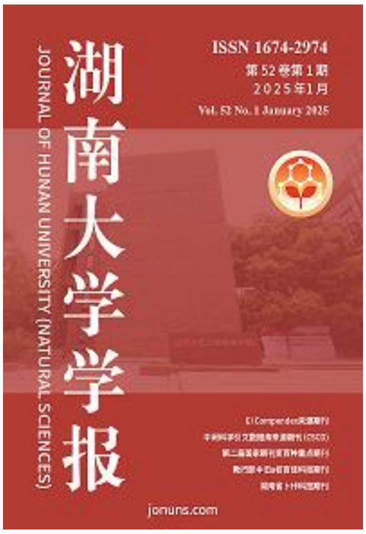

Journal of Hunan University

(Natural Sciences)

Vol. 52 No. 1

January 2025

Available online at

https://jonuns.com

Original research article

Open Access Article

https://doi.org/10.55463/issn.1674-2974.52.1.14

Optimization of Tuberculosis Diagnosis Using the Support Vector Machine Method on Health Data of Central Sulawesi Province

Hartayuni Sain[1]*, Firda Fadri[2], Mohammad Fajri[1], Rais[1], Defi Yusti Faidah[3]

[1]Statistics Study Program, Faculty of Mathematics and Natural Sciences, Tadulako University, Palu, Indonesia

[2]Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Jember University, Jember, Indonesia

[3]Department of Statistics, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Sumedang, Indonesia

* Corresponding author: hartayuni.sain@untad.ac.id

Article History:

Received: December 2, 2024
Reviewed: December 29, 2024
Revised: January 10, 2025
Accepted: January 17, 2025
Published: February 28, 2025

Abstract: Tuberculosis (TB) is a disease caused by infection with the bacterium Mycobacterium tuberculosis in the lungs. TB is a major global health problem, ranking as the second leading cause of death from infectious diseases worldwide. Given that TB is one of the infectious diseases that remains a public health problem, early and accurate detection is key in controlling its spread. This study uses historical data of TB patients from several hospitals in Central Sulawesi as research samples. One of the classification methods is the Support Vector Machine (SVM) method. This method was chosen because of its ability to classify non-linear data and its potential to produce high prediction accuracy with complex data. This study aims to improve the accuracy and efficiency of Tuberculosis (TB) disease diagnosis in Central Sulawesi Province through the application of the SVM method. The parameters used in this study were age, gender, body temperature, shortness of breath, chest pain, sputum examination, and final diagnosis. The results of this study show that handling the problem of imbalanced data with an approach at the data level using the Adaptive Synthetic Sampling (adasyn) method approach and SVM as a classification method for two classes in this study shows that this method can classify each research data used with an accuracy rate of more than

95%. The classification results show that the larger the data used for testing, the greater the classification accuracy. The best classification accuracy value is obtained from the scheme 4 division of training and testing data, namely 90%:10% with a value of 100%. This shows that the method used in this study can be used to help diagnose TB disease based on a patient's medical data. From this research, we introduce a web-based information system that can be used to conduct early detection for TB patients independently, where they are unwilling or embarrassed to come to the hospital.

Keywords: TB disease, Classification, Support Vector Machine, accuracy, Web-based application information system.

使用支持向量机方法对中苏拉威西省卫生数据进行结核病诊断优化

摘要：结核病 (TB) 是一种由肺部感染结核分枝杆菌引起的疾病。结核病是全球主要的健康问题，是全球第二大传染病死亡原因。鉴于结核病是仍然是公共卫生问题的传染病之一，早期和准确检测是控制其传播的关键。本研究使用中苏拉威西省几家医院的结核病患者历史数据作为研究样本。分类方法之一是支持向量机 (SVM) 方法。之所以选择这种方法，是因为它能够对非线性数据进行分类，并且有可能对复杂数据产生高预测精度。本研究旨在通过应用 SVM 方法提高中苏拉威西省结核病 (TB) 疾病诊断的准确性和效率。本研究使用的参数是年龄、性别、体温、呼吸急促、胸痛、痰液检查和最终诊断。本研究结果表明，采用数据层面的方法处理数据不平衡问题，本研究中采用自适应合成抽样（adasyn）方法和SVM作为两类分类方法，结果表明该方法可以对所使用的每个研究数据进行分类，准确率超过95%。分类结果表明，用于测试的数据越大，分类精度越高。方案4训练和测试数据的划分获得了最好的分类精度值，即90%：10%，值为100%。这表明本研究中使用的方法可用于根据患者的医疗数据帮助诊断结核病。通过这项研究，我们介绍了一个基于网络的信息系统，可用于在结核病患者不愿或不好意思来医院时独立进行早期检测

