletados durante um processo real. Nossa abordagem foi avaliada usando o banco de dados espectral RRUFF, compreendendo dados de amostras minerais. Como esperado, o desempenho de classificação apresentado pelo classificador Autoencoder (99,17%) foi superior ao encontrado por outros algoritmos de aprendizado de máquina frequentemente utilizados.

Palavras-chave: Espectros de Raman; Técnicas de Aprendizado de Máquinas; Classificação de Minerais.

APPLICATION OF MACHINES LEARNING ALGORITHMS FOR MINERALOGICAL CLASSIFICATION USING RAMAN SPECTRUM

Abstract

Machine learning methods have found many applications in Raman spectroscopy, especially for the identification of chemical species. This work presents a standard recognition approach for the classification of minerals using Raman spectra, focusing on the six main components of iron ore: domolite, goethite, hematite, magnetite, quartz and siderite. Specifically, the classifiers based on Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), k-Nearest Neighbors - kNN) and Neural Networks Autoencoders were used. In addition, the treatment of the database was performed, mainly in the correction of the baseline of the spectrum and the balancing of the samples through the generation of data with noises, simulating data collected in the process. We evaluated our approach using the RRUFF spectral database, comprising mineral sample data. As expected, the classification performance presented by the Autoencoder classifier (99.17%) was superior to that found by other frequently used machine learning algorithms.

Keywords: Raman Spectrum; Machine Learning Techniques; Minerals Classification.

¹ Engenheiro Químico, Programa de Pós-Graduação em Engenharia de Controle e Automação, Instituto Federal do Espírito Santo, Serra, Espírito Santo, Brasil.

² Químico, Dr, Professor EBTT, Engenharia de Controle e Automação, Instituto Federal do Espírito Santo, Serra, ES Brasil

³ Engenheiro Eletricista, Dr, Professor EBTT, Engenharia de Controle e Automação, Instituto Federal do Espírito Santo, Serra, ES Brasil

⁴ Engenheiro Eletricista, Dr, Professor EBTT, Engenharia de Controle e Automação, Instituto Federal do Espírito Santo, Serra, ES Brasil.

1 INTRODUÇÃO

Há anos, a espectroscopia vem sendo utilizada em diversos segmentos da indústria como, por exemplo, na produção de bauxita, com o objetivo de explorar as diferenças na composição química e na estrutura do material [1], na indústria de fármacos para controle de qualidade e validação da pureza dos medicamentos [2], na indústria agropecuária para caracterização da qualidade do solo, na medicina em processos de investigação não invasiva de componentes moleculares [3], entre outras aplicações.

De acordo com os trabalhos desenvolvidos por [4] e [5], a crescente aplicação desta técnica possui as seguintes vantagens: análises não destrutivas, ou seja, não requer aplicação de produtos químicos na amostra; equipamentos compactos de fácil transporte e manuseio; alta taxa de amostragem e facilidade de realização de ensaios *in situ* em materiais não preparados.

Embora vários estudos já tenham demonstrado o potencial da espectroscopia em diversas áreas, poucos estudos foram realizados para caracterização do minério de ferro para elaboração de um modelo global que possa ser aplicado na predição da mineralogia. Dentro deste contexto, a espectroscopia Raman fornece informações complementares aos espectros infravermelhos, sendo relevante para a caracterização mineralógica *in situ* do minério de ferro. Segundo [5] os espectros gerados por Raman são representativos para os principais minerais de óxido de ferro de interesse econômico e das fases, pois apresentam natureza distinta do espectro e permitem identificação e quantificação dos principais óxidos de ferro e outros minerais.

A quantidade de informação contida em um espectro é muito grande, o que gera o desafio para as empresas da aplicação em processos industriais de forma online. Este desafio vem sendo estudado e superado com os avanços tecnológicos provenientes da 4ª Revolução Industrial. A indústria 4.0 busca promover a automação de todos os processos de manufatura aumentando consequentemente a produtividade das linhas de produção e a competitividade por meios de fábricas inteligentes [6].

