Advanced Tuberculosis Detection System for Preserving Public Health

Dr. Nidhi Mishra[1], Ghorpade Bipin Shivaji[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.*

[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS

Tuberculosis, Public Health, Crow Search Driven Tuned Logistic Regression (CS-TLR).

ABSTRACT

Tuberculosis is a major public health threat that requires efficient and early detection technologies. Challenges associated with real-world implementation include information variability, model availability, and operations in different clinical circumstances. In this paper, we introduce a Crow Search Driven Tuned Logistic Regression (CS-TLR) tuberculosis identification system to sustain early identification of the disease and increase the levels of diagnostic accuracy to help sustain public health. The model attempts to enhance the overall prediction performance of logistic regression by optimizing its parameters by using CSA's optimization capability. We gathered the Kaggle dataset with a variety of clinical and demographic characteristics of tuberculosis patients. We evaluate the performance of the suggested method employing standard parameters including AUC (90.5%), precision (89.9%), recall (90.1%), and F1-score (89.7%). The findings suggest that it could be an important public health surveillance tool by enabling early diagnosis and treatment of TB.

1. Introduction

TB is transmitted easily by air and affects many people globally, which classifies it as a significant threat to public health with negative repercussions. Therefore, techniques that can help in early identification are essential to fight its spread and minimize the impact on communities globally. However, conventional strategies to diagnose the diseases can demand complex laboratory procedures, and sometimes they are unavailable, especially in rural settings or developing countries [1]. These include the enhancement of the systems of automatic TB detection that can significantly change the diagnosis of the disease and evaluation of the treatment in patients due to the availability of new technology and computational methods [15]. The technology seeks to enhance accuracy in the detections, accelerate the rate at which diseases are diagnosed, and enable medical practitioners to intervene at an early stage of the disease process due to the automation of this process [2]. With these precautions, the risk of contamination of TB is minimized amongst communities and the production for the patients involved is enhanced [12]. The system also solves the problems of delivery and infrastructure that can be problematic when diagnosing TB; it was designed to be effective in low resource settings [4]. With its conductive capability and accessible features, healthcare practitioners can enhance the quality of treatments to enhance their patient's quality of life and hasten the diagnosis process [3]. It tends to early diagnosis and treatment, which contributes to other goals of public health excluding illnesses and maintaining community health besides trying to tackle challenges that clinics encounter concerning TB [5]. The positive outcomes that can stem from the increased development and use of such technologies to reduce TB rates and increase healthcare equity are important to note, affirming the need for continued growth in diagnostics and medicine [6].

This paper introduced a CS-TLR TB detection system to establish early detection and accurate diagnosis to support to public health maintenance.

In section 2, the related works are presented. Section 3 indicates the methodology. Section 4 presents the result and the conclusion is presented in section 5.

Related works

A passive TB detection method based on an advanced deep learning (DL) model was proposed in the investigation of Nafisah and Muhammad [13]. Experimental findings support the proposition that segmented chest X-ray (CXR) images offered greater efficiency than raw chest CXR images.

A computer-aided DL-based integrated method was developed by Xie et al. [7] in the search for multiple types of tuberculosis on chest radiographs. The proposed computer-assisted system was better than the existing one, which could help public health professionals to screen for TB in endemic areas and enable radiologists to identify patients.

The examination confirmed that multidrug-resistant Mycobacterium tuberculosis (M.TB) posed a serious threat to public health around the world (Gröschel et al.) [8]. The motive of the Genomics platform (GenTB), a web application that was movable and had passed through benchmarking, changed into assisting public health and clinical practitioners in know-how the intricacy of M.TB whole-genome sequencing (WGS) facts [9].

The diagnosis of tuberculosis was a potentially difficult public health issue. The most popular imaging modality to evaluate pulmonary tuberculosis (PTB) was CXR, which assessed the effectiveness of machine learning (ML) for medical research inspired by recent advances in artificial intelligence (Parreira et al.) [14]. The diagnostic device was able to satisfactorily differentiate pulmonary (PTB) from other lung diseases and accurately detected and processed abnormal CXR.

