❒2857

Classifying lymphoma and tuberculosis case reports using machine learning algorithms

Moanda Diana Pholo[1], Yskandar Hamam[2], Abdel Baset Khalaf[3], Chunling Du[4]

[1,4]Department of Computer Engineering, Tshwane University of Technology, South Africa
[2]Department of Electrical Engineering, Tshwane University of Technology, South Africa
[3]School of Science & Engineering Technology, Durham College, Canada

Article Info

ABSTRACT

Available literature reports several lymphoma cases misdiagnosed as tuberculosis, especially in countries with a heavy TB burden. This frequent misdiagnosis is due to the fact that the two diseases can present with similar symptoms. The present study therefore aims to analyse and explore TB as well as lymphoma case reports using Natural Language Processing tools and evaluate the use of machine learning to differentiate between the two diseases. As a starting point in the study, case reports were collected for each disease using web scraping. Natural language processing tools and text clustering were then used to explore the created dataset. Finally, six machine learning algorithms were trained and tested on the collected data, which contained 765 lymphoma and 546 tuberculosis case reports. Each method was evaluated using various performance metrics. The results indicated that the multi-layer perceptron model achieved the best accuracy (93.1%), recall (91.9%) and precision score (93.7%), thus outperforming other algorithms in terms of correctly classifying the different case reports.

Corresponding Author:

Moanda Diana Pholo
Department of Computer Systems Engineering
Tshwane University of Technology
2 Aubrey Matlala Rd, Block L, Soshanguve, Gauteng, South Africa
Email: pholodiana@gmail.com

1. INTRODUCTION

According to the United Nations Programme on HIV/AIDS (UNAIDS), tuberculosis (TB) is the deadliest infectious disease worldwide [1]. It is therefore a major threat to global health. In an effort to combat this threat, countries like South Africa administer empiric TB treatment to patients likely to be suffering from the disease [2], [3]. This means that the patients receive treatment while awaiting for their TB laboratory results.

TB symptoms include fatigue, fever, dyspnea and night sweats. On radiological images, the disease can present as masses and fold thickenings [4]. However these symptoms are not unique to tuberculosis, leading to possible misdiagnoses. One disease that is often misdiagnosed as TB is lymphoma, a cancer which occurs when lymphocytes inside the lymph nodes multiply too fast or live too long [5], [6].

Many cases of lymphoma have been diagnosed as TB, as reported by [7]-[9], with the misdiagnosed patients receiving TB treatment while their cancer progresses. This is why the current study investigates the use of machine learning and natural language processing (NLP) in differentiating between the two diseases.

Machine learning (ML) is a sub-field of artificial intelligence (AI) which aims to process data, identify patterns in them and learn those patterns, without being explicitly programmed to do so [10], [11]. It

has proven to improve diagnosis prediction, as reported by [12]. NLP, on the other hand, focuses on extracting information from unstructured texts and converting it into a format that computers can process [13]. It has been successfully implemented in decision support systems for areas such as risk stratification, symptom identification and medical diagnosis [14], [15].

Although generic text classifiers exist, they generally are not tuned for the scientific data analysis [16]. This is why specific NLP systems have been used to classify patients with TB since the early 1990s [17] [18], and differentiate between TB and other pulmonary diseases [19]. To the best of our knowledge, there is however no system differentiating specifically between TB and lymphoma. Hence the overall purpose of our study is to create an NLP system to classify lymphoma and TB diagnosis. The system could serve for screening purposes and help reduce the misdiagnosis rate between the two diseases.

In our previous paper [20], we classified the two diseases using case reports collected from ScienceDirect. The features in each report were extracted using TF-IDF as well as Amazon Medical Comprehend, which is an NLP API for medical feature extraction. The current paper aims to; 1) analyse the collected case reports using NLP and clustering, 2) explore their different characteristics, 3) and identify documents which are not case reports of either diseases using machine learning algorithms.

This will help us collect additional relevant case reports from various sources, and design a more robust training dataset to be used in differentiating TB and lymphoma. All algorithms in this study are implemented using the "sklearn" Python module [21], using the default parameters, with no parameter tuning. Another limitation of this study is that it disregards the semantic value when extracting terms from the collected text. The rest of this paper is organised as; section 2 discusses the methods used in this study while section 3 presents the results of the various experiments performed. Finally, we discuss the results obtained in section 4 and conclude this paper in section 5.