Na Indústria 4.0 produtos e sistemas de produção estão conectados a redes globais de produção por meio da internet das coisas (IOT) e sistemas ciber-físicos (CPS). Estas redes inteligentes de produção são capazes de se comunicar umas com as outras, interpretar dados coletados e desencadear ações automáticas em tempo real para otimização e controle do processo [7]. Assim, com o avanço do aprendizado de máquinas a possibilidade de aplicações industriais para a espectrometria de Raman de forma automática também vem crescendo de forma que as técnicas de classificação têm sido estudadas por serem essenciais na obtenção da resposta em tempo real de uma amostragem espectral [8].

Uma dificuldade encontrada por [8] no processamento do espectro natural é a interferência da linha de base do sinal natural causado principalmente pela fluorescência podendo ser mais intensa que a dispersão do efeito Raman, o que afeta negativamente o desempenho dos classificadores. Neste aspecto, há um grande esforço nessa área para correção da linha de base, mas continua um desafio para os pesquisadores, especialmente para confecção de um sistema automático. Uma revisão detalhada e comparação dos métodos de correção da linha de base pode ser encontrada em [9].

Neste contexto, este trabalho apresenta uma abordagem de reconhecimento de padrões para a classificação de minerais quanto sua estrutura química obtida por

meio de espectros de Raman conforme proposto por [8], com foco nos 6 componentes principais do minério de ferro: domolita, goethita, hematita, magnetita, quartzo e siderita. Especificamente, foram utilizados os classificadores baseados em Máquinas de Vetores Suporte (do inglês, *Support Vector Machines* - SVM), Análise de Discriminante Linear (do inglês, *Linear Discriminant Analysis* - LDA), *k*-Vizinhos Mais Próximos (do inglês *k-Nearest Neighbors* - kNN) e Redes Neurais Autoencoders. Além disso, é tratado também a importância do tratamento da base de dados, principalmente na correção da linha de base do espectro e do balanceamento das amostras através da geração de dados com ruídos, simulando dados coletados no processo.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta o banco de dados dos espectros de Raman utilizados no estudo, como ele foi adquirido, processado e os detalhes dos algoritmos dos classificadores empregados neste estudo. Na Seção 3 apresentamos os resultados experimentais comparando os algoritmos utilizados neste trabalho. Finalmente, a Seção 4 concluímos o artigo, identificando o classificador que apresentou melhor desempenho e fornecemos algumas propostas de complementação para futuras pesquisas.

2 DESENVOLVIMENTO

O sistema proposto pode ser organizado em várias partes interoperáveis, conforme ilustrado na Figura 1 e explicado nas subseções seguintes.

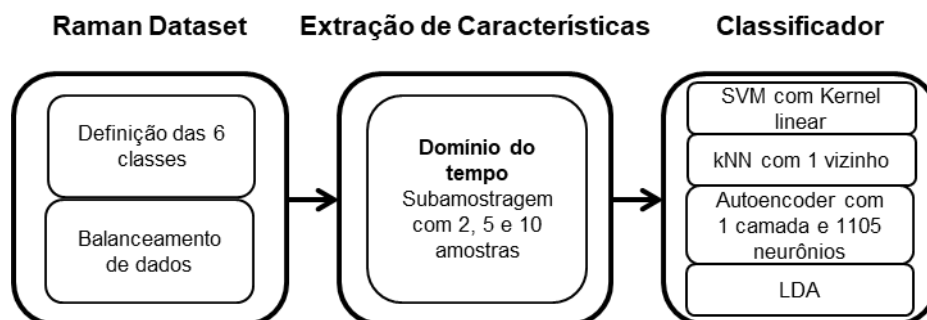


Figura 1: Arquitetura do sistema

2.1 Base de Dados Raman

Neste artigo, a base de dados utilizada foi a mesma empregada por [8]. Este conjunto de dados denominado Projeto RRuff é um conjunto completo de dados espectrais de alta qualidade a partir de minerais bem caracterizados que fornecem um padrão para mineralogistas, geocientistas, gemologistas e o público em geral para a identificação e caracterização de minerais. Como essa base de dados contém 9433 espectros de minerais, definiu-se neste estudo apenas os seis minerais com maior relevância e predominância na composição do minério de ferro e que apresentam forte interação no espectro de Raman [5].