关键词：结核病，分类，支持向量机，准确率，基于网络的应用信息系统

1. Introduction

Tuberculosis or TB is a disease caused by infection with the bacterium Mycobacterium tuberculosis, primarily affecting the lungs. TB is a major global health problem, ranking as the second leading cause of death from infectious diseases worldwide. By 2022, the World Health Organization (WHO) estimated that 10 million people globally suffer from TB, resulting in 1.5 million deaths annually [1].

Indonesia is a country with the highest burden of TB in the world, ranking second globally in TB cases. The estimated number of people who fall ill from tuberculosis reaches 845.000 with a mortality rate of 98.000 equivalent to 11 deaths per hour. Of these cases, only 67% were detected and treated, leaving 283.000 TB patients untreated and at risk of spreading the infection to those around them [2].

The prevalence study on TB disease transmission conducted by the Central Sulawesi Provincial Health Office over the past three years recorded 10.207 cases. In 2020, there were 2.874 new TB cases reported in Central Sulawesi. This number increased to 3.143 cases in 2021 and 3.456 cases in 2022. As of September 2022, the region with the highest number of TB cases was Palu City, with 718 cases, followed by Banggai District with 579 cases, and Parigi Moutong District with 421 cases [3].

The classification of TB is crucial for ensuring an accurate diagnosis and appropriate treatment. Traditionally, TB classification relies on physical examinations, chest X-rays, and laboratory tests [4]. However, these methods have several limitations: 1)The interpretation of physical examination and chest X-ray results can vary depending on the doctor conducting the assessment; 2)Laboratory tests may not always accurately detect TB, particularly in patients with latent infections.

The support vector machine (SVM) is a machine learning algorithm widely used for data classification [5]. SVM has proven effective in various classification applications, including disease classification. This effectiveness is due to the SVM's ability to find a globally optimal solution and consistently reach the same solution in each run [6]. Compared to traditional classification methods, SVM can achieve higher accuracy and efficiently analyze large datasets [7][8].

Related to the problem of handling TB disease, the task of a doctor will be greatly helped when there is a system that can help doctors diagnose TB disease. The purpose of this research is not to replace the role of a doctor, but to provide recommendations or possible diagnostic results based on patient symptoms. Research on the classification of TB disease is still under development. Research conducted by [9] used artificial neural networks (ANN) for the classification of TB disease, but the accuracy value has not provided better results. Further research, [10] used k-nearest neighbor (KNN) in TB classification and gave an accuracy value of 78.66% and an AUC value of 0.806, which identified the model as a good classification. In this study, it is expected that the use of SVM methods for TB disease classification can help improve the quality of TB diagnosis and treatment. This can help reduce the mortality from TB and improve the quality of life of TB patients.

2. Materials and Methods

2.1. Tuberkulosis (TBC)

Tuberculosis (TB) is a contagious disease caused by the Mycobacterium tuberculosis bacteria [11]. It primarily affects the lungs but can also spread to other organs, a condition known as extrapulmonary TB [13],[14]. Pulmonary TB, the most common form, occurs in approximately 80% of cases. TB can infect anyone, especially those in the economically productive age group of 15 to 50 years old [15],[16]. Factors such as poverty, unhealthy lifestyles, poor environmental conditions, and lack of awareness of TB contribute to its spread. Early and appropriate treatment is crucial to prevent severe complications and fatalities. While TB is often associated with lung infections, it can also target organs such as the kidneys, bones, and brain [12]. Diagnosing extrapulmonary TB can be challenging, as symptoms may mimic other diseases.

2.2. Support Vector Machine

Support Vector Machines (SVM), first introduced by [5], are a powerful machine learning technique capable of handling both classification and regression tasks. SVMs excel at finding the optimal global solution, ensuring consistent results across multiple runs. The algorithm functions by mapping the training data into a higher-dimensional feature space. Within this space, an optimal hyperplane is constructed to maximize the margin between different data classes, effectively separating them into distinct categories (-1, +1).