2. RESEARCH METHOD
2.1. Data collection
Figure 1 gives a summary of the methodology applied. To create our dataset, we automatically scraped tuberculosis and lymphoma case reports from ScienceDirect through their search API using the following search terms; "tuberculosis case report", "tuberculosis case report". The case reports were restricted based on title, as described in [20]. For each search result returned, we retrieved the full article using ScienceDirect's Full-text retrieval API, then extracted the second section as the case report. This was achieved using a Python library called Beautiful Soup. A summary of the data collection process is shown in Figure 2.

2.2. Data pre-processing
The first part of preparing the data for our machine learning algorithms was done using "natural language toolkit" (NLTK), a Python module for NLP. This process consisted of the following steps:
- Contractions expansion; using the 'contractions' Python package, known shortened combinations of words were expanded back to their original form.
- Tokenization; each document was split into a series of words. Punctuation, numbers and special characters were then removed and letters converted to lower case.
- Stopwords removal; recurrent English words which convey little to no information, such as articles and pronouns, were removed from the text. NLTK's stopwords list was extended to also include terms such as 'patient', 'disease', 'using', 'figure', 'fig', 'clinic', 'hospital', 'et','al'. These terms appeared in multiple texts without bringing information necessary to our classification task.
- Lemmatization; using NLTK's WordNetLemmatizer algorithm, words were reduced to their root form in an effort to group together similar words (e.g.; plural words were converted to their singular form).

The result data were then converted from free text into a vector space using term frequency-inverse document frequency. Extra pre-processing consisted of extracting the age and gender of each patient. The detailed feature extraction process is reported in [20].

2.3. Data exploration
Using the "scikit-learn" library in Python, k-means++ clustering was applied to the vectorised dataset in order to group together similar case reports. The algorithm is described as [22]:
a. Choose k initial centroids
   For k iterations:
- For each data point, calculate the Euclidian distance with the closest centroid.
- Choose a centroid using a distribution specified by the squared Euclidean distances.

b. Do
−   Assign each data entry to the cluster specified by the closest centroid.
−   Compute the mean of each cluster and assign that value as the new centroid. until the centroids stay the same between consecutive iterations.
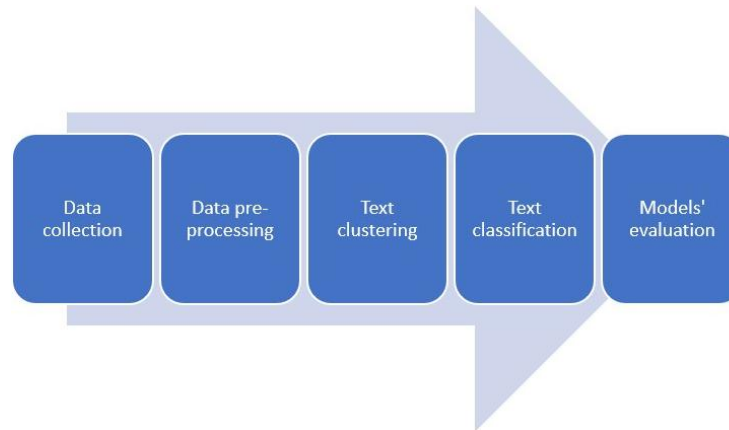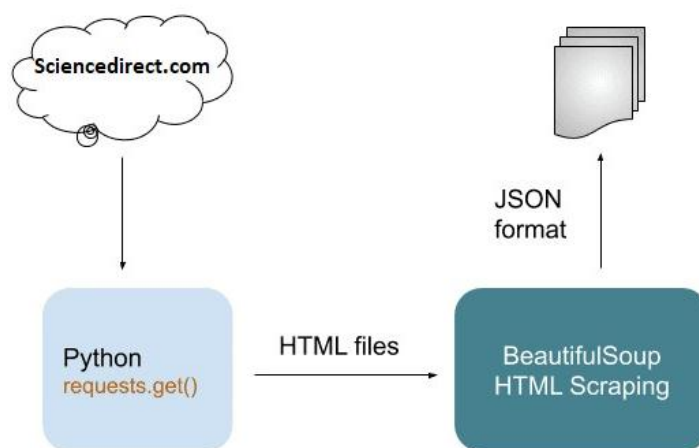

Figure 1. Methodology

Figure 2. Web scraping using Beautiful Soup and Python

The optimal number of clusters (k) was decided based on silhouette scores, which measure how cohesive and distinguishable clusters are. In (1) is the formula used:

$$SC = \frac{b - a}{max(a, b)} \tag{1}$$

where a is the average distance between each data point and other data points in the same cluster and b is the average distance between each data point and other data points in the closest cluster. For each data point, a and b are calculated as follows:

