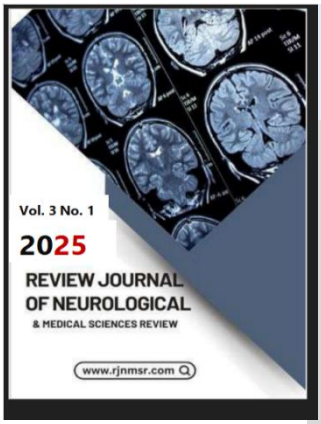


Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

Optimizing Tuberculosis Diagnosis: A Comparative Study of Machine Learning Algorithms and Feature Selection Methods Disorder: A Cross-Sectional Study

Abdul Qayoom Pitafi[1]
Statistical Officer, Sindh Bureau of Statistics, Karachi
Soomal Jhatial[2]
Department of Computer Science, University of Sindh, Jamshoro, Sindh, Pakistan
Kifayat Ullah[3]
Department of Mathematics and Statistics, Institute of Business Management Karachi
Shamim Jhatial[4]
Department of Statistics, University of Sindh, Jamshoro, Sindh, Pakistan
Muhammad Ismail[5*]
Department of Statistics, Quaid-i-Azam University, Islamabad Pakistan
Correspondence: ismailkhanmuhammad37@gmail.com

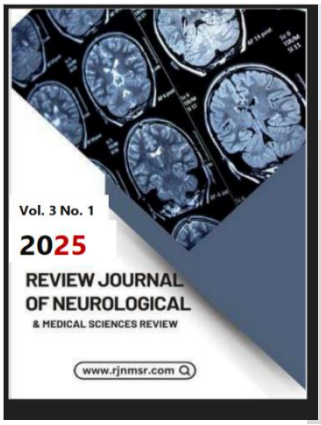
Abstract

The disease known as tuberculosis (TB) is extremely contagious and can be fatal if left untreated. When someone with tuberculosis coughs, sneezes, or speaks, airborne droplets are released into the air. Over the course of a year, one untreated TB patient can infect ten to fifteen others. This study investigates the risk factors associated with tuberculosis (TB) and explores the predictive capabilities of various machine learning algorithms. The research utilizes a dataset comprising 452 patient records from two hospital named as Mardan International Hospital and Khyber Hospital Mardan in district Mardan, encompassing 12 characteristics. The binary response variable differentiates between TB-positive and TB-negative cases. The study employs a range of machine learning techniques, including classification trees, random forests, k-nearest neighbors, random k-nearest neighbors, neural networks, and logistic regression. Feature selection was performed to identify the most relevant predictors. Model performance was evaluated using an independent test data set, assessing metrics such as classification accuracy, sensitivity, specificity, and kappa statistic. Despite its simplicity, the classification tree model demonstrated superior performance across most evaluation metrics compared to more complex algorithms, regardless of the number of selected features.

Keywords: Tuberculosis (TB), Machine Learning, Feature selection, Classification accuracy

Introduction

Tuberculosis (TB) derives its name from the Latin word "tubercula," meaning a small lump or nodule. This term refers to the small, scar-like lesions that form in the tissues of individuals infected with the disease. TB is a bacterial infection primarily

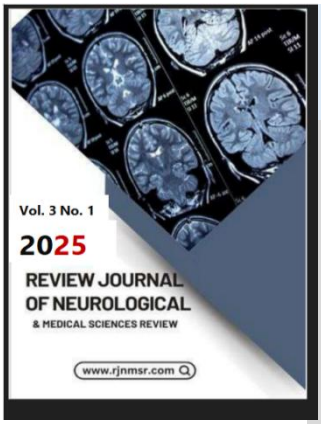


Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

affecting the lungs, although it can also spread to other parts of the body, including lymph nodes, kidneys, bones, and the brain [1]. Tuberculosis (TB) remains a significant global health concern, claiming an estimated 1.6 million lives annually [2]. It is estimated that around 25% of the world's population has been infected with the tuberculosis bacteria [3]. Tuberculosis (TB) remains a significant global health threat, contributing substantially to mortality worldwide. Notably, India, China, Indonesia, Pakistan, and the Philippines collectively bear a disproportionate burden of the disease, accounting for over 50% of global TB cases [4]. While TB has a long history, dating back to ancient times, encouraging progress has been made in recent decades. Since the year 2000, there has been a steady decline in the number of newly reported TB cases globally [5]. Tuberculosis (TB) primarily manifests in two forms: pulmonary TB, affecting the lungs, and extra pulmonary TB, impacting other organs. Pulmonary TB is the most common form, occurring in approximately 90% of cases [6]. Individuals with pulmonary TB may experience persistent cough, chest pain, and the production of phlegm (sputum). In severe cases, coughing up blood (hemoptysis) can occur [7]. Rarely, the infection can spread to the pulmonary artery, leading to significant bleeding. The upper lobes of the lungs are more frequently affected by TB due to factors like better airflow and potentially less efficient lymphatic drainage. Extra pulmonary tuberculosis (EPTB) is tuberculosis outdoor of the lungs [8]. This form is greater commonplace in people with weakened immune systems, along with those with HIV/AIDS, younger children, and those with sure persistent ailments. Extra pulmonary TB can have an effect on numerous organs, such as the pleura (lining of the lungs), relevant anxious device (brain and spinal cord), lymphatic device, genitourinary machine (kidneys and urinary tract), and bones and joints [9]. The causative agent of TB is Mycobacterium tuberculosis (MTB), a gradual-developing, aerobic bacterium. It has a completely unique outer membrane containing a waxy substance known as mycolic acid. While MTB can live to tell the tale and multiply inside host cells, it can additionally be cultured in laboratory settings. Several factors increase the danger of TB contamination. HIV infection is a major hazard thing globally, with about thirteen% of humans with TB also being inflamed with HIV [10]. This co-contamination is specifically regular in Sub-Saharan Africa. In the absence of HIV, the lifetime chance of developing active TB sickness from a prior infection is anticipated to be among 5-10%. Socioeconomic elements play a vital position in TB transmission. Overcrowding, malnutrition, and poverty boom the hazard of exposure and sickness development [11]. Individuals at excessive risk include people who inject capsules, residents of overcrowded settings (prisons, homeless shelters), those with restricted get entry to healthcare, sure ethnic minorities, kids in close contact with infected individuals, and healthcare people serving high-hazard populations [12]. Other tremendous risk elements encompass continual lung illnesses which includes persistent obstructive pulmonary sickness (COPD) and smoking (smokers have roughly double the chance of TB as compared to non-people who smoke). Additionally, situations like diabetes, alcohol abuse, and immunosuppressive treatments can boom susceptibility to TB infection [13]. Tuberculosis (TB) remains a sizeable public health task in Pakistan. A predicted