Diante disso, os seis minerais de maior relevância adotados neste estudo foram: dolomita, goethita, hematita, magnetita, quartzo e siderita. É importante destacar que esses minerais encontram-se de forma desbalanceada na base de dados, ou seja, do total de 64 espectros têm-se 16 espectros para dolomita, 4 para goethita, 10 para hematita, 8 para magnetita, 20 para o quartzo e 6 para a siderita e, de acordo com [10] e [11], para que o classificador não fique polarizado, fez-se necessário a criação

de novos espectros para as classes escolhidas de forma a balancear a base de dados.

Desta forma, padronizou-se 20 espectros para cada classe de mineral, sendo gerados novos espectros a partir dos sinais naturais da base de dados incluindo ruídos gaussianos por meio da função “awgn” do software Matlab [12] fixando o *signal-to-noise ratio* (SNR) em 1 e variando o *Sigpower* de 7 a 60 dBW.

Outro fator importante dos espectros gerados e obtidos da base de dados é a presença de um componente denominado linha de base de um espectro de Raman [8] e [9]. Este fenômeno é causado principalmente por fluorescência que pode ser mais intenso do que o Raman em várias ordens de magnitude e interfere no desempenho dos sistemas de aprendizado de máquina. Apesar do considerável esforço nessa área, a correção de linha de base é considerada como um problema desafiador, especialmente para desenvolvimento de um sistema de classificação totalmente automático. Neste sentido, o objetivo de qualquer técnica de subtração de fluorescência é a remoção de todas as contribuições deste fenômeno, sem efeito sobre os espectros de Raman. Desta forma, realizou-se a correção da linha de base como descrito por [13] utilizando a função “msbackadj” do software Matlab [14] conforme pode ser observado na Figura 2.

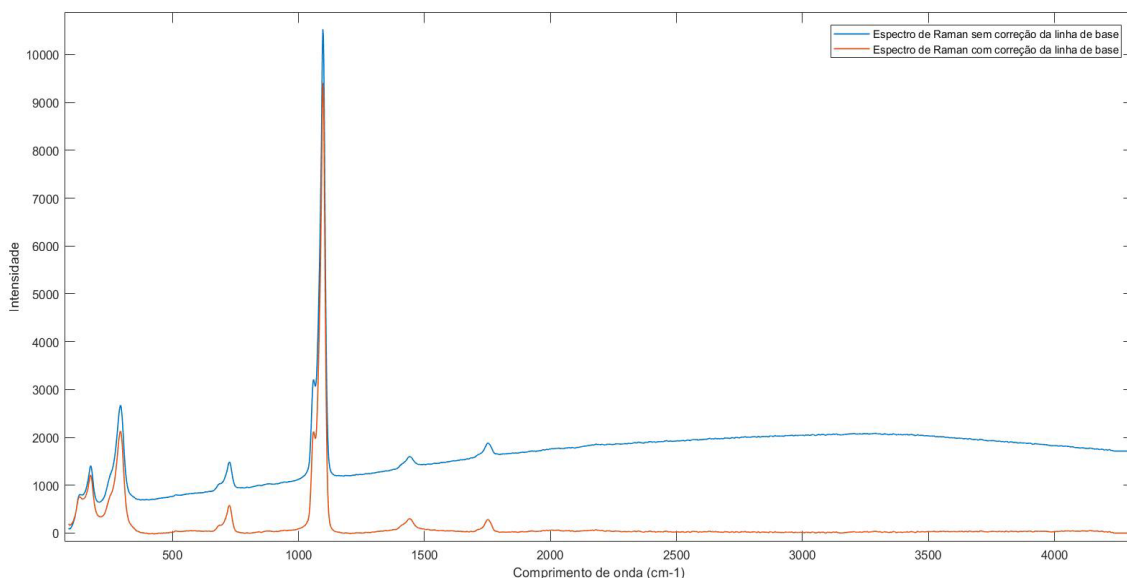


Figura 2: Espectro de Raman para a goethita gerado a partir da base de dados *versus* o espectro correspondente com a correção da linha de base aplicando a função “msbackadj”