Given a set of $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}\}$, with $\boldsymbol{x_i} \in \boldsymbol{R^n}$, $i = 1, \cdots, n$. Known that $\boldsymbol{X}$ is in a particular pattern, that if $\boldsymbol{x_i}$ is a member of a certain class then $\boldsymbol{x_i}$ given able (target) $y_i = +1$, else $y_i = -1$. Hence, data will be given as pair $(\boldsymbol{x_1}, y_1)$, $(\boldsymbol{x_2}, y_2)$, …, $(\boldsymbol{x_n}, y_n)$ which are a training vector set from two classes that will be classified by SVM,

$$(\boldsymbol{x_i}, y_i), \boldsymbol{x_i} \in \boldsymbol{R^n}, \quad y_i \in \{-1, 1\}, \ i = 1, \cdots, n, \tag{1}$$

A separating hyperplane is defined from a normal vector parameter called $\boldsymbol{w}$, and parameter defines relative plane position toward coordinate center called b, the two parameters are shown as follows:

$$(\boldsymbol{w}^T.\boldsymbol{x}) + b = 0 \tag{2}$$

In defining the canonical-shaped hyperplane separating, it must satisfy this constraint,

$$y_i \left[(\boldsymbol{w}^T.\boldsymbol{x_i}) + b\right] \geq 1, \ i = 1, 2, \cdots, n \tag{3}$$

while the optimal hyperplane is obtained by maximizing the margin $\frac{2}{\|\boldsymbol{w}\|}$ or by minimizing the following function:

$$\boldsymbol{\Phi}(\boldsymbol{w}) = \frac{1}{2} \|\boldsymbol{w}\|^2 \tag{4}$$

then the optimization problem can be solved by the Lagrange function,

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{l} \alpha_i \{y_i [(\boldsymbol{w^T}.\boldsymbol{x_i}) + b] - 1\} \tag{5}$$

where α_i is Lagrange multiplier. The Lagrange function is a primal space; thus, it is necessary to transform it into a dual space so that the function can be easier and more efficient to solve. Hence, solving the dual space can be obtained as follows:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i^T}.\boldsymbol{x}_j) - \sum_{i=1}^{l} \alpha_i \tag{6}$$

Subject to,

$$\alpha_i \geq 0, \qquad i = 1, 2, \cdots, n \text{ dan } \sum_{i=1}^{n} \alpha_i y_i = 0$$

Thus, the classification using the following formula [9],

$$f(x) = h(\hat{\boldsymbol{w}}^T.\boldsymbol{x} + \hat{b}) \tag{7}$$

$$\text{where } h(x) = \begin{cases} -1 & x < -1 \\ x & -1 \leq x \leq 1 \\ 1 & 1 \end{cases} \tag{8}$$

$$\boldsymbol{w} = \sum_{i=1}^{n} \hat{\alpha}_i y_i \boldsymbol{x}_i \text{ and } \hat{b} = -\frac{1}{2} \boldsymbol{w}(\boldsymbol{x}_r + \boldsymbol{x}_s) \tag{9}$$

2.3. Evaluation of the Model Performance

Classification accuracy is measured at the accuracy rate. This metric evaluates the overall performance of the classification model. A higher accuracy rate indicates a more effective model for correctly categorizing data points.

$$Accuracy\ rate = \frac{tp + tn}{tp + fp + fn + tn} \quad (10)$$

where tp is the number of true positives, fp is the number of false positives, tn is the number of true negatives, and fn is the number of false negatives.

3. Data Description

The research variables consisting of response variables and predictor variables from each data used in this study are as follows; The response variable used in this study is the Final Diagnosis, which is a variable with categories; 0 = "negative TB" and 1 = "positive TB". The predictor variables used are Patient Age ($\boldsymbol{X_1}$), Gender ($\boldsymbol{X_2}$), Fever Condition ($\boldsymbol{X_3}$) which is categorized into mild, moderate and high, Shortness of Breath ($\boldsymbol{X_4}$) which is categorized into mild and acute/chronic, Chest Pain ($\boldsymbol{X_5}$) which is categorized into mild and acute and Sputum Type ($\boldsymbol{X_6}$) which is categorized into clear, yellowish green and bloody.