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

600,000 new TB cases are suggested yearly, rating Pakistan fifth globally in terms of TB burden. Tragically, over 70,000 individuals succumb to the disorder each yr. Around an envisioned 15000 cases of the youngsters is affected yearly from tuberculosis [14]. Pakistan is likewise predicted to have the 4th maximum incidence of multidrug-resistant. This poses a severe threat to public fitness and necessitates sturdy treatment and prevention strategies [15]. The Global Fund has been a critical supply of financial support for TB control applications in Pakistan, contributing about US four million between 2017 and 2018. These finances are channeled through each public and personal sectors, with NGOs such as Indus and Mercy Corps gambling crucial roles in software implementation. The budget is in most cases utilized by the provincial TB applications and other non-public carrier carriers [16]. The World Health Organization (WHO) and Stop TB Partnership offer important aid to the countrywide and provincial TB applications in Pakistan [17]. Pakistan, as different social areas, spends a small meager of its GDP on health including excessive shortage of healthcare experts [18]. With a health practitioner-to-affected person ratio of about 1:3000, get entry to fine healthcare remains constrained. This insufficient healthcare infrastructure exacerbates the unfold of infectious sicknesses, which include TB [19]. Addressing the TB epidemic in Pakistan requires a multi-pronged approach. This includes strengthening the healthcare gadget, growing get right of entry to best healthcare offerings, enhancing TB diagnosis and remedy, and enforcing powerful prevention techniques [20]. The objective of this paper is to investigate the threat factors associated with tuberculosis (TB) and to increase and examine predictive fashions for the disease the use of numerous systems learning strategies. The goal was to decide the best system getting to know algorithms (consisting of class timber, random forests, k-nearest pals, and others) for correctly predicting TB status. Feature selection become employed to pick out the most applicable predictors, and version overall performance was assessed the use of metrics inclusive of type accuracy, sensitivity, specificity, and kappa statistic. Also, an association between TB and threat elements are evaluated to test its importance.

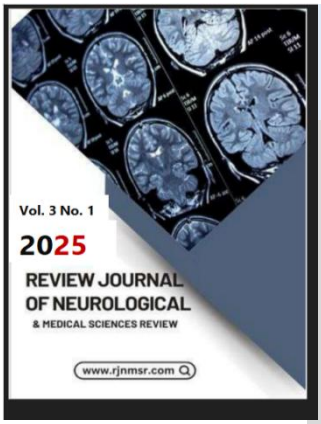
Purpose of the Study

The goal of this study is

- Identify and examine the most effective gadget mastering algorithms for predicting tuberculosis (TB) in patients.
- Determine the maximum critical chance elements related to TB based on patient characteristics.
- Develop and investigate predictive models for TB that could probably improve early prognosis and optimize useful resource allocation inside the healthcare device.

Materials and Methods

The main objective of the study was to identify the most likely risk factors associated with tuberculosis (TB) in patients arriving at two main hospitals named as Mardan International Hospital and Khyber Hospital Mardan in the Mardan division. A total of 452 patients were examined, and their personal and medical information was collected. The occurrence of TB in each patient was studied in relation to various potential risk factors, including balanced diet, smoking, diabetes, intestinal disorders,



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

exposure to sunlight, living in congested areas, family household income, and close contact with an infectious patient. Additionally, demographic variables such as age, gender, and residence were also considered in the analysis. The study aimed to provide a comprehensive understanding of the factors contributing to TB infection, which could assist in enhancing prevention and treatment strategies.

Variable Explanation

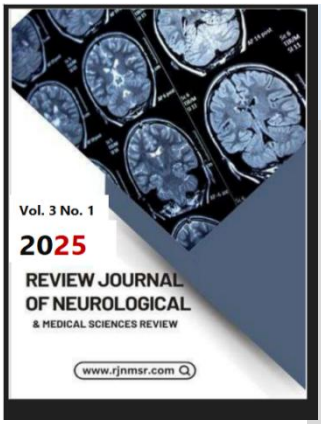
The study examined various variables to identify the most significant risk factors contributing to tuberculosis (TB) in patients. The following variables were used:

Response Variable

- Tuberculosis (TB): The response variable for this study indicates the presence or absence of TB in patients. TB was categorized into two groups: TB positive (coded as '1') and TB negative (coded as '0').