2.2 Extração de Características

A escolha de um bom subconjunto de dados é importante em qualquer processo de classificação por várias razões. Se o conjunto de recursos escolhido não incluir todas as informações necessárias para discriminar amostras pertencentes às diferentes classes o desempenho do classificador pode ser insatisfatório independentemente da eficácia do algoritmo de aprendizagem [15]. Por outro lado, o tamanho do conjunto de recursos usados para caracterizar os objetos determina o espaço de pesquisa a ser explorado durante a fase de aprendizado do classificador. Sendo assim, dados com informações irrelevantes tornam o espaço de varredura do classificador maior, aumentando a complexidade do processo [15]. Por fim, o custo computacional da classificação depende do número de recursos usados para descrever os padrões. Assim, neste artigo realizou-se três subamostragens para

cada classe, sendo analisado todo o comprimento do espectro de Raman a cada duas amostras (*sample 2*), a cada cinco amostras (*sample 5*) e a cada dez amostras (*sample 10*), analisando-se a influência desta técnica na acurácia média para cada classificador.

2.3 Classificadores

Como proposto por [8], a ideia inicial foi verificar qual técnica de reconhecimento de padrões aumenta o desempenho de um sistema de classificação para espectros minerais. No entanto, assim como a pesquisa realizada por [8] que fez uma comparação entre os classificadores kNN, *Gradient Boosting*, *Random Florest*, SVM Linear e RBF, Correlação e CNN (*Convolutional Neural Network*), procuramos encontrar o melhor algoritmo para o problema proposto.

Desta forma, neste trabalho foram adotadas as seguintes metodologias: SVM com *kernel* linear, kNN com 1 vizinho, LDA e Rede Neural Autoencoder para lidar com o processo de classificação. Uma breve explicação desses classificadores é apresentado a seguir. É importante mencionar que todos os algoritmos foram implementados utilizando-se o software MATLAB e os resultados serão apresentados na Seção 3.

2.3.1 Rede Neural Autoencoder. Uma rede neural de autoencoder, também chamada de auto-associador, é um algoritmo de aprendizado não supervisionado que aplica a retropropagação definindo os valores alvo como iguais às entradas. Um problema com essa abordagem é que, se não houver outra restrição, um autoencoder com entrada n -dimensional poderia apenas aprender a função identidade, ou seja, estaria apenas copiando a entrada [16]. Surpreendentemente, quando o número de unidades ocultas é grande (ou maior que o número de atributos de entrada), ainda é possível descobrir estruturas interessantes, impondo outras restrições à rede. Em particular, se impusermos uma restrição de dispersão nas unidades ocultas, o autoencoder ainda descobrirá uma estrutura interessante nos dados. Como proposto por [17], para conseguir isso, um termo de penalidade extra próximo de zero ($= 0,05$) é adicionado ao objetivo de otimização, ou seja, deseja-se restringir os neurônios para ficarem inativos na maior parte do tempo e serem ativados somente quando necessário. Além disso, experimentos relatados em [17] sugerem que, na prática, quando treinados com gradiente descendente estocástico, autoencoders não lineares com um número maior de unidades (neurônios) ocultas do que as entradas (chamadas de supercompleto) produzem representações úteis.