M.TB, a kind of bacteria, was the infectious agent that caused TB Bakır and M Babalık [10] in the investigation. TB was indeed a major global public health concern, with a high prevalence in underdeveloped nations. A 98% success rate was attained with the suggested CNN model.

2. Methodology

Data collection

We gathered the dataset from the kaggle source Random Sample of NIH Chest X-ray Dataset (kaggle.com) there are 30,805 distinct patients' X-ray images (112,120) with disease labels included in this NIH chest X-ray database.

crow search-driven tuned logistic regression (CS-TLR)

Crow search algorithm (CSA)

A crow is thought to be the smartest bird in the world. Considering their bodily size, they have the largest brains. Their brain is positioned somewhat below the human brain in terms of brain-to-body ratio. Crow intelligence is demonstrated by a plethora of data. They are capable of manufacturing tools and have shown consciousness in mirror assessments. With their ability to recall faces, crows can alert one another to the approach of an intruder. They can also recall the hiding site of their food up to numerous months later, use instruments, and interact in complex ways.

It has been observed that crows examine other birds, find out where they keep their food, and then take it when that individual leaves. If a crow has stolen anything, it will move its hiding spots and take additional security measures to make sure it doesn't get stolen again. They can anticipate a thief's actions based on their personal experience as thieves, and they can figure out the best way to keep their caches safe from theft. The following is a compilation of CSA concepts.

1. Crows are a flock-living species.
2. Crows commit their hiding spots' locations to memory.
3. Crows follow one another when they steal.
4. Crows guard their reserves against potential thieves.

It is believed that there are several crows in a d-dimensional setting. The vector $w^{j,iter}(j = 1,2,\dots,M; iter = 1,2,\dots,iter_{max})$, where $w^{j,iter} = [w_1^{j,iter}, w_2^{j,iter}, \dots\dots, w_c^{j,iter}]$ $iter_{max}$ indicates here crow j is at any given time in the searching region and is the higher amount of iterations. Each crow possesses a retention in which it retains the location of its hidden spot. The location of the crow's hidden place is displayed by $n^{j,iter}$. Thus far, this is the best location j have been able to secure. Every crow indeed retains a memory of the location of its finest experience. Crows wander about their surroundings, looking for better locations to hide and eat.

Assume that the crow i wishes to access its hidden spot, $n^{i,iter}$ at iteration iter. At this iteration, crow j chooses to approach crow i hiding spot by following i. Two outcomes are possible in this situation:

Situation 1: Crow i is unaware that crow j is trailing it. Consequently, crow j determinations go closer to crow i hidden spot. In this instance, crow j new location is determined as follows:

$$w^{j,iter+1} = w^{j,iter} + q_j \times fl^{j,iter} \times \left(n^{i,iter} - w^{j,iter}\right) \tag{1}$$

In this situation, the flight duration of crow j at each iteration is indicated by $fl^{j,iter}$, and q_j is an arbitrary integer with normal dispersion among 0 and 1. Because of this condition and how fl affects search power, lower values of fl indicate local search $w^{j,iter}$, but large values produce global search $w^{j,iter}$.

Situation 2: Crow j follows it, and crow i is aware of this. Thus, to prevent its cache from existence lifted, crow i will false crow j by moving to a diverse location in the search region. In their entirety, situations 1 and 2 can be stated as follows:

$$w^{j,iter+1} = \begin{cases} w^{iter} + q_j \times fl^{j,iter} \times \left(n^{i,iter} - w^{j,iter}\right) & q_i \geq AP^{j,iter} \\ \qquad a\ random\ position & otherwise \end{cases} \tag{2}$$

Here $AP^{i,iter}$ indicates the conscious proportion of crow i at iteration iter, and q_i is a random integer with normal dispersion among 0 and 1.