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \tag{2}$$

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i,j) \tag{3}$$

2.4. Text classification

We implemented the following algorithms as a benchmark: logistic regression, k-Nearest Neighbours (KNN), artificial neural network (ANN), Naive Bayes, support vector machines (SVM) and perceptron. There are brief descriptions of each algorithm:

a. Decision trees; this method returns a tree-like structure, where internal nodes perform tests based on attribute values and each branch represents the outcome of the test. The tree ends in leaf nodes, which are associated with the most probable decision. [23]. Instances are classified by traversing the tree and applying rules at each internal node until a decision node is reached [24].

b. Artificial neural network; an ANN consists of layers of artificial neurons which are connected with each other. Input data traverse the layers, which process it and output a result. [25, 26]. Each neuron receives the input data from neurons in the previous layer, and each neuron-to-neuron connection has a weight representing its strength [25].

We used a multi layer perceptron (MLP) of one hidden layer with 100 hidden units. This ANN determines the input weights of each linear model as follows [27]:

− Initialize w=0
− Go through the data points { xi, yi }
− if a data point is misclassified then $w \leftarrow w + \alpha sign(f(x_i))x_i$
− Until all the data are correctly classified

c. Naive Bayes; Naive Bayes is a simple, statistics-based method, which predicts a class (Y) for a new example (X) based on the largest a posteriori probability, previous experience and event probability [28].

The probability of X belonging to a class c is given by the following formula.

$$P(c|X) = \frac{P(X|c)P(Y)}{P(X)} \tag{4}$$

where:

$P(c)$: probability of class c
$P(X)$: probability of the predictors X
$P(X|c)$: probability of having X features given class c
$P(c|X)$: probability of an instance X belonging to class c given the value of its dependent variables [29]

d. Support vector machines; using a dataset of n features, a Support Vector Machine (SVM) attempts to find a decision boundary which maximises the margin between two observed classes [30]. This makes it a robust choice for binary classification. In the simplest case, SVMs must come up with a linear classifier of the form [31]:

$$f(x) = w^T x + b \tag{5}$$

One method of determining the input weights is the perceptron algorithm described above.

e. k-Nearest Neighbours; this method classifies a new instance by finding the k most similar instances in an existing dataset. The similarity is determined using metrics such as Euclidean distance or Mahalanobis distance [32]. With two feature vectors $A=(x_1,x_2,...,x_m)$ and $B=(y_1,y_2,...,y_m)$, representing two data points with m features, the Euclidean distance is calculated as:

$$distance(A,B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \tag{6}$$

We evaluated the performance of each algorithm using classification accuracy, precision and recall. Accuracy evaluates the ratio of correctly classified instances. On the other hand, precision gives us the ratio of true positives among all instances classified as postive. Finally, recall computes the ratio of positive instances that were correctly classified.

For each evaluation metric above, the performance of each algorithm was estimated using cross-validation. The dataset was randomly split into 5 subsets then each algorithm run 5 times, with 4 subsets used as for training and one used for testing.

3.   RESULT AND DISCUSSION

The search terms submitted to the ScienceDirect API provided 6080 and 4034 articles for tuberculosis and lymphoma, respectively. After automatic title review, 546 TB and 765 lymphoma case reports were kept for our study. Figure 3 gives us a quick preview of some features obtained using TF-IDF.

old	girl	admitted	pediatric	disease	department	...	difficile	bristol	ewing	german	hypermetabolism	waldeyers	exceptionally
0.0	0.0	0.000000	0.024603	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.042663	0.158406	0.0	0.180864	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.000000	0.111094	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 3. Text features screenshot

Looking at Figure 4, we see that highest average silhouette occurs when n=3. Considering three (3) clusters is therefore optimal in this case, since it minimises similarities between different clusters while maximising similarities within each cluster. This means that case reports are less likely of being assigned to the wrong cluster.
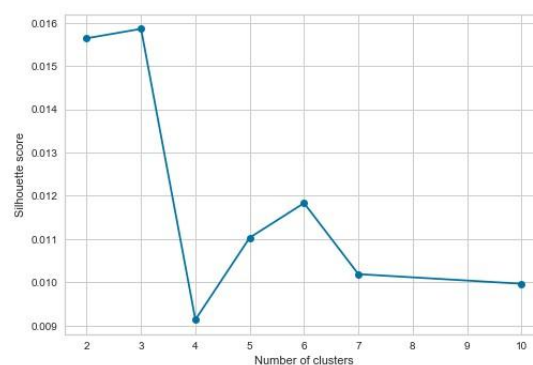