Figure 1 shows a flow chart of the research methodology used.

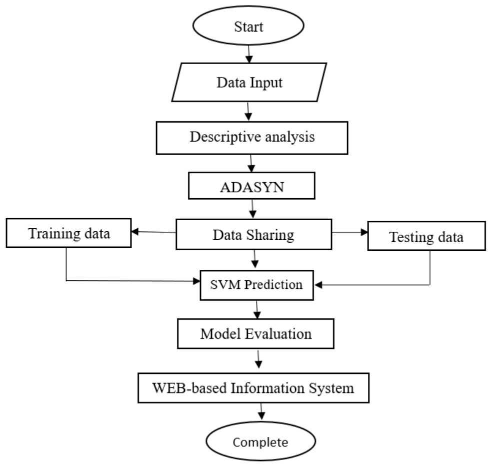

Figure 1. The research flow chart (developed by the authors)

4. Results

4.1. Data Description and Comparison

The first step is to divide the training and testing data. In this section, four training data division schemes were identified: testing data, namely as follows:

a. 60% : 40% division scheme
b. 70% : 30% division scheme
c. 80% : 20% sharing scheme
d. 90% : 10% division scheme

Table 1. Description of TB data used in the study (compiled by the authors)

Data Type	Data Description	Amount of Class Data	Minority Class	Majority Class
TBC	Tuberculosis or TB is a disease caused by infection of the Mycobacterium tuberculosis bacteria in the lungs. This condition is sometimes referred to as pulmonary TB. These data consists of 6 research variables and 2 classes, namely, negative TB and Positive TB.	980	399 (40.71%)	581 (59.29%)

Next, we will go through the data class balancing stage using the Adasyn method. The next stage is to perform classification with the SVM method using the Radial Basic Kernel (RBF) function with parameter values σ = 1, 10, and 100, and the cost parameters for optimization are C = 1, 10, and 100. The classification results in the SVM method for all specified schemes are as follows (Tables 2-8, Figures 2-5).

Table 2. Comparison of Scheme 1 TB data before and after applying Adisyn (compiled by the authors)

Division Scheme	Data Description	Training		Testing	
		0	1	0	1
60% : 40%	Before Adasyn	253 (43.02%)	335 (56.98%)	146 (37.24%)	246 (62.76%)
	After Adasyn	327 (49.40%)	335 (50.60%)	229 (48.21%)	246 (51.79%)

Table 3. Scheme 1 TB data classification accuracy rate with SVM (%) (compiled by the authors)

Data	C	σ		
		1	10	100
TBC scheme 1	1	95.41	97.45	97.45
	10	97.45	97.45	97.45
	100	97.45	97.45	97.45

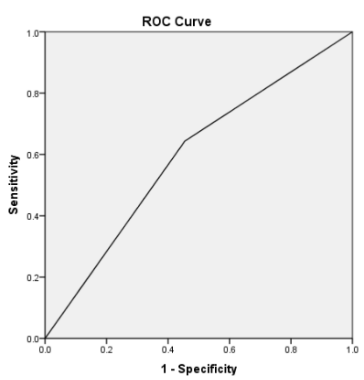

Figure 2. ROC curve for 60%:40% division scheme (developed by the authors)

Table 4. Comparison of Scheme 2 TB data before and after applying Adisyn (compiled by the authors)

Division Scheme	Data Description	Training		Testing	
		0	1	0	1
70% : 30%	Before Adasyn	266 (38.78%)	420 (61.22%)	133 (45.24%)	161 (54.76%)
	After Adasyn	400 (48.78%)	420 (51.22%)	149 (48.06%)	161 (51.94%)

Table 5. Scheme 2 TB data classification accuracy rate with SVM (%) (compiled by the authors)

Data	C	σ		
		1	10	100
TBC scheme 2	1	97.28	98.30	98.30
	10	98.30	98.30	98.30
	100	98.30	98.30	98.30