Metaheuristic methods ought to offer a harmonious blend of intensification and diversity. In CSA, awareness probability (AP) is the primary control variable that governs intensification and diversity. When the AP value drops, CSA is more likely to focus its search efforts on a small area where an acceptable option is currently located. Thus, employing low AP values results in increased intensification. Conversely, as the AP value rises, there is a decreased probability of exploring in the vicinity of already viable solutions, and as a result, CSA tends to search globally (randomization). Consequently, using high AP values promotes variety.

TLR

The magnitude of independent factors that influence the probability of dependent factors occurring can be ascertained using the LR model, a statistical analytic tool for variable classification. Numerous studies on the factors contributing to TB and its modeling have made extensive use of this model. The contribution of the contributing factors to TB was estimated using a binary LR model. TB is defined by the model as setting the dependent parameter z as 1. For the explanatory parameters $w_1, w_2, \dots\dots w_m$, $O(z = 1|w_1, w_2, \dots\dots, w_m)$ or $O(z = 1|w_j)$ represents the chance O that TB will occur. This is the formulation for the traditional LR model.

$$logit(O) = In\left(\frac{O}{1-O}\right) = \propto + \sum_{j=1}^{m} \beta_j w_j \tag{3}$$

When TB occurs, i.e., when one parameter varies by one degree while all other independent parameters remain fixed, the logarithmic fluctuation in probabilities O is shown by β_j, the LR coefficient for every explaining parameter (EP), and the unchanged component α. The chance ratio of TB will rise proportionately with each increased value of the EP, according to the positive regression coefficient. In comparison, the negative regression coefficient suggests that for every extra value of the EP, the odds ratio of urban expansion will fall proportionately.

It built a TLR model for evaluating the factors driving TB using the field strength (FS) parameter, based on the prediction of the TLR model. The chance of an individual i in TB being transformed to $z = 1$ is represented by O_i, and it can be computed as follows:

$$O_i = O_i(z = 1|w_j, E_i) = \frac{\exp(\propto + \beta_j w_j + \delta E_i)}{1 + exp(\propto + \beta_j w_j + \delta E_i)} \quad (4)$$

The following formula can be used to create the TLR model:

$$logit(O_i) = In\left(\frac{O_i}{1 - O_i}\right) = \propto + \sum_{j=1}^{m} \beta_j w_j + \delta E_i \quad (5)$$

Where the remaining variables are identical to those in Equation 4 and δ are the FS parameters regression coefficient.

CS-TLR

In the TB detection framework, the hybridization of CS and TLR offers a remarkable expedient for early tuberculosis identification and management, ultimately assisting in the sustainability of public health globally. The amalgamation of CS and TLR capitalizes on the advantages of both methodologies: CS's flair for investigating varied parameter preparations and TLR's capability to maximize the predictive power of logistic regression. This CS-TLR method enhances prognosis accuracy and promotes public health by combining sophisticated optimization strategies with dependable type algorithms for early identification and remedy:

3. Results and discussion

A laptop running Windows 11 with an Intel i5 7th Gen processor, 16 GB of RAM, and a Python 3.10.1 environment are used to perform our proposed approach. The proposed method is evaluated in terms of recall, precision, area under the curve (AUC), and F1-score and compared with the existing approaches, which are Random Forest (RF), K-Nearest Neighborhood (KNN), and Support Vector Machine (SVM) [11].

The AUC measure is essential to assess the analytical validity of the program and its potential impact on public health through early detection and intervention methods. A comparison of AUC is presented in Figure 1. Our suggested FSO-RF approach performed (90.5%), in contrast to (86.1%), (84.8%), and (89.2%) of the current techniques such as SVM, KNN, and RF. The outcomes demonstrate that the suggested strategy outperforms existing methods.