Figure 4. Silhouette analysis

3.1. Cluster analysis

Figure 5 shows a word cloud for each cluster, which help visualise the most important words per cluster. The most frequent words in Cluster 1, such as "hodgkin" and "cell" suggest that this cluster mainly contains lymphoma case reports. Examples of lymphoma cases that were assigned to this cluster include those reported by [33]-[35]. TB cases were mostly allocated to Cluster 2. These cases include those reported by [36]-[38]. However, it also contained articles discussing tuberculosis, which were not excluded during title review but were not case reports [39], [40]. The documents in Cluster 0 were neither tuberculosis nor lymphoma case reports. After analysis, it was found that this cluster consisted of many cases of diseases wrongly diagnoses as TB, as reported by [41], [42]. The cluster also contained cases where had another disease on top of tuberculosis or lymphoma.

Analysing the age of patients in the different clusters revealed that lymphoma patients were in average older than TB patients, with respective mean ages of about 53 and 40 years old as shown in Figure 6. This is consistent with previous findings indicating that lymphoma cases tend to occur in older patients [43], [44]. We also notice that lymphoma cases had a higher proportion of reported male patients. After pre-processing the text and vectorising, we obtained 7088 features to be fed into machine learn ing algorithms. Table 1 shows the average cross-validation performances of each algorithm in terms of accuracy, recall and precision.
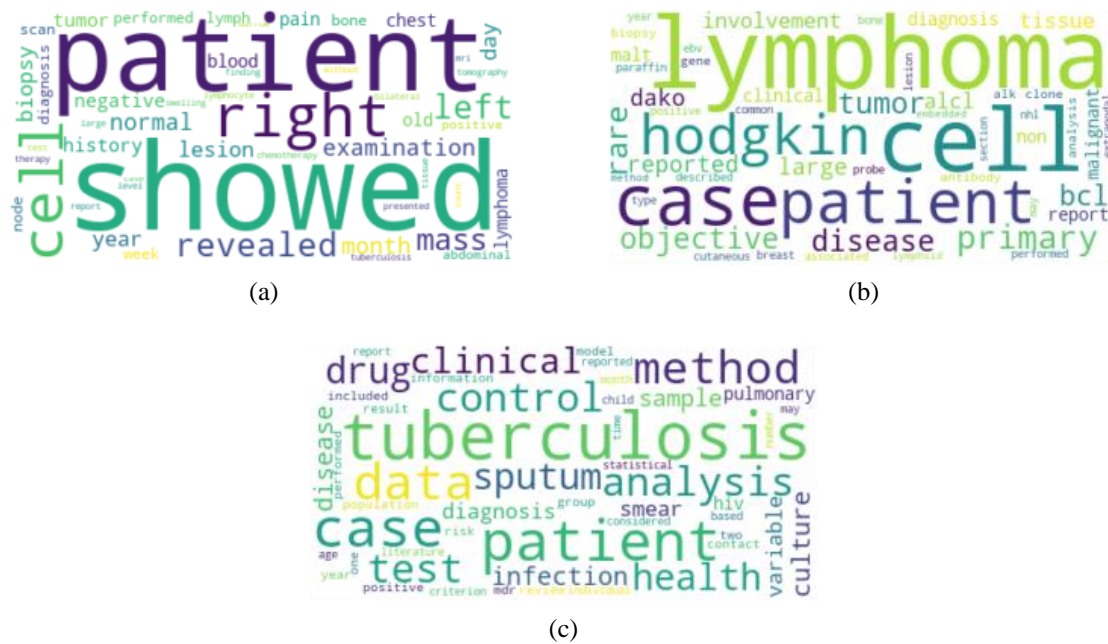
(a)   (b)

(c)

Figure 5. Clusters' word clouds, (a) cluster 0: others, (b) cluster 1: lymphoma, (c) cluster 2: TB