Independent Variables (Risk Factors)

- Residence: Classified as both urban or rural based totally on the affected person's vicinity of house. Urban turned into coded as '1', and rural as '2'.
- Gender: TB happens greater regularly in males in comparison to females, except at some stage in early life when women are more affected. Gender became classified as male (coded as '1') and lady (coded as '2').
- Age: In Pakistan, the majority of energetic TB instances are in the productive age group (15-59 years). Although age is a non-stop variable, it became divided into four classes for this study: 0-14 years (coded as '1'), 15-34 years (coded as '2'), 35-54 years (coded as '3'), and fifty-five years or older (coded as '4').
- Population of a Household: Overcrowded households, with as a minimum 3 people, boom the threat of TB transmission because of negative hygiene and air flow. Households with three or more participants have been labeled as overcrowded (coded as '1'), while people with two or fewer individuals have been categorized as not overcrowded (coded as '2').
- Economic Status of a Household: Economic condition is closely linked with TB, as poverty can boom vulnerability to contamination. Households with a monthly profit of Rs. 6000 or much less have been considered terrible (coded as '1'), even as those with higher incomes have been labeled as not bad (coded as '2').
- Diabetes: Diabetes is a giant comorbidity in TB sufferers, especially in low-income nations. Patients have been categorized as having diabetes (coded as '2') or no longer having diabetes (coded as '1').
- Smoking: Smoking is understood to increase the risk of TB and have an effect on treatment outcomes. This variable turned into categorized into people who smoke (coded as '1') and non-people who smoke (coded as '2').
- Diet: An unbalanced eating regimen can weaken the immune machine, growing vulnerability to TB. Patients with a wholesome weight loss program were coded as '1', and people with a terrible food plan have been coded as '2'.
- Medical Care: Access to exact hospital treatment is critical for coping with TB and its comorbidities. Good hospital therapy turned into coded as '1', and terrible hospital therapy as '2'.



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

- Living in a Refugee Camp: Living in unsanitary and crowded conditions, along with in refugee camps, increases the threat of TB infection. Patients dwelling in refugee camps have been coded as '1', at the same time as others were coded as '2'.
- Close Contact with an Infectious Patient: Close touch with someone infected with TB is a sizeable danger thing for growing the ailment. Patients with near contact were coded as '1', while those without near contact have been coded as '2'.

Methods

The term "Machine Learning" (ML) was first added with the aid of Arthur Samuel in 1959 [21]. Today, ML has turn out to be a fundamental approach for identifying patterns in massive datasets throughout numerous fields, along with facial popularity, scientific prognosis, image detection, banking, and marketing. ML makes use of exceptional techniques to analyze and interpret big quantities of statistics, locating underlying styles and building predictive fashions [22].

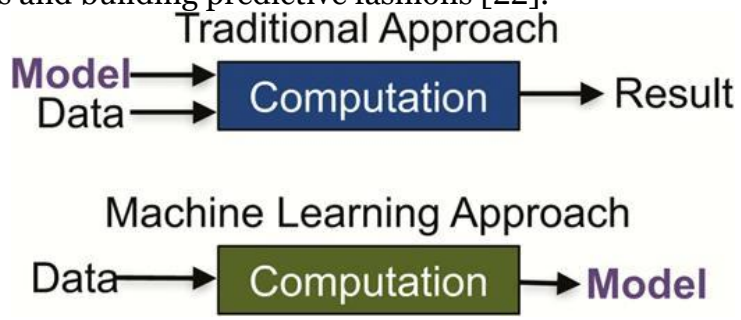


Figure 1 Comparison between the traditional approach and machine learning approach

Machine getting to know allows computers to analyze and improve from statistics without being explicitly programmed. Rather than difficult-coding regulations, ML algorithms build practical intelligence via using models trained on historic records to carry out responsibilities. The subject of ML is intently linked to statistics, with each focusing on getting to know from information and making predictions [23].

Types of Machine Learning

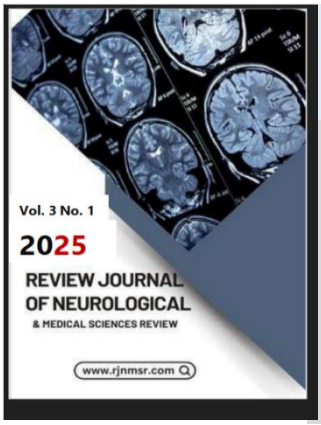
The Machine learning has classified into three types:

Supervised Learning

In supervised learning, the purpose is to predict the ideal final results primarily based on labeled input data [24]. Examples consist of classifying emails as unsolicited mail or non-spam or predicting whether or not affected person has a selected ailment primarily based on clinical facts. Supervised gaining knowledge of may be classified into:

- Classification: Predicts categorical outcomes, such as whether an email is spam.
- Regression: Predicts continuous variables, such as forecasting house prices.

Popular supervised learning algorithms include Linear Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Decision Trees.