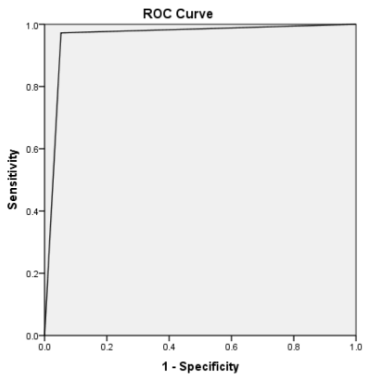

Figure 3. ROC Curve for 70%:30% division scheme (developed by the authors)

Table 6. Comparison of Scheme 3 TB data before and after applying Adisyn (compiled by the authors)

Division Scheme	Data Description	Training		Testing	
		0	1	0	1
80% : 20%	Before Adasyn	329 (41.97%)	455 (58.03%)	70 (35.71%)	126 (64.29%)
	After Adasyn	411 (47.46%)	455 (52.54%)	136 (51.91%)	126 (48.09%)

Table 7. Scheme 3 TB data classification accuracy rate with SVM (%) (compiled by the authors)

Data	C	σ		
		1	10	100
TBC scheme 3	1	98.47	98.47	98.47
	10	98.47	98.47	98.47
	100	98.47	98.47	98.47

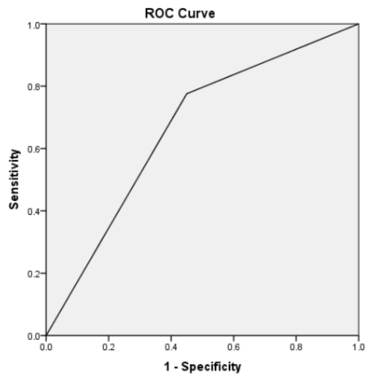

Figure 4. ROC Curve for 80%:20% division scheme (developed by the authors)

Table 8. Comparison of Scheme 4 TB data before and after applying Adisyn (compiled by the authors)

Division Scheme	Data Description	Training		Testing	
		0	1	0	1
90% : 10%	Before Adasyn	366 (41.50%)	516 (58.50%)	33 (33.68%)	65 (66.32%)
	After Adasyn	454 (46.80%)	516 (53.20%)	67 (50.76%)	65 (49.24%)

Table 9. Scheme 4 TB data classification accuracy rate with SVM (%) (compiled by the authors)

Data	C	σ		
		1	10	100
TBC scheme 4	1	100	100	100
	10	100	100	100
	100	100	100	100

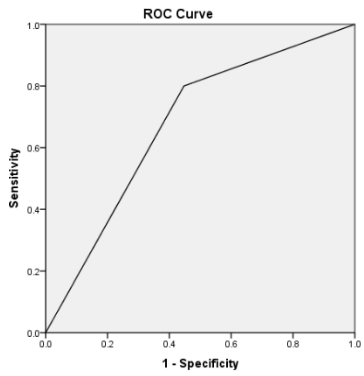

Figure 5. ROC Curve for 90%:10% division scheme (developed by the authors)

4.2. Web System Implementation

Subsequently, we have produced a web-based information system as an output of this research, entitled “Sistem Informasi Tuberkulosis (TBC) dengan Metode Support Vector Machine”. Our website is designed with several objectives. The general objectives are: to develop a web-based information system to assist in the early diagnosis of Tuberculosis, to improve the accuracy and efficiency of the TB diagnosis process, and to expand community access to TB diagnosis services. The specific objectives are: to build a comprehensive database of TB symptoms and diseases, to develop accurate and reliable diagnostic algorithms, to provide an easy-to-use and attractive user interface, and to integrate the system with existing health information systems.

We have designed the following visualization of the website. Figure 6 shows the initial view of our design. This is the homepage. This page is designed to make it easier for users, so they do not need to log in. By simply clicking on “Mulai Analisis” (Start Analysis), users can directly proceed to the next step.

Figure 6. Homepage (developed by the authors)

Figure 7 shows the next step where users are expected to input their biodata and fill in all the sub-districts according to the laboratory results. After that, they continue by clicking on the “Simpan” button (Save).

Figure 7. Biodata input (developed by the authors)

Figure 8 shows the diagnostic results. Patients or users only need to enter their name according to the patient data, and the prediction results for the patient or user will appear, which is in the form of “TBC atau Tidak TBC” (TB or Not TB).