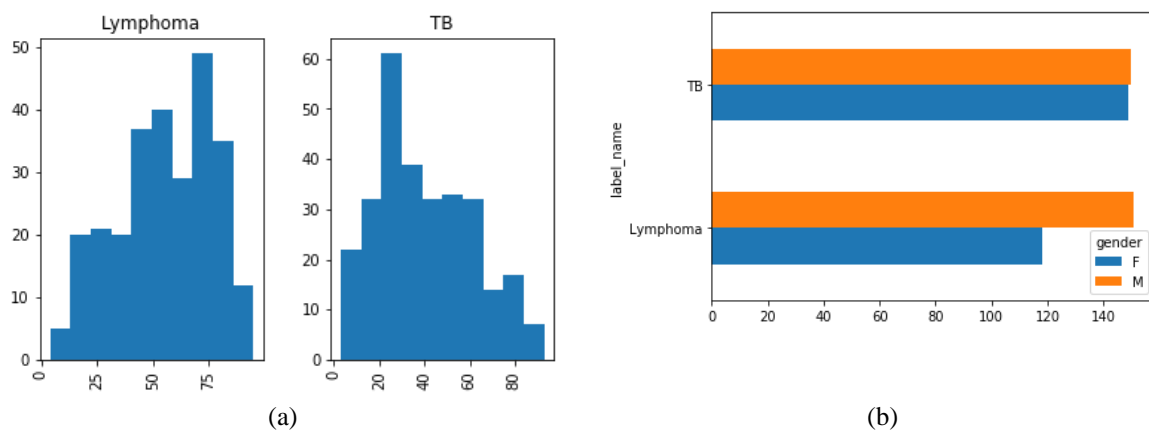(a)   (b)

Figure 6. Patients' overview, (a) age distribution, (b) gender distribution

Table 1. Algorithms' performances

	Accuracy	Precision	Recall
Logistic Regression	86.6%	88.1%	87.4%
kNN	70.5%	76.7%	66.6%
Decision Trees	92.3%	91%	93%
Naive Bayes	86.5%	83.7%	88.7%
SVM	45.6%	15.2%	33.3%
Perceptron	93.1%	91.9%	93.7%

Performance evaluation of the various algorithms showed that the Multi-Layer Perceptron algorithm best identified the correct class of case reports (with 93.1% accuracy). This method also achieved the highest recall score (94.1%) and the highest positive predictive value, with a precision score of 95.4%. It therefore minimised the possibility of misclassifying a case report and maximise the number of documents from a given class to be identified correctly.

These results show that machine algorithm can differentiate between TB and lymphoma case reports with high accuracy. Given that the reported results are cross-validation scores, it is likely that the trained model will perform well on unseen case reports. If implemented to classify case reports, it can help feed the right data into a diagnosis or referral support system. Such a system can be used to screen patients and detect lymphoma cases earlier, potentially improving the patients' prognosis. This could be extremely useful in diagnosing cancer in people with HIV-related lymphoma, who tend to show non-specific symptoms [45]. It is important to note that SVM performed very poorly, most likely due to the fact that the algorithm's default sklearn parameters were used. Future research will therefore look into tuning the algorithm and selecting their optimal parameters.

4. CONCLUSION

Since tuberculosis symptoms are shared by many other diseases, there is a high probability of misdiagnosis, especially in areas with restricted resources. And although there are various diagnosis machine learning systems, this study focuses on collecting and exploring data for a system dedicated to differentiating between tuberculosis and lymphoma.

As a starting point, the study used web scraping to collect available TB and lymphoma case reports, then used unsupervised methods to explore these latter. Case reports were assigned to one of three clusters: lymphoma, TB and "others". The results obtained after applying various classification algorithms on the dataset showed that the MLP model outperformed other algorithms when it came to accuracy, recall as well as precision, making it most likely to classify a case report correctly. This provides us with a tool for collecting additional case reports from different while ensuring the quality of the collected data.

Future research will aim to improve the MLP and decision tree models by tuning their hyper-parameters. The pre-processing will also compare the performance when using stemming instead of lemmatization, since words like "abdomen" and "abdominal" are still seen as different concepts using the latter method. We will further collect case reports for the extraction of semantic features, such as patient symptoms. The resulting feature space will then be used to train a TB/Lymphoma screening support system.