2.3.2 Support Vector Machines. Resumidamente, esse algoritmo classifica os dados encontrando o limite de decisão linear (ou não-linear), também conhecido como hiperplano, que separa todos os pontos de dados de uma classe daqueles da outra classe. Além disso, as SVMs podem utilizar uma função *kernel* para transformar dados não linearmente separáveis, onde um limite de decisão linear pode ser encontrado, fornecendo um bom desempenho de generalização para a classificação [18]. Neste trabalho foram realizados testes com os *kernels* linear e função de base radial, sendo que os resultados obtidos por este último são piores do que os resultados obtidos quando se utiliza o *kernel* linear, que foi escolhido para compor o classificador SVM. Além disso, os melhores resultados foram obtidos quando padronizamos as entradas. Todos os outros parâmetros foram deixados como padrão.

2.3.3 k-Nearest Neighbor. Resumidamente, o kNN categoriza objetos com base nas classes de seus vizinhos mais próximos no conjunto de dados. Em outras palavras, as predições realizadas pelo kNN assumem que objetos próximos uns dos outros são semelhantes. Neste caso, métricas de distância como a Euclidiana, o *city block*, o cosseno e o Chebychev são usadas para encontrar o vizinho mais próximo [19]. Neste trabalho, com alguns testes empíricos, os melhores resultados foram encontrados usando um vizinho mais próximo e a distância euclidiana.

2.3.4 Linear Discriminant Analysis. A análise discriminante classifica os dados encontrando combinações lineares de recursos. Esse tipo de algoritmo assume que classes diferentes geram dados baseados em distribuições gaussianas. A análise discriminante linear (LDA) é também conhecida como o discriminante de Fisher, nomeado por seu inventor [20]. Uma vantagem em se utilizar um classificador linear é que ele não requer o ajuste de nenhum parâmetro, então podemos usá-lo como um algoritmo de *benchmark*.

3 RESULTADOS EXPERIMENTAIS

Como já dito anteriormente, o desempenho dos algoritmos propostos foi avaliado utilizando-se um conjunto de seis classes de minerais, onde foi realizado o tratamento da base de dados balanceando-se as classes determinadas e corrigindo-se o efeito da linha de base. Além disso, adotou-se como extração de característica três tipos de amostragens diferentes conforme descrito na Seção 2.2.

Na fase de classificação foram utilizados um SVM com kernel linear, parâmetros padrão e entradas padronizadas, um kNN com distância euclidiana e $k = 1$, um LDA e um Autoencoder com uma camada oculta, 1105 neurônios e uma função de transferência log-sigmóide. A rede neural autoencoder foi treinada com o uso de gradiente descendente estocástico com 100 épocas, *L2-weight regularization*, *sparsity regularization* e *sparsity proportion* de 0,004, 8 e 0,15, respectivamente. Outros parâmetros foram definidos empiricamente à medida que produziam os melhores resultados. Todos os resultados experimentais foram gerados usando o MATLAB © rodando em um notebook i7-7700HQ com NVIDIA GeForce GTX 1050TI, 16 Gb de RAM e um SSD de 256 Gb.

Devido ao fato da base de dados conter apenas 20 espectros por classe foi utilizada a técnica de validação cruzada *leave-one-out* para dividir o conjunto de dados em conjuntos de treinamento e teste. Especificamente, seleciona-se aleatoriamente um espectro para compor a base de teste e utiliza-se o restante para o treinamento do classificador. Essa técnica é comumente empregada em situações onde há poucos dados disponíveis, entretanto, demanda uma alta taxa de processamento computacional.

A Tabela 1 apresenta a acurácia global para cada classificador proposto de acordo com cada subamostragem escolhida.

Tabela 1. Desempenho dos classificadores para acurácia global para cada subamostragem proposta

Leave-one-out	Subamostra	Autoencoder	kNN	SVM	LDA
Acurácia	2	97,50 ± 2,8339	97,50 ± 2,8339	88,33 ± 5,8271	63,33 ± 8,7471
Acurácia	5	97,50 ± 2,8339	98,33 ± 2,3237	85,83 ± 6,3296	57,50 ± 8,9731
Acurácia	10	99,17 ± 1,6501	96,67 ± 3,2583	85,83 ± 6,3296	51,67 ± 9,0707

É interessante destacar que também foi avaliado o tempo de treinamento para cada subamostragem. Assim, nota-se na Tabela 1 que na primeira subamostragem o tempo de processamento (teste/treinamento) foi de 4,45 horas, na segunda e na terceira o tempo foi respectivamente 1,77 e 0,94 horas.