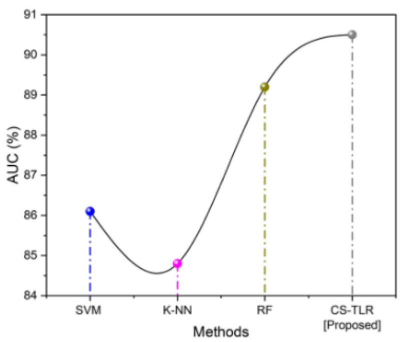

Figure 1 Result of AUC

Precision evaluates the system's capacity to accurately detect tuberculosis cases throughout all positive predictions. This is important to reduce false positives and guarantee proper diagnosis and treatment, which in turn effectively protects public health. A comparison of precision is presented in Figure 2. Our suggested FSO-RF approach performed (89.9%), in contrast to (78.4%), (83.9%), and (72.7%) of the current techniques such as SVM, KNN, and RF. The outcomes demonstrate that the suggested strategy outperforms existing methods.

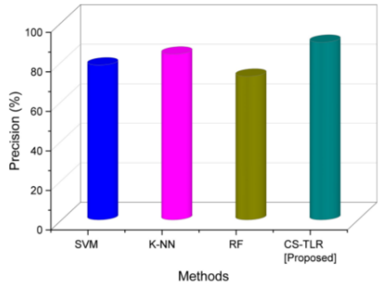

Figure 2 Result of precision

Recall assesses the system's capacity to identify all pertinent cases, which is important for public health since it guarantees early detection and treatment, reducing the risk of spread and enhancing patient outcomes. A comparison of recall is presented in Figure 3. Our suggested FSO-RF approach performed (90.1%), in contrast to (79.6%), (79.2%), and (78.2%) of the current techniques such as SVM, KNN, and RF. The outcomes demonstrate that the suggested strategy outperforms existing methods.

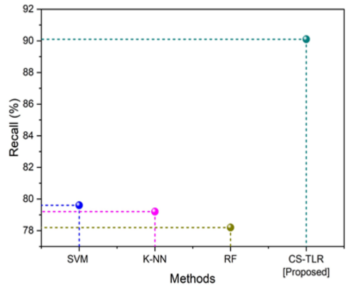

Figure 3 Result of recall

F1-score establishes a compromise between reducing false positives and accurately identifying TB cases in the system, which is essential for preserving the reliability and effectiveness of public health. A comparison of the f1-score is presented in Figure 4. Our suggested FSO-RF approach performed (89.7%), in contrast to (78.6%), (78.1%), and (78.3%) of the current techniques such as SVM, KNN, and RF. The outcomes demonstrate that the suggested strategy outperforms existing methods.

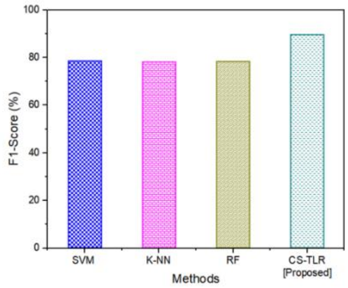

Figure 4 Result of f1-score

4. Conclusion and future scope

The present work introduces the CS-TLR tuberculosis detection method as a potentially effective strategy to address the public health threat posed by tuberculosis. Our approach provides important information using the Kaggle dataset that includes various clinical and demographic information about TB patients. With an AUC of 90.5%, precision of 89.9%, recall of 90.1%, and F1-score of 89.7%, the system showed excellent performance. Its effectiveness in early ailment identification and unique diagnosis both of which are critical for improving public fitness consequences is validated via these indicators. It may additionally come across difficulties with facts safety, algorithmic bias prevention, and the illustration of numerous networks. For public health interventions to be carried out efficiently, resilience throughout diverse TB signs and symptoms and demographic determinants ought to be ensured. Future scopes encompass extending to multi-modal statistics sources, enhancing real-time statistics integration, the usage of AI for early prognosis, combining telemedicine for remote locations, and ensuring scalable and reachable public fitness solutions.