REFERENCES

[1] UNAIDS, "Tuberculosis-good progress, but not enough," 2020, [Online], Available: www.unaids.org/en/resources/presscentre/featurestories/2020/march/20200323, Accessed 25-October-2020.
[2] R. Jhamb, S. V. Madhu, and A. Juneja, "Non hodgkins lymphoma masquerading as tuberculosis," *Journal of Blood Disorders and Transfusion*, vol. 7, no. 3, 2016, doi: 10.4172/2155-9864.1000349.
[3] K. McCarthy, K. Fielding, G. J. Churchyard, and A. D. Grant, "Empiric tuberculosis treatment in south african primary health care facilities-for whom, where, when and why: Implications for the development of tuberculosis diagnostic tests," *PloS one*, vol. 13, no. 1, 2018, doi: 10.1371/journal.pone.0191608.
[4] A. B. Uy, A. M. Garcia, A. Manguba, and A. Loyola, "Tuberculosis: the great lymphoma pretender," *International Journal of Cancer Research and Molecular Mechanisms*, vol. 2, no. 1, 2016, doi: 10.16966/2381-3318.123.
[5] B. Liu, Y. Liu, J. Zhang, Y. Zeng, and W. Wang, "Application of the synergetic algorithm on the classification of lymph tissue cells," *Computers in biology and medicine*, vol. 38, no. 6, pp. 650-8, June 2008, doi: 10.1016/j.compbiomed.2008.03.003.
[6] Lymphoma Research Foundation, "About lymphoma," 2016, [Online], Available: https://lymphoma.org/aboutlymphoma/. Accessed 27-September-2020.
[7] B. Puvaneswaran and B. Shoba, "Misdiagnosis of tuberculosis in patients with lymphoma," *The South African Medical Journal*, vol. 103, no. 1, pp. 32-33, 2013.
[8] N. O. Ndjeka, R. Matji, and G. A. Ogunbanjo, "An approach to the diagnosis, treatment and referral of tuberculosis patients: The family practitioners' role," *SA Family Practice*, vol. 50, no. 4, pp. 44-49, 2008.
[9] L. Antonangelo, F. S. Vargas, E. H. Genofre, C. M. N. de Oliveira, L. R. Teixeira, and R. K. B. de Sales, "Differentiating between tuberculosis-related and lymphoma-related lymphocytic pleural effusions by measuring clinical and laboratory variables: is it possible?" *Jornal Brasileiro de Pneumologia*, vol. 38, no. 2, pp. 181-187, 2012, doi: 10.1590/S1806-37132012000200006.
[10] O. H. Hamid, N. L. Smith and A. Barzanji, "Automation, per se, is not job elimination: How artificial intelligence forwards cooperative human-machine coexistence," *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, 2017, pp. 899-904, doi: 10.1109/INDIN.2017.8104891.
[11] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborova, "Machine learning and the physical sciences," *Reviews of Modern Physics*, vol. 91, no. 4, 2019, doi: 10.1103/RevModPhys.91.045002.
[12] Y. Mu, M. Huang, C. Ye, and Q. Wu, "Diagnosis prediction via recurrent neural networks," *International Journal of Machine Learning and Computing*, vol. 8, no. 2, 2018, doi: 10.18178/ijmlc.2018.8.2.673.
[13] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of indonesia's capital city," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1620-1630, 2020, doi: 10.11591/eei.v9i4.2352.
[14] A. Agah, "Medical applications of artificial intelligence," CRC Press, 2013.

[15] M. Chary, S. Parikh, A. F. Manini, E. W. Boyer, and M. Radeos, "A review of natural language processing in medical education," *Western Journal of Emergency Medicine*, vol. 20, no. 1, p. 78, 2019, doi: 10.5811/westjem.2018.11.39725.

[16] J. Shruthi and S. Swamy, "A prior case study of natural language processing on different domain," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, pp. 4928-4936, 2020, doi: 10.11591/ijece.v10i5.pp4928-4936.

[17] N. L. Jain, C. A. Knirsch, C. Friedman, and G. Hripcsak, "Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports." in *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association, 1996, p. 542.

[18] G. Hripcsak, C. A. Knirsch, N. Jain, and A. Pablos-Mendez, "Automated tuberculosis detection," *Journal of the American Medical Informatics Association: JAMIA*, vol. 4, pp. 376-81, 09 1997.

[19] G. Hripcsak, J. H M Austin, P. Alderson, and C. Friedman, "Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports," *Radiology*, vol. 224, pp. 157-63, 08 2002, doi: 10.1148/radiol.2241011118.

[20] M. D. Pholo, Y. Hamam, A. Khalaf and C. Du, "Combining TD-IDF with symptom features to differentiate between lymphoma and tuberculosis case reports," *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1-4, doi: 10.1109/GlobalSIP45357.2019.8969317.

[21] F. Pedregosa, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[22] T. Gupta and S. P. Panda, "A comparison of k-means clustering algorithm and clara clustering algorithm on iris dataset," *International Journal of Engineering & Technology*, vol. 7, no. 4, pp. 4766-4768, 2018, doi: 10.14419/ijet. v7i4.21472.

[23] L. Rokach and O. Maimon, "Decision Trees," In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, vol. 6, pp. 165-192, 2005, doi: 10.1007/0-387-25465-X_9.