É possível observar também da Tabela 1 que a acurácia global para o classificador autoencoder e kNN ficaram muito próximos sendo que com uma subamostragem de 2 amostras os classificadores apresentaram o mesmo resultado. Especificamente, o classificador Autoencoder alcançou $99,17\% \pm 1,6501$ (melhor resultado) contra $51,67\% \pm 9,0707$ (pior resultado) do classificador de LDA. Esses resultados confirmam nossa hipótese de que um desempenho uniformemente bom é obtido por um classificador mais complexo, em termos da capacidade de distinguir entre padrões, à medida que a complexidade dos dados de treinamento/teste também aumenta. Assim, o melhor classificador foi determinado analisando-se a acurácia global *versus* tempo de processamento que, neste caso, foi o autoencoder com uma subamostragem de 10 amostras.

A partir da definição anterior, foram elaboradas as matrizes de confusão apresentadas na Figura 3.

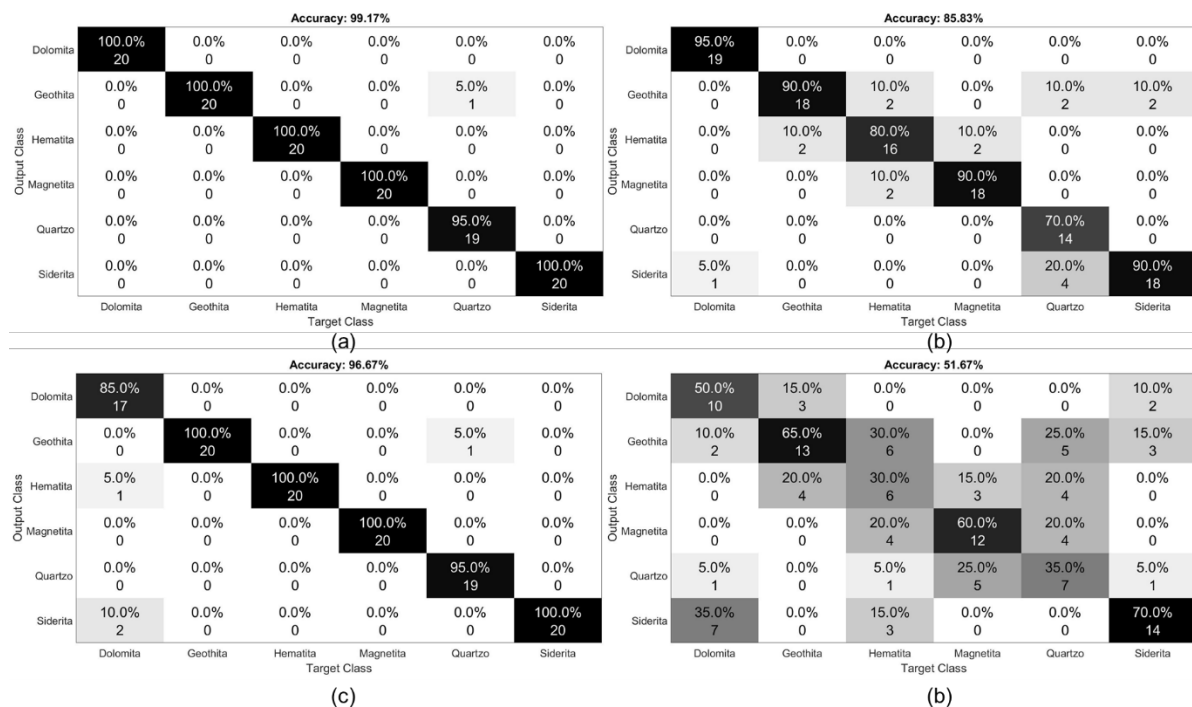


Figura 3: Matrizes de confusão para (a) Autoencoder, (b) SVM, (c) kNN e (d) LDA. Cada matriz é a média das matrizes de confusão utilizando-se o *leave-one-out* e subamostragem de 10