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

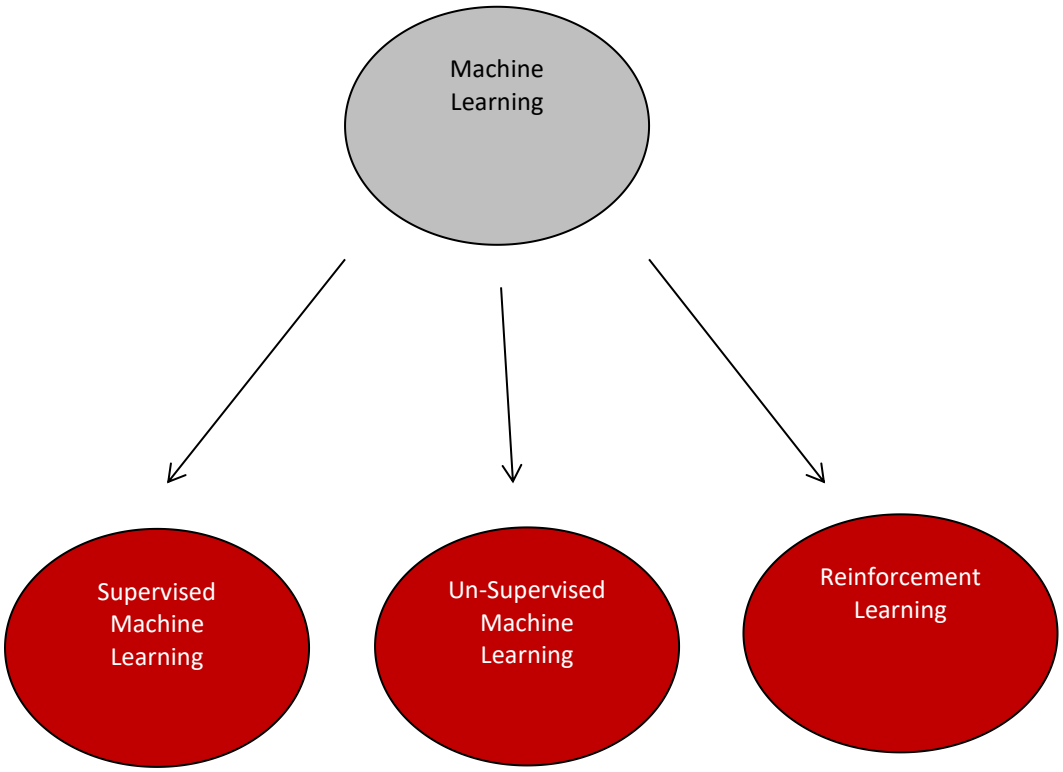


Figure 2 Types Of Machine Learning Approach

Unsupervised Learning

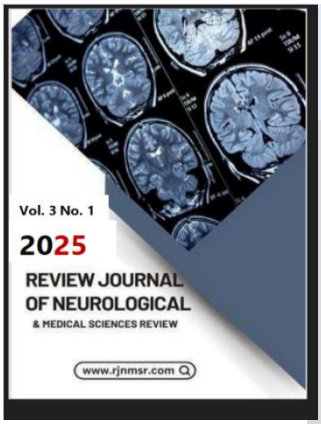
Unsupervised studying entails reading data without classified outcomes, aiming to find out hidden patterns [25]. Common obligations consist of clustering, in which similar information points are grouped collectively. Algorithms like K-means clustering, Self-Organizing Maps, and Hierarchical Clustering are examples of unsupervised learning.

Reinforcement Learning

In reinforcement mastering, the set of rules learns from interactions with its environment. It improves its overall performance through receiving comments on moves taken, adjusting destiny choices to maximize lengthy-time period rewards [26]. Examples encompass robotics and game AI.

In our study of tuberculosis (TB) threat prediction, numerous machine leaning has been used, each bringing specific strengths to the evaluation.

Logistic Regression (LR): Logistic Regression (LR) is a simple, yet powerful set of rules used for binary type obligations, making it properly perfect for predicting whether an person is liable to TB [27]. It models the probability of an individual belonging to a selected magnificence, primarily based on input functions which includes clinical records, socio-financial conditions, and demographic elements. The model uses a logistic feature (sigmoid) to output probabilities, and if the probability exceeds a fixed threshold, the person is classified as at risk. Logistic regression is



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

effective for its interpretability, as it gives clear insights into how each component influences the chance of TB.

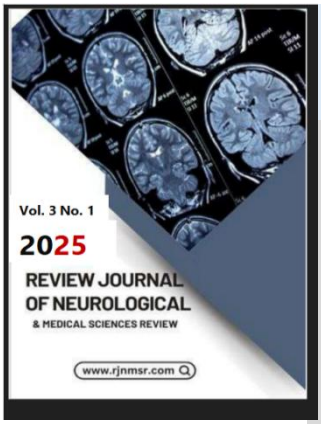
K-Nearest Neighbors (K-NN): K Nearest Neighbor is a plain set of rules that stores all to be had cases in addition to classifying new cases on the idea of similarity measures [28]. By calculating the distance between the enter statistics and each different facts factor, K-NN selects the k nearest friends and assigns the most frequent magnificence amongst them. This method works well in cases where the relationship between features and consequences is complicated and non-linear. However, K-NN may be computationally highly priced, particularly with huge datasets, and is touchy to noisy or inappropriate capabilities.

Decision Trees: Decision Trees are some other effective devices for class obligations, such as TB chance prediction. The selection tree splits the data into subsets based totally on feature values, developing a tree-like structure. Each internal node represents a choice about a characteristic, and each leaf node represents the final classification [29]. This algorithm is intuitive and clean to interpret, making it reachable for understanding how decisions are made. However, choice timber can effortlessly over fit if not nicely tuned, and they may no longer carry out properly with very noisy data.

Random Forests: Random Forests build upon selection bushes by means of developing an ensemble of a couple of decision bushes. Each tree is skilled on a random subset of the data and capabilities, and their outputs are blended to make the final prediction [30]. This approach reduces the overfitting problem normally encountered with character decision timber and improves predictive accuracy. Random forests are especially powerful for handling large datasets with complex interactions between capabilities, making them a notable choice for TB chance prediction.

Support Vector Machines (SVM): Support Vector Machine (SVM), or Support Vector Network, a supervised getting to know model with associated gaining knowledge of algorithms is used for class and regression analysis of data. The SMV set is the choice boundary line to maximize the distinction among the two classes [31]. This set of rules can cope with complicated datasets with difficult relationships among functions, making it nicely suitable for predicting TB risk. However, SVM can be sensitive to the choice of kernel and the parameters used, and it is able to not scale well with very large datasets.