[24] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015, doi: 10.11919/j.issn.1002-0829.215044.

[25] M. H. M. Zaman, M. M. Mustafa, M. Hannan, and A. Hussain, "Neural network based prediction of stable equivalent series resistance in voltage regulator characterization," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 1, pp. 134-142, 2018, doi: 10.11591/eei.v7i1.857.

[26] Z. Qatawneh, M. Alshraideh, N. Almasri, L. Tahat, and A. Awidi, "Clinical decision support system for venous thromboembolism risk classification," *Applied Computing and Informatics*, vol. 15, no. 1, pp. 12-18, 2019, doi: 10.1016/j.aci.2017.09.003.

[27] A. Zisserman, "Lecture 2: The SVM classifier," 2015, [Online] Available: https://www.robots.ox.ac.uk/az/lectures/ml/2011/lect2.pdf. [Accessed 22-December-2020], Accessed 22-December-2020.

[28] N. M. Isa, A. Amir, M. Ilyas, and M. Razalli, "Motor imagery classification in brain computer interface (bci) based on eeg signal by using machine learning technique," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 1, pp. 269-275, 2019, doi: 10.11591/eei.v8i1.1402.

[29] S. Mocherla, A. Danehy, and C. Impey, "Evaluation of naive bayes and support vector machines for wikipedia," *Applied Artificial Intelligence*, vol. 31, no. 9-10, pp. 733-744, 2017, doi: 10.1080/08839514.2018.1440907.

[30] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "Text classification based on gated recurrent unit combines with support vector machine," *International Journal of Electrical & Computer Engineering*, vol. 10, 2020, doi: 10.11591/ijece.v10i4.pp3734-3742.

[31] K. Davagdorj, S. Yu, S. Kim, P. Huy, J.-H. Park, and K. Ryu, "Prediction of 6 months smoking cessation program among women in korea," *International Journal of Machine Learning and Computing*, vol. 9, pp. 83-90, 02 2019, doi: 10.18178/ijmlc.2019.9.1.769.

[32] K. Venkatachalam and N. Karthikeyan, "Effective feature set selection and centroid classifier algorithm for web services discovery," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 2, pp. 441-450, February 2017, doi: 10.11591/ijeecs.v5.i2.pp441-450.

[33] E. Flores, N. Aydin, D. Vu, and S. Misra, "A case series of diffuse large b-cell lymphoma and burkitt lymphoma presenting with peritoneal lymphomatosis," *International Journal of Surgery Case Reports*, vol. 28, pp. 262-265, 2016, doi: 10.1016/j.ijscr.2016.10.007.

[34] K.-M. Wong, C.-S. Chang, C.-C. Wu, and H.-L. Yin, "Hodgkin's lymphoma-related vanishing bile duct syndrome: A case report and literature review," *The Kaohsiung Journal of Medical Sciences*, vol. 29, no. 11, pp. 636-641, 2013, doi: 10.1016/j.kjms.2013.05.002.

[35] H. J. Bae, G. R. Chon, D. J. Kim, S. H. Lee, and J.-Y. Ahn, "A case of intravascular large b-cell lymphoma of lung presenting with progressive multiple nodules on chest computed tomography," *Respiratory Medicine Case Reports*, vol. 21, pp. 108 - 112, 2017, doi: 10.1016/j.rmcr.2017.04.009.

[36] J. Park and S. Kyung, "Unilateral primary tuberculosis presenting as an extratarsal chalazion," *Canadian Journal of Ophthalmology*, vol. 52, no. 1, pp. e1-e3, 2017, doi: 10.1016/j.jcjo.2016.10.006.

[37] M. Kimura, H. Araoka, H. Baba, C. Okada, Y. Murase, A. Takaki, S. Mitarai, and A. Yoneyama, "First case of sexually transmitted asymptomatic female genital tuberculosis from spousal epididymal tuberculosis diagnosed by active screening," *International Journal of Infectious Diseases*, vol. 73, pp. 60-62, 2018, doi: 10.1016/j.ijid.2018.05.021.

[38] S. K. Swain and M. C. Sahu, "Isolated nasopharyngeal tuberculosis in a 12 year old boy - a case report," *Egyptian Pediatric Association Gazette*, vol. 66, no. 3, pp. 78-80, 2018, doi: 0.1016/j.epag.2018.05.003.

[39] E. A. Dodor and S. J. Kelly, "Manifestations of tuberculosis stigma within the healthcare system: The case of sekondi-takoradi metropolitan district in ghana," *Health Policy*, vol. 98, no. 2-3, pp. 195-202, 2010, doi: 10.1016/j.healthpol.2010.06.017.