Reference

[1] O. Hrizi, K. Gasmi, I. Ben Ltaifa, H. Alshammari, H. Karamti, M. Krichen, L. Ben Ammar, and M.A. Mahmood, “Tuberculosis disease diagnosis based on an optimized machine learning model”, *Journal* of Healthcare Engineering, *2022(1)*, p.8950243, 2022. https://doi.org/10.1155/2022/8950243

[2] K. Munadi, K., Muchtar, N. Maulina, and B. Pradhan, “Image enhancement for tuberculosis detection using deep learning”, *IEEE* Access, 8, pp.217897-217907, 2020. https://doi.org/10.1109/ACCESS.2020.3041867

[3] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. *Indian Journal of Information Sources and Services*, *14*(2), 146–152. https://doi.org/10.51983/ijiss-2024.14.2.21

[4] A. Wong, J.R.H.Lee, H. Rahmat-Khah, A. Sabri, A. Alaref, and H. Liu, “TB-Net: a tailored, self-attention deep convolutional neural network design for detection of tuberculosis cases from chest X-ray images”, *Frontiers in Artificial Intelligence, 5*, p.827299, 2022. https://doi.org/10.3389/frai.2022.827299

[5] P. Msonda, S.A. Uymaz, and S.S. Karaağaç, “Spatial pyramid pooling in deep convolutional networks for automatic tuberculosis diagnosis”, Traitement du Signal, 2020. https://doi.org/10.18280/ts.370620

[6] Alamer, L., Alqahtani, I. M., & Shadadi, E. (2023). Intelligent Health Risk and Disease Prediction Using Optimized Naive Bayes Classifier. Journal of Internet Services and Information Security, 13(1), 01-10.

[7] Y. Xie, Z. Wu, X. Han, H. Wang, Y. Wu, L. Cui, J. Feng, Z. Zhu, and Z. Chen, “Computer-Aided System for the Detection of Multicategory Pulmonary Tuberculosis in Radiographs”, *Journal of Healthcare Engineering, 2020(1)*, p.9205082, 2020. https://doi.org/10.1155/2020/9205082

[8] M.I. Gröschel, M.Owens, L. Freschi, R. Vargas, M.G. Marin, J. Phelan, Z. Iqbal, A. Dixit, and M.R. Farhat, “A user-friendly genome-based predictor for tuberculosis resistance powered by machine learning”, *Genome medicine*, *13*, pp.1-14, 2021. https://doi.org/10.1186/s13073-021-00953-4

[9] Sofiene, M., Souhir, C., Yousef, A., & Abdulrahman, A. (2023). Blockchain Technology in Enhancing Health Care Ecosystem for Sustainable Development. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14(3), 240-252.

[10] Ç. Bakır, and M. Babalık, “Detection of Tuberculosis Disease with Convolutional Neural Networks”, *In Proceedings of the International Conference on Advanced Technologies 11*, pp. 49-54, 2023. https://doi.org/10.58190/icat.2023.17

[11] V.V Chellam, V.Veeraiah, A. Khanna, T. Sheikh, S. Pramanik, and D. Dhabliya, “A Machine Vision-Based Approach for Tuberculosis identification in chest X-Rays images of patients”, *Lecture notes in networks and systems*, pp.23-32, 2023.

[12] A. Iqbal, M. Usman, and Z. Ahmed, “An efficient deep learning-based framework for tuberculosis detection using chest X-ray images”, *Tuberculosis*, 136, p.102234, 2022. https://doi.org/10.1016/j.tube.2022.102234

[13] S.I. Nafisah, and G. Muhammad, “Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence”, *Neural Computing and Applications*, *36(1)*, pp.111-131, 2024. https://doi.org/10.1007/s00521-022-07258-6

[14] P.L. Parreira, A.U. Fonseca, F. Soares, M.B. Conte, and M.F. Rabahi, “Chest X-ray evaluation using machine learning to support the early diagnosis of pulmonary TB”, The *International Journal of Tuberculosis and Lung Disease, 28(4)*, pp.171-175, 2024. https://doi.org/10.5588/ijtld.23.0230

[15] Bobir, A.O., Askariy, M., Otabek, Y.Y., Nodir, R.K., Rakhima, A., Zukhra, Z.Y., Sherzod, A.A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. *Natural and Engineering Sciences*, *9*(1), 72-83.