Nota-se novamente na Figura 3 que o autoencoder apresentou os melhores resultados seguido do classificador kNN. Além disso, como esperado, um classificador mais simples (LDA) não apresentou bons resultados, isso deve-se principalmente pela natureza não linear do espectro dos minerais. É interessante notar também na Figura 3 que os classificadores autoencoder e kNN foram capazes de classificar corretamente inclusive classes semelhantes, dentro do contexto da composição química, como a magnetita e a hematita.

A partir das matrizes de confusão apresentadas na Figura 3 procurou-se analisar com mais detalhes cada classificador. Para isso, gerou-se a Tabela 2 que apresenta a precisão e a sensibilidade de cada classificador. De maneira geral, pode-se definir precisão como a capacidade do classificador fornecer os mesmos resultados quando repetido, e sensibilidade como a capacidade do classificador identificar corretamente as classes corretas.

Tabela 2. Desempenho dos classificadores utilizando-se o *leave-one-out* para cada classe de mineral proposta

<i>Leave-one-out</i>	Análise Estatística	Classes					
		Dolomita	Goethita	Hematita	Magnetita	Quartzo	Siderita
Autoencoder	Precisão	100,00	95,24	100,00	100,00	100,00	100,00
	Sensibilidade	100,00	100,00	100,00	100,00	95,00	100,00
kNN	Precisão	100,00	95,24	95,24	100,00	100,00	90,91
	Sensibilidade	85,00	100,00	100,00	100,00	95,00	100,00
SVM	Precisão	100,00	75,00	80,00	90,00	100,00	78,26
	Sensibilidade	95,00	90,00	80,00	90,00	70,00	90,00
LDA	Precisão	66,67	44,83	35,29	60,00	46,67	58,33
	Sensibilidade	50,00	65,00	30,00	60,00	35,00	70,00

É possível notar na Tabela 2 que, como esperado, os piores valores de precisão e sensibilidade foram obtidos pelo LDA para todas as classes de minerais deste artigo, constatando assim que os espectros de Raman não podem ser classificados por um algoritmo linear simples. Além disso, como esperado, os melhores resultados foram encontrados pelo classificador Autoencoder.

Apesar dos resultados encontrados serem superiores ao obtido por [8] é importante ressaltar que este estudo utilizou apenas 6 classes de minerais ao invés de todo banco de dados, entretanto a grande assertividade do autoencoder para classificação de espectros de Raman com a correção da linha de base nos incentiva avançar nos estudos.

4 CONCLUSÃO

Neste trabalho foi apresentada uma metodologia para o reconhecimento de minerais através da análise dos espectros de Raman utilizando-se a subamostragem das amostras aliadas a técnicas de aprendizado de máquinas SVM, LDA, kNN e Redes Neurais Autoencoder. O balanceamento dos dados com a inclusão dos ruídos foi uma técnica que apresentou resultados satisfatórios, ou seja, além de evitar a polarização do classificador, o sinal de espectro gerado com ruídos é similar aos espectros gerados num processo industrial com diversas influências externas. Além disso, os resultados obtidos para a escolha do melhor classificador demonstram que o autoencoder apresentou o melhor desempenho para o treinamento e teste com uma acurácia de 99,17% e com o menor tempo de processamento (aproximadamente 56 minutos). Como esperado, os resultados obtidos pelo LDA confirmam que não se trata de um problema de classificação linear. Ademais, é interessante numa etapa futura aumentar o número de classes considerando outros minerais relevantes que compõem o minério de ferro, além de se utilizar outras técnicas de classificação como o *stacking*, utilizado para a combinação de classificadores.