[40] K. Ahmad, "Nearly 70% of tuberculosis cases remain undetected," *The Lancet Infectious Diseases*, vol. 2, no. 6, p. 319, 2002, doi: 10.1016/S1473-3099(02)00301-8.

[41] J. Yu, J. Shen, L. Wang, R. Zhu, and J. Wu, "A case report of atypical sarcoidosis misdiagnosed as tuberculosis," *Radiology of Infectious Diseases*, vol. 3, no. 1, pp. 40 - 43, 2016, doi: 10.1016/j.jrid.2016.01.005.

[42] L. Roche, C. Gulman, and M. O'Kane, "A case of genital pyoderma gangrenosum successfully treated with cyclosporine without relapse of established follicular lymphoma," *JAAD Case Reports*, vol. 4, no. 5, pp. 474-476, 2018, doi: 10.1016/j.jdcr.2018.01.004.

[43] C. Thieblemont, S. Bernard, and T. Molina, "Management of aggressive lymphoma in very elderly patients," *Hematological oncology*, vol. 35, no. Suppl Suppl 1, p. 49, 2017, doi: 10.1002/hon.2413.

[44] B. Boll and H. G¨ orgen, "The treatment of older hodgkin lymphoma patients," *British journal of haematology*, vol. 184, no. 1, pp. 82-92, 2019, doi: 10.1111/bjh.15652.

[45] P. Buyego, L. Nakiyingi, H. Ddungu, S. Walimbwa, D. Nalwanga, S. Reynolds, and R. Parkes-Ratanshi, "Possible misdiagnosis of HIV associated lymphoma as tuberculosis among patients attending Uganda Cancer Institute," *AIDS Research and Therapy*, vol. 14, no. 1, 3 2017, doi: 10.1186/s12981-017-0139-x.

BIOGRAPHIES OF AUTHORS

Moanda D. Pholo, born in Kinshasa (D.R.C.) in 1987, is a PhD candidate at the Tshwane University of Technology. She has a Masters degree in Intelligent Industrial Systems from the same university as well as an Msc in electronics from Ecole Sup´ erieure d'Ing´ enieurs en´ Electrotechnique et´ Electronique, obtained in´ 2014. She currently works as a data scientist in Pretoria, South Africa. She has published a number of scientific articles, with the latest publication appearing in the GlobalSIP 2019 Proceedings. Mrs Pholo has recently received the award of "Best work in the category of AI/ big data" at the 3rd ICEHTMC in Italy.

Yskandar Hamam completed his bachelor's degree at the American University of Beirut in 1966. He then obtained M.Sc. and Ph.D. degrees from the University of Manchester Institute of Science and Technology, respectively in 1970 and 1972, and the "Diplome d'Habilitationˆ a Diriger` des Recherches" degree from the Universite des Sciences et Technologies de Lille in 1998. He has´ conducted research and been a lecturer in various countries including England, Brazil, Lebanon, Belgium, and France. He was also an Active Member in modelling and simulation societies and the President of EUROSIM and the Head of the Control Department and the Dean of the Faculty with ESIEE Paris (France). From 2007 to 2012 he served as the Scientific Director of the French South African Institute of Technology, South Africa. He is currently an Emeritus Professor with the Department of Electrical Engineering, TUT as well as at ESIEE-Paris, France. He has co-authored four books, and 40 chapters in edited books. He has also authored or co-authored more than 400 papers in peer reviewed archived journals and in peer-reviewed conference proceedings

Abdel Baset Khalaf is a Professor of Healthcare Technology Management at Durham College in Canada and has 35 years of experience in Clinical Engineering and Health Technology Management. In 2011, he was appointed as a member of the ministerial advisory committee on Health Technology for 5 years. In 2004, he was awarded a master's degree in Clinical Engineering at Tshwane University of Technology. In 2012, he was awarded his PhD from Universite De Versailles Saint-Quentin-EnYvelines, France. He is the author and co-author of numerous research papers in medical equipment maintenance models and strategies as well as a book in mathematical maintenance models for medical equipment published by Scholars Press in Germany.

Chunling Du received an M.S. degree in electrical engineering from Tshwane University of Technology, Pretoria, South Africa, in 2010, followed by a Ph.D. in electrical engineering at same university. She currently works for the Department of Computer Engineering, Tshwane University of Technology, lecturing formal classes and supervising postgraduate students. Dr Du's research interests include image processing, AI, industrial control and pattern recognition.