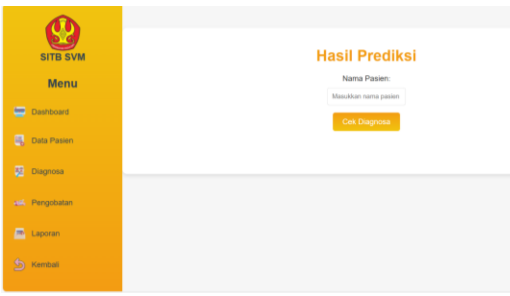

Figure 8. Diagnosis results (developed by the authors)

Figure 9 is the treatment recommendation. In this section, patients or users diagnosed with TB will be advised to seek further treatment according to the recommendations of a specialist.

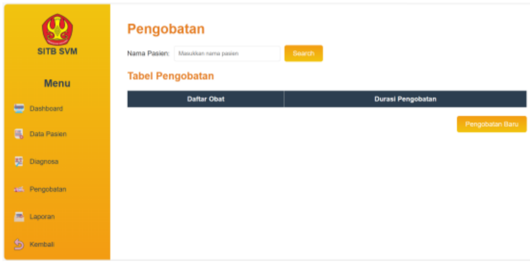

Figure 9. Treatment recommendations (developed by the authors)

5. Discussion

The classification results show that the larger the data used for testing, the greater the classification accuracy. The best classification accuracy value is obtained from scheme 4 division of training and testing data, namely 90%: 10% with a value of 100%. This shows that the method used in this study can be used to help diagnose TB disease based on a patient's medical data. Based on the results and conclusions obtained from this research, several things need to be developed. Increase the amount of data from several hospitals in Central Sulawesi so that further research can be done mapping areas based on the number of patients with TB so that it can be a reference for the Health Office or other related parties to focus more on handling TB cases. The next data that can be used as a continuation of this research is thorax photo data (thorax X-ray) even though this data is very difficult to obtain.

6. Conclusion

Handling the problem of data imbalance with the Adaptive Synthetic Sampling (adasyn) approach and SVM as a classification method for two classes in this study shows that the method can classify each research data used with an accuracy rate of more than 95%. This shows that the method used in this study can be used to help diagnose TB disease based on patient medical data. From this research, we developed and introduced a web-based information system that can be used to conduct early detection for TB patients independently, where they are unwilling or embarrassed to come to the hospital.

Declarations

Author Contributions

Conceptualization, H.S.; methodology, H.S., and F.F.; validation, H.S., F.F., and M.F.; formal analysis, H.S.; statistical analysis, F.F.; investigation, H.S., F.F., and M.F.; data curation, H.S.; writing—original draft preparation, all authors contributed equally; writing—review and editing, H.S.; visualization, R., and D.Y.F; supervision, H.S., and F.F.; project administration, H.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

Funding information is not available.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this manuscript. In addition, ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies, have been completely observed by the authors.

References

[1] WHO, *Global Tuberculosis Report*. 2023.

[2] WAHID W.W.C., ARSY F. F., SETIAWATI N., VICRI R. N., MIRANDA A.V., LESTARI B.W., and APRILLIA K.R. Increasing Financing for Tuberculosis Programs in Indonesia. *RD*I Policy Brief., 2023, 6, 20230619.

[3] RIZKANINGSIH R., and MUSTAFA M. Hubungan Kondisi Fisik Lingkungan Rumah dengan Kejadian TBC (Tuberculosis). *Jurnal Promotif Preventif*, 2023, 6(2): 335-343. https://doi.org/10.47650/jpp.v6i2.784

[4] REA, G., SPERANDEO, M., LIETO, R., BOCCHINO, M., QUARATO, C. M. I., FERAGALLI, B., VALENTE, T., SCIOSCIA, G., GIUFFREDA, E., FOSCHINO BARBARO, M. P., & LACEDONIA, D. Chest Imaging in the Diagnosis and Management of Pulmonary Tuberculosis: The Complementary Role of Thoraci Ultrasound. *Frontiers in Medicine*, 2021, 8: 753821. https://doi.org/10.3389/fmed.2021.753821

[5] VAPNIK V. N. *The Nature of Statistical Learning Theory*. Springer, N.Y., 2000. https://doi.org/10.1007/978-1-4757-3264-1