REFERÊNCIAS

- 1 Carter J.; Auyong K.; Dixon L. Determination of Iron Ore Mineralogy using Fourier Transform Infrared Spectroscopy: a Chemometric Approach. In: MATLAB Conference, 2017, Perth: Bureau Veritas, 2017.
- 2 Souza J S.; Ferrão M F. Aplicação da espectroscopia no infravermelho no controle de qualidade de medicamentos contendo diclofenaco de potássio. Parte I: Dosagem por regressão multivariada. *Revista Brasileira de Ciências Farmacêuticas* , 2006, v 42. n 3: 437-445
- 3 Lima A.; Bakker J. Espectroscopia no infravermelho próximo para monitorização da perfusão tecidual. *Revista Brasileira de Terapia Intensiva* , 2011, v 23. n 3: 341-351
- 4 Barcellos D C. Caracterização do carvão vegetal através do uso de espectroscopia no infravermelho próximo. Viçosa: Universidade Federal de Viçosa, MG, 2007.
- 5 Ramanaidou E. et al. Characterization of iron ore by visible and infrared reflectance and, Raman spectroscopies. In: Iron Ore: Mineralogy, Processing and Environmental Sustainability. Woodhead Publishing, 2015. 191-228.
- 6 Lu, Yang. Industry 4.0: a survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, 2017, v. 6: 1-10
- 7 Gölzer, Philipp, Patrick Cato, and Michael Amberg. "Data Processing Requirements of Industry 4.0-Use Cases for Big Data Applications." ECIS. 2015.
- 8 Liu, Jinchao, et al. "Deep convolutional neural networks for Raman spectrum recognition: a unified solution." *Analyst*, 2017, 142.21: 4067-4074.
- 9 Schulze, Georg, et al. "Investigation of selected baseline removal techniques as candidates for automated implementation." *Applied spectroscopy*, 2005, 59.5: 545-574.
- 10 Almeida, Leandro Maciel, et al. "Uma ferramenta para extração de padrões." *Revista Eletrônica de Iniciação Científica* 3 (2003).
- 11 Mazurowski, Maciej A., et al. "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance." *Neural networks*, 2008, 21.2-3: 427-436.
- 12 MathWorks Documentation Center, [acesso em 28 jul. 2018]. Disponível em: http://www.mathworks.com/help/comm/ref/awgn.html?s_tid=doc_ta
- 13 Utsunomiya S, Fujita Y, Tanaka S, Kajihara S, Aoshima K, Oda Y and Tanaka K 2014 Signal processing algorithm development for Mass++ (ver. 2): platform software for mass spectrometry *IP SJ Trans. Bioinforma.* 7 24–9
- 14 MathWorks Documentation Center, [acesso em 28 jul. 2018]. Disponível em: <http://www.mathworks.com/help/bioinfo/ref/msbackadj.html>
- 15 Mohamad, Muhammad'Arif, et al. "A review on feature extraction and feature selection for handwritten character recognition." *International Journal of Advanced Computer Science and Applications*, 2015, 6.2: 204-212.
- 16 Bengio, Y.: *Learning Deep Architectures for AI*. Found. Trends Mach. Learn., Now Publishers Inc, 2009, .vol. 2, n 1: 1-127.
- 17 Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In: B. Schölkopf, J. C. Platt, and T. Hoffman (eds) *Proceedings of the 19th International Conference on Neural Information Processing Systems 2006, NIPS'06*, MIT Press, Cambridge, MA, USA, 2006: 153-160.
- 18 Burges, C. J. C.: *A Tutorial on Support Vector Machines for Pattern Recognition*. *Data Min. Knowl. Discov.* 2, vol. 2, pp. 121-167 (1998).
- 19 Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, (1967).
- 20 Fisher, R. A.: *The Use of Multiple Measurements in Taxonomic Problems*. *Annals of Eugenics*, vol. 7, pp. 179–188. Blackwell Publishing Ltd (1936).