Neural Networks (NN): Neural Networks (NN) are relatively flexible models stimulated by the human mind's structure, able to accomplish complicated patterns in information. They encompass layers of interconnected nodes (neurons), where every node performs a simple mathematical operation [32]. Neural networks are particularly powerful in detecting problematic, non-linear relationships among capabilities, which is treasured in TB risk prediction whilst a couple of interacting factors need to be considered. However, they require big amounts of information to carry out nicely and are computationally intensive. The "black-container" nature of neural networks can also cause them to less interpretable as compared to simpler fashions like logistic regression or decision trees.



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

By applying these diverse devices getting to know algorithms, we purpose to become aware of the simplest method for predicting TB threat, considering elements consisting of version accuracy, interpretability, and computational efficiency [33]. Several R programs are employed to aid the analysis, together with the randomForest package for building decision tree ensembles for class and regression, the kernlab and e1071 programs for implementing Support Vector Machines (SVM) for classification and regression, and the caret package deal for streamlining version schooling and assessment. Additionally, packages which includes FNN for instant nearest neighbor seek and Mlbench for benchmarking device learning troubles are used. The MASS package affords equipment for statistical analysis, while the SDM Tools bundle assists in visualizing and comparing model results. Each algorithm offers awesome advantages, and their mixed use will help ensure a robust and complete evaluation of TB threat factors, optimizing the fashions and performing correct predictions.

After applying these techniques, we evaluated their performance using several metrics:

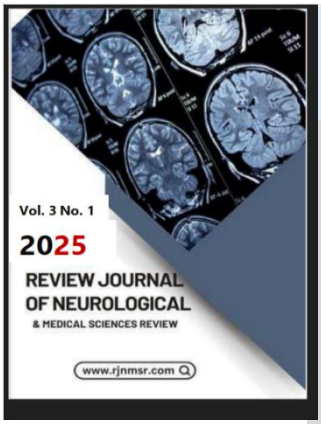
- Accuracy: The proportion of correctly classified instances (both TB positive and negative) out of the total instances.
- Specificity: The ability of the model to correctly identify TB-negative cases (true negatives).
- Sensitivity: The ability of the model to correctly identify TB-positive cases (true positives).
- Kappa Statistic: A measure of agreement between the predicted and actual classifications, adjusted for chance agreement.

Each of these metrics helped us assess and compare the predictive performance of the machine learning models used in the study.

Results

Table 1: Performance metrics for various machine learning algorithms under 70% training and 30% testing data

70% training and 30% testing data							
Features	**Metrics**	**Logistic**	**k-NN**	**Tree**	**RF**	**SVM**	**NN**
5	Accuracy	0.730	0.754	**0.772**	0.732	0.763	0.544
	Specificity	0.891	0.947	**0.995**	0.853	0.948	0.549
	Sensitivity	0.169	0.099	0.008	0.323	0.137	**0.526**
	Kappa	0.079	0.061	0.003	**0.184**	0.105	0.083
6	Accuracy	0.734	0.755	**0.769**	0.734	0.760	0.545
	Specificity	0.906	0.950	**0.995**	0.864	0.959	0.552
	Sensitivity	0.146	0.101	0.007	0.300	0.097	**0.517**
	Kappa	0.069	0.067	0.002	**0.177**	0.069	0.078
7	Accuracy	0.762	0.760	**0.774**	0.746	0.772	0.584



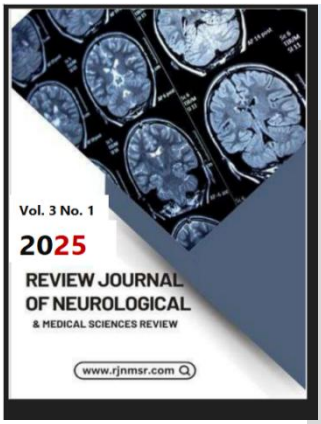
Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

	Specificity	0.943	0.947	**0.998**	0.874	0.954	0.606
	Sensitivity	0.145	0.123	0.003	0.310	0.164	**0.507**
	Kappa	0.111	0.091	0.000	**0.199**	0.150	0.121
	Accuracy	**0.773**	0.753	0.771	0.746	0.769	0.628
8	Specificity	0.950	0.943	**0.993**	0.879	0.957	0.681
	Sensitivity	0.173	0.106	0.009	0.297	0.134	**0.444**
	Kappa	0.156	0.064	0.003	**0.193**	0.116	0.135
	Accuracy	**0.775**	0.755	0.772	0.756	0.773	0.611
9	Specificity	0.947	0.944	**0.991**	0.890	0.972	0.660
	Sensitivity	0.184	0.104	0.014	0.296	0.088	**0.442**
	Kappa	0.166	0.063	0.005	**0.207**	0.082	0.110
	Accuracy	0.771	0.753	**0.773**	0.757	0.772	0.612
10	Specificity	0.945	0.943	**0.993**	0.897	0.978	0.663
	Sensitivity	0.183	0.101	0.010	0.275	0.064	**0.439**
	Kappa	0.161	0.057	0.003	**0.195**	0.058	0.108
	Accuracy	0.771	0.753	**0.772**	0.754	0.765	0.624
11	Specificity	0.942	0.945	**0.994**	0.900	0.980	0.684
	Sensitivity	0.182	0.097	0.011	0.256	0.028	**0.422**
	Kappa	0.159	0.056	0.005	**0.179**	0.011	0.110
	Accuracy	0.766	0.751	**0.769**	0.755	0.764	0.620
12	Specificity	0.937	0.945	**0.990**	0.904	0.978	0.671
	Sensitivity	0.186	0.091	0.015	0.249	0.034	**0.445**
	Kappa	0.154	0.048	0.006	**0.177**	0.016	0.119