[6] SAIN H. ,and PURNAMI S. W. Combine Sampling Support Vector Machine for Imbalanced Data Classification. *Procedia Computer Science*, 2015, 72: 59–66, https://doi.org/10.1016/j.procs.2015.12.105

[7] CERVANTES J., GARCIA-LAMONT F., RODRÍGUEZ-MAZAHUA L., and LOPEZ A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 2020, 408: 189–215, https://doi.org/10.1016/j.neucom.2019.10.118

[8] HAMED B. A., IBRAHIM O. A. S., and ABD EL-HAFEEZ T. Optimizing classification efficiency with machine learning techniques for pattern matching. *Journal of Big Data*, 2023, 10(1): 124, https://doi.org/10.1186/s40537-023-00804-6

[9] FIBRIANTO A., MAGDALENA R., & FUADAH Y.N. Classification of normal pulmonary conditions, tuberculosis disease (tbc) and pleura effusion on humans using artificial neural network-Backpropagation. *e-Proceeding of Engineering*, 2018, 5(3): 5071–5078. https://core.ac.uk/reader/299924570

[10] WARDANI R. S., and PURWANTO. Model diagnosis tuberkulosis menggunakan k-nearest neighbor berbasis seleksi atribut External Monitoring Mission. *Proceedings of the 2nd University Research Coloquium*, 2015, 284-290. https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/1601

[11] SHOFIA E. N., REGASARI R., PUTRI M., and ARWAN A. Sistem Pakar Diagnosis Penyakit Demam : DBD, Malaria dan Tifoid Menggunakan Metode K-Nearest Neighbor – Certainty Factor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2017, 1(5), 426–435. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/122

[12] AINI N. and HATTA H. R. Sistem pakar pendiagnosa penyakit tuberkulosis. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 2017, 12(1): 56–63. http://dx.doi.org/10.30872/jim.v12i1.224

[13] CUDAHY P., and SHENOI S. V Diagnostics for pulmonary tuberculosis. *Postgraduate Medical Journal*, 2016, 92(1086): 187–193, https://doi.org/10.1136/postgradmedj-2015-133278

[14] LEWINSOHN, D. M., LEONARD, M. K., LOBUE, P. A., COHN, D. L., DALEY, C. L., DESMOND, E., KEANE, J., LEWINSOHN, D. A., LOEFFLER, A. M., … & WOODS, G. L. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 2017, 64(2), e1–e33. https://doi.org/10.1093/cid/ciw694

[15] CHENG, J., SUN, Y. N., ZHANG, C. Y., YU, Y. L., TANG, L. H., PENG, H., PENG, Y., YAO, Y. X., HOU, S. Y., LI, J. W., ZHAO, J. M., XIA, L., XU, L., XIA, Y. Y., ZHAO, F., WANG, L. X., & ZHANG, H. Incidence and risk factors of tuberculosis among the elderly population in China: a prospective cohort study. *Infectious diseases of poverty*, 2020, 9(1), 13. https://doi.org/10.1186/s40249-019-0614-9

[16] GEBREWELD, F. H., KIFLE, M. M., GEBREMICHEAL, F. E., SIMEL, L. L., GEZAE, M. M., GHEBREYESUS, S. S., MENGSTEAB, Y. T., & WAHD, N. G. Factors influencing adherence to tuberculosis treatment in Asmara, Eritrea: a qualitative study. *Journal of Health, Population, and Nutrition*, 2018, 37(1), 1. https://doi.org/10.1186/s41043-017-0132-y

参考文:

[1] 世界卫生组织，全球结核病报告。2023。

[2] WAHID W.W.C.、ARSY F.F.、SETIAWATI N.、VICRI R.N.、米兰达 A.V.、LESTARI B.W. 和 APRILLIA K.R.增加印度尼西亚结核病防治项目的融资。 RDI 政策简报，2023，6，20230619。