The table 1 presents the performance metrics of various machine learning algorithms such as Logistic Regression, k-Nearest Neighbors (k-NN), Decision Tree (Tree), Random Forest (RF), Support Vector Machine (SVM), and Neural Networks (NN) on a 70% training and 30% testing data split, using feature sets ranging from 5 to 12. Accuracy, Specificity, Sensitivity, and Kappa are evaluated for each model. In terms of accuracy, Logistic Regression and Support Vector Machine generally perform the best, with accuracy peaking around 0.775 when using 9 features, while Neural Networks consistently show lower accuracy, ranging from 0.544 to 0.624. When looking at specificity, which measures the model's ability to correctly classify negative cases, k-NN and SVM excel, maintaining high values (up to 0.972), while Neural Networks show lower specificity values (around 0.549 to 0.684). Sensitivity, which reflects the ability to correctly identify positive instances, is highest for Random Forest, particularly with 5 features, though overall it remains lower than



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

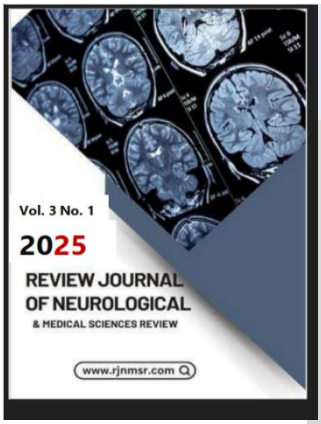
Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

other metrics across models. Neural Networks exhibit moderate sensitivity, peaking at 0.526 with 5 features, but again fall short compared to other models. Kappa, which corrects for random chance agreement, shows the best performance in Random Forest and SVM models, peaking at 0.207 and 0.150, respectively, while Neural Networks have the lowest Kappa values, indicating poorer agreement. In conclusion, Random Forest, k-NN, and SVM are the top-performing models in terms of accuracy and specificity, whereas Neural Networks consistently underperform, especially with respect to accuracy and sensitivity.

Table 2: Performance metrics for various machine learning algorithms under 50% training and 50% testing data

50% training and 50% testing data							
Features	**Metrics**	**Logistic**	**k-NN**	**Tree**	**RF**	**SVM**	**NN**
5	Accuracy	0.694	0.752	**0.771**	0.730	0.766	0.644
	Specificity	0.821	0.943	**0.988**	0.846	0.958	0.719
	Sensitivity	0.248	0.098	0.022	0.333	0.113	**0.379**
	Kappa	0.086	0.053	0.009	**0.186**	0.084	0.110
6	Accuracy	0.733	0.753	**0.770**	0.731	0.762	0.579
	Specificity	0.898	0.953	**0.991**	0.859	0.960	0.611
	Sensitivity	0.168	0.076	0.016	0.299	0.094	**0.463**
	Kappa	0.083	0.039	0.007	**0.168**	0.066	0.080
7	Accuracy	0.754	0.755	**0.771**	0.740	0.770	0.603
	Specificity	0.922	0.947	**0.990**	0.866	0.952	0.643
	Sensitivity	0.182	0.098	0.019	0.308	0.152	**0.456**
	Kappa	0.126	0.060	0.008	**0.188**	0.129	0.105
8	Accuracy	0.769	0.754	**0.770**	0.742	0.767	0.635
	Specificity	0.939	0.948	**0.988**	0.873	0.957	0.695
	Sensitivity	0.187	0.090	0.022	0.297	0.123	**0.422**
	Kappa	0.157	0.050	0.010	**0.185**	0.100	0.122
9	Accuracy	**0.772**	0.753	0.770	0.750	0.770	0.627
	Specificity	0.940	0.950	**0.989**	0.885	0.971	0.685
	Sensitivity	0.201	0.080	0.023	0.288	0.082	**0.424**
	Kappa	0.175	0.040	0.011	**0.193**	0.071	0.114
10	Accuracy	0.769	0.754	**0.770**	0.750	0.769	0.647
	Specificity	0.935	0.948	**0.986**	0.892	0.975	0.722
	Sensitivity	0.198	0.083	0.026	0.264	0.061	**0.384**
	Kappa	0.164	0.042	0.012	**0.176**	0.048	0.111
11	Accuracy	0.766	0.752	**0.770**	0.747	0.765	0.617
	Specificity	0.933	0.949	**0.988**	0.892	0.978	0.665
	Sensitivity	0.197	0.078	0.023	0.254	0.038	**0.442**
	Kappa	0.160	0.037	0.010	**0.166**	0.021	0.108
12	Accuracy	0.765	0.753	**0.771**	0.752	0.766	0.632
	Specificity	0.929	0.949	**0.990**	0.898	0.977	0.695
	Sensitivity	0.204	0.079	0.019	0.255	0.045	**0.413**



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

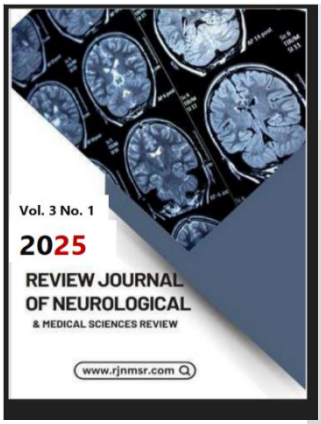
E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

	Kappa	0.162	0.039	0.008	**0.175**	0.028	0.109

In Table 2, where 50% training and 50% testing data is used, the Logistic Regression, k-Nearest Neighbors (k-NN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Neural Networks (NN) models were evaluated based on accuracy, specificity, sensitivity, and the Kappa statistic. The Decision Tree (Tree) model consistently shows the highest accuracy, reaching a peak value of 0.771 at 5 features. It also performs best in terms of specificity, with a maximum value of 0.988 at 5 features. While the Neural Networks (NN) model excels in sensitivity, reaching the highest value of 0.379 at 5 features, indicating its ability to detect positive cases more effectively compared to other models [34-36]. As the number of features increases, the Decision Tree (Tree) remains strong, with accuracy and specificity values near 0.770 and 0.99, respectively, at 6, 7, and 8 features. The k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) models perform similarly in terms of accuracy and specificity, with slight variations. The Random Forest (RF) model is the leader in terms of the Kappa statistic, a measure of agreement beyond chance, showing the best performance across all feature sets. At 9 features, Logistic Regression achieves the highest accuracy of 0.772, while the Decision Tree (Tree) and Random Forest (RF) models maintain accuracy values of 0.770. The specificity for the Decision Tree (Tree) is 0.989, and the sensitivity for the Neural Networks (NN) model is 0.424, demonstrating its high capability in identifying positive instances. The Kappa statistic for Random Forest (RF) increases to 0.193, indicating improved performance with more features. In scenarios with 10, 11, and 12 features, the Decision Tree (Tree) maintains a steady accuracy around 0.770, while the Random Forest (RF) continues to show the best Kappa statistic, ranging from 0.166 to 0.176. The Neural Networks (NN) model continues to have the highest sensitivity, peaking at 0.442 at 11 features. In summary, while Logistic Regression performs best in terms of accuracy at 9 features and Neural Networks (NN) shows the highest sensitivity, the Decision Tree (Tree) and Random Forest (RF) models are the most consistent performers across different metrics. Decision Tree (Tree) is favored for its high accuracy and specificity, especially with fewer features, while Random Forest (RF) stands out for its superior Kappa statistic, making both models the top choices for this study.

Table 3: Performance metrics for various machine learning algorithms under 30% training and 70% testing data

Features	Metrics	Logistic	*k*-NN	Tree	RF	SVM	NN
5	Accuracy	0.714	0.748	**0.768**	0.720	0.763	0.596
	Specificity	0.852	0.946	**0.985**	0.843	0.956	0.636
	Sensitivity	0.239	0.072	0.029	0.304	0.111	**0.459**
	Kappa	0.107	0.024	0.012	**0.155**	0.079	0.099
6	Accuracy	0.733	0.746	**0.769**	0.722	0.762	0.608
	Specificity	0.887	0.944	**0.986**	0.851	0.954	0.662
	Sensitivity	0.203	0.069	0.027	0.283	0.105	**0.422**



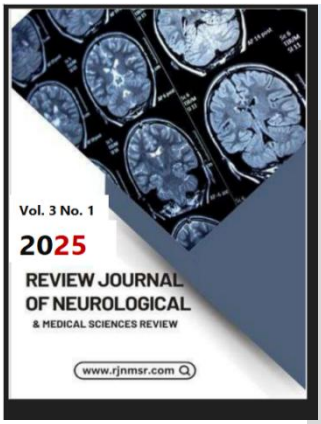
Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

	Kappa	0.106	0.017	0.010	**0.142**	0.070	0.085
	Accuracy	0.749	0.747	**0.768**	0.736	0.767	0.625
7	Specificity	0.909	0.946	**0.988**	0.867	0.957	0.686
	Sensitivity	0.206	0.069	0.021	0.290	0.123	**0.419**
	Kappa	0.135	0.020	0.007	**0.171**	0.097	0.105
	Accuracy	0.761	0.746	**0.767**	0.737	0.765	0.641
8	Specificity	0.918	0.943	**0.981**	0.871	0.957	0.708
	Sensitivity	0.226	0.072	0.035	0.283	0.114	**0.411**
	Kappa	0.170	0.019	0.014	**0.167**	0.088	0.119
	Accuracy	0.763	0.749	**0.768**	0.743	0.766	0.628
9	Specificity	0.920	0.947	**0.983**	0.882	0.967	0.687
	Sensitivity	0.224	0.069	0.030	0.270	0.075	**0.423**
	Kappa	0.171	0.022	0.013	**0.169**	0.055	0.110
	Accuracy	0.759	0.749	**0.768**	0.744	0.766	0.646
10	Specificity	0.915	0.949	**0.984**	0.887	0.972	0.719
	Sensitivity	0.229	0.063	0.027	0.255	0.059	**0.396**
	Kappa	0.168	0.016	0.009	**0.159**	0.040	0.112
	Accuracy	0.755	0.749	**0.768**	0.743	0.763	0.661
11	Specificity	0.911	0.950	**0.986**	0.893	0.973	0.746
	Sensitivity	0.222	0.061	0.024	0.234	0.049	**0.368**
	Kappa	0.157	0.015	0.009	**0.144**	0.029	0.113
	Accuracy	0.755	0.749	**0.770**	0.746	0.765	0.662
12	Specificity	0.907	0.950	**0.990**	0.896	0.976	0.745
	Sensitivity	0.237	0.063	0.018	0.236	0.045	**0.374**
	Kappa	0.167	0.018	0.006	**0.151**	0.028	0.117

In Table 3, Logistic Regression, k-Nearest Neighbors (k-NN), Decision Tree (Tree), Random Forest (RF), Support Vector Machine (SVM), and Neural Networks (NN) were trained using 30% of the dataset for training and the remaining 70% for testing. The table compares the models using various performance metrics: accuracy, specificity, sensitivity, and Kappa statistic, with different numbers of features (ranging from 5 to 12). From the results, the Decision Tree model consistently achieved the highest accuracy for any number of features, with a maximum accuracy of 77% when 12 features were used. The specificity for the Decision Tree ranged between 98.1% and 99%, outperforming the other models. Both k-NN and SVM showed similar performance, with slight variations in accuracy and specificity [37-39]. Neural Networks had the highest sensitivity, reaching 45.9% when 5 features were used. However, its sensitivity decreased as the number of features increased. Random Forest performed best in terms of the Kappa statistic, peaking at 0.171 when 7 features were used. This indicates that Random Forest provided relatively stable performance across the different feature sets. Overall, the Decision Tree model was the top performer in terms of accuracy and specificity, especially with 12 features, while Neural Networks excelled in sensitivity, and Random Forest provided the best Kappa statistic for 7 features [40].