[3] RIZKANINGSIH R. 和 MUSTAFA M. 家庭环境物理状

况与结核病发病率的关系。预防促进杂志，2023， 6(2): 335-343。 https://doi.org/10.47650/jpp.v6i2.784
[4] REA, G., SPERANDEO, M., LIETO, R., BOCCHINO, M., QUARATO, C. M. I., FERAGALLI, B., VALENTE, T., SCIOSCIA, G., GIUFFREDA, E., FOSCHINO BARBARO, M. P., 和 LACEDONIA, D. 胸部影像学在肺结核诊断和治疗中的作用：胸部超声的补充作用。医学前沿，2021，8：753821。https://doi.org/10.3389/fmed.2021.753821
[5] VAPNIK V. N. 统计学习理论的本质。 Springer，N.Y.，2000。https://doi.org/10.1007/978-1-4757-3264-1
[6] SAIN H. 和 PURNAMI S. W. 结合抽样支持向量机进行不平衡数据分类. Procedia 计算机科学，2015，72：59–66，https://doi.org/10.1016/j.procs.2015.12.105
[7] CERVANTES J.、GARCIA-LAMONT F.、RODRÍGUEZ-MAZAHUA L. 和 LOPEZ A. 支持向量机分类综合调查：应用、挑战和趋势。 神经计算，2020，408：189–215，https://doi.org/10.1016/j.neucom.2019.10.118
[8] HAMED B. A.、IBRAHIM O. A. S. 和 ABD EL-HAFEEZ T. 使用机器学习技术优化模式匹配的分类效率。大数据杂志，2023，10(1)：124，https://doi.org/10.1186/s40537-023-00804-6
[9] FIBRIANTO A.、MAGDALENA R. 和 FUADAH Y.N. 使用人工神经网络反向传播对人类正常肺部疾病、结核病 (tbc) 和胸膜积液进行分类。电子工程学报，2018，5(3)：5071–5078。 https://core.ac.uk/reader/299924570
[10] WARDANI R.S.， 和 PURWANTO。基于外部监测任务属性选择的k近邻结核病诊断模型。。第二届大学研究研讨会论文集，2015，284-290。https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/1601
[11] SHOFIA E. N.、REGASARI R.、PUTRI M. 和 ARWAN A. 使用 K-最近邻 - 确定性因子方法诊断发热疾病的专家系统：登革热、疟疾和伤寒。信息技术与计算机科学发展杂志，2017，1(5)，426–435。 https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/122
[12] AINI N. 和 HATTA H. R. 诊断结核病的专家系统。Mulawarman 信息学：计算机科学科学期刊，2017，12(1)：56–63。 http://dx.doi.org/10.30872/jim.v12i1.224
[13] CUDAHY P. 和 SHENOI S.V 肺结核诊断。研究生医学杂志，2016，92（1086）：187–193，https：//doi.org/10.1136/postgradmedj-2015-133278
[14] LEWINSOHN，D.M.，LEONARD，M.K.，LOBUE，P.A.，COHN，D.L.，DALEY，C.L.，DESMOND，E.，KEANE，J.，LEWINSOHN，D.A.，LOEFFLER，A.M.，⋯ 和 WOODS，G.L. 美国胸科学会/美国传染病学会/疾病控制与预防中心官方临床实践指南：成人和儿童结核病诊断。临床传染病：美国传染病学会官方出版物，2017，64（2），e1–e33。https://doi.org/10.1093/cid/ciw694
[15] CHENG, J., SUN, Y. N., ZHANG, C. Y., YU, Y. L., TANG, L. H., PENG, H., PENG, Y., YAO, Y. X., HOU, S. Y., LI, J. W., ZHAO, J. M., XIA, L., XU, L., XIA, Y. Y., ZHAO, F., WANG, L. X., 和 ZHANG, H. 中国老年人口结核病发病率及危险因素：一项前瞻性队列研究。贫困传染病，2020 年，9(1)，13。https://doi.org/10.1186/s40249-019-0614-9
[16] GEBREWELD, F. H.、KIFLE, M. M.、GEBREMICHEAL, F. E.、SIMEL, L. L.、GEZAE, M. M.、GHEBREYESUS, S. S.、MENGSTEAB, Y. T. 和 WAHD, N. G. 影响厄立特里亚阿斯马拉结核病治疗依从性的因素：一项定性研究。《健康、人口与营养杂志》，2018，37(1)，1。https://doi.org/10.1186/s41043-017-0132-y

Word count: 3,447 words, excluding references.

Peer review information:
The reviewer reports submitted in first round: 3 reports
The reviewer reports submitted in 2nd round: 1 report
The revision rounds stages: 2 rounds
Final revised version submitted: January 10, 2025