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

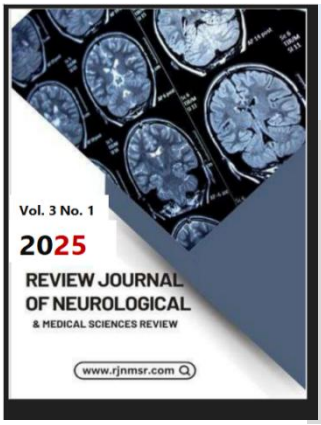
Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

Table 4: Distribution of TB Cases by Various Risk Factors

Risk Factor	Category	TB Positive	TB Negative	Total
Gender	Female	217	26	243
	Male	133	76	209
Age	0-14	41	12	63
	15-34	127	60	187
	35-54	92	22	114
	55+	78	20	88
Residential Area	Urban	261	60	321
	Rural	89	42	131
Intestinal Disorder	Yes	285	88	373
	No	65	14	79
Family Income	Low	325	89	414
	Middle	20	8	28
	High	5	5	10
Smoking	Yes	21	8	29
	No	329	94	423
Balanced Diet	Poor Diet	293	76	369
	Healthy Diet	57	26	83
Exposure to Sunlight	Yes	135	67	202
	No	215	35	250
Congested Living Area	Yes	268	67	335
	No	82	35	117
Diabetes	Yes	79	25	104
	No	271	77	348
Contact with Infectious Patient	Close Contact	239	68	307
	No Contact	111	34	145
HIV	Yes	11	1	12
	No	339	101	440

The table 4 presents a comparison of TB-positive and TB-negative cases across various risk factors such as gender, age, residential area, and health-related conditions. It shows that females are more likely to test positive for TB than males,



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

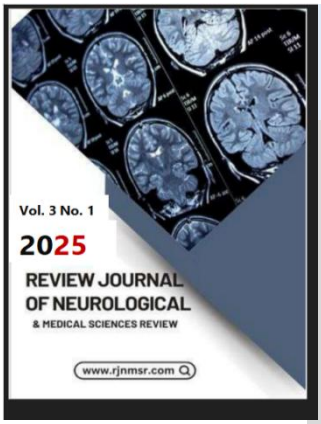
E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

with 217 females testing positive compared to 133 males. Age also plays a significant role, as individuals aged 15-34 and 35-54 form the majority of TB-positive cases. Urban areas account for a larger portion of TB-positive individuals compared to rural areas. Other prominent factors include intestinal disorders, low family income, and smoking, all of which contribute to higher rates of TB positivity. Health and lifestyle factors also show significant correlations with TB. Those following a poor diet, lacking exposure to sunlight, living in congested areas, and having close contact with infectious TB patients are more likely to test positive. The presence of diabetes and HIV further increases the likelihood of testing positive for TB [41]. Overall, the data highlights that TB is more prevalent among individuals with lower socioeconomic status, unhealthy living conditions, and existing health complications, suggesting that improving living conditions and addressing health issues could help reduce TB incidence.

Table 5: Statistical Analysis of Various Risk Factors

Risk Factor	Category	Chi-square	P-value	D.F	Odds Ratio
Gender	Male vs Female	42.35	0.000	1	4.77
Age	0-14, 15-34, 35-54, 55+	6.529	0.317	3	-
Residential Area	Urban vs Rural	9.52	0.002	1	2.05
Intestinal Disorder	Yes vs No	1.286	0.257	1	0.70
Family Income	Low, Middle, High	51.5	0.002	2	-
Smoking	Yes vs No	0.446	0.504	1	0.75
Balanced Diet	Poor Diet vs Healthy Diet	4.46	0.030	1	1.76
Exposure to Sunlight	Yes vs No	23.49	0.000	1	0.33
Congested Living Area	Yes vs No	4.88	0.027	1	1.71
Diabetes	Yes vs No	3.457	0.047	1	0.90
Contact with Infectious Patient	Close Contact vs No Contact	0.95	0.758	1	1.08
HIV	Yes vs No	1.429	0.232	1	3.28

The table 5 summarizes the statistical analysis of various risk factors associated with a disease. Significant associations were found for gender, residential area, family income, balanced diet, exposure to sunlight, congested living area, and diabetes. Males are 4.77 times more likely to be affected than females, while urban residents have 2.05 times higher odds compared to rural ones. A poor diet and congested living conditions increase risk by 1.76 and 1.71 times, respectively, and less sunlight exposure is associated with a higher risk (odds ratio = 0.33). Family income also shows a significant relationship with the disease. While diabetes has a borderline effect (odds ratio = 0.90), age, intestinal disorders, smoking, contact with infectious patients, and HIV status show no significant associations. Figure 3 presents multiple pie charts illustrating various socio-demographic and health-related factors



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

associated with TB-positive cases. The data highlights key risk factors such as gender, where females account for 62% of cases, while males make up 38%. Age distribution shows that most TB-positive individuals fall within the 15-34 age range (37.6%), followed by the 35-54 group (27.2%). A significant majority (74.6%) reside in urban areas. Other notable factors include low family income (92.9%), poor diet (83.7%), and congested living conditions (76.6%). Additionally, 68.3% of patients had close contact with infectious individuals, while only a small percentage (6%) reported smoking.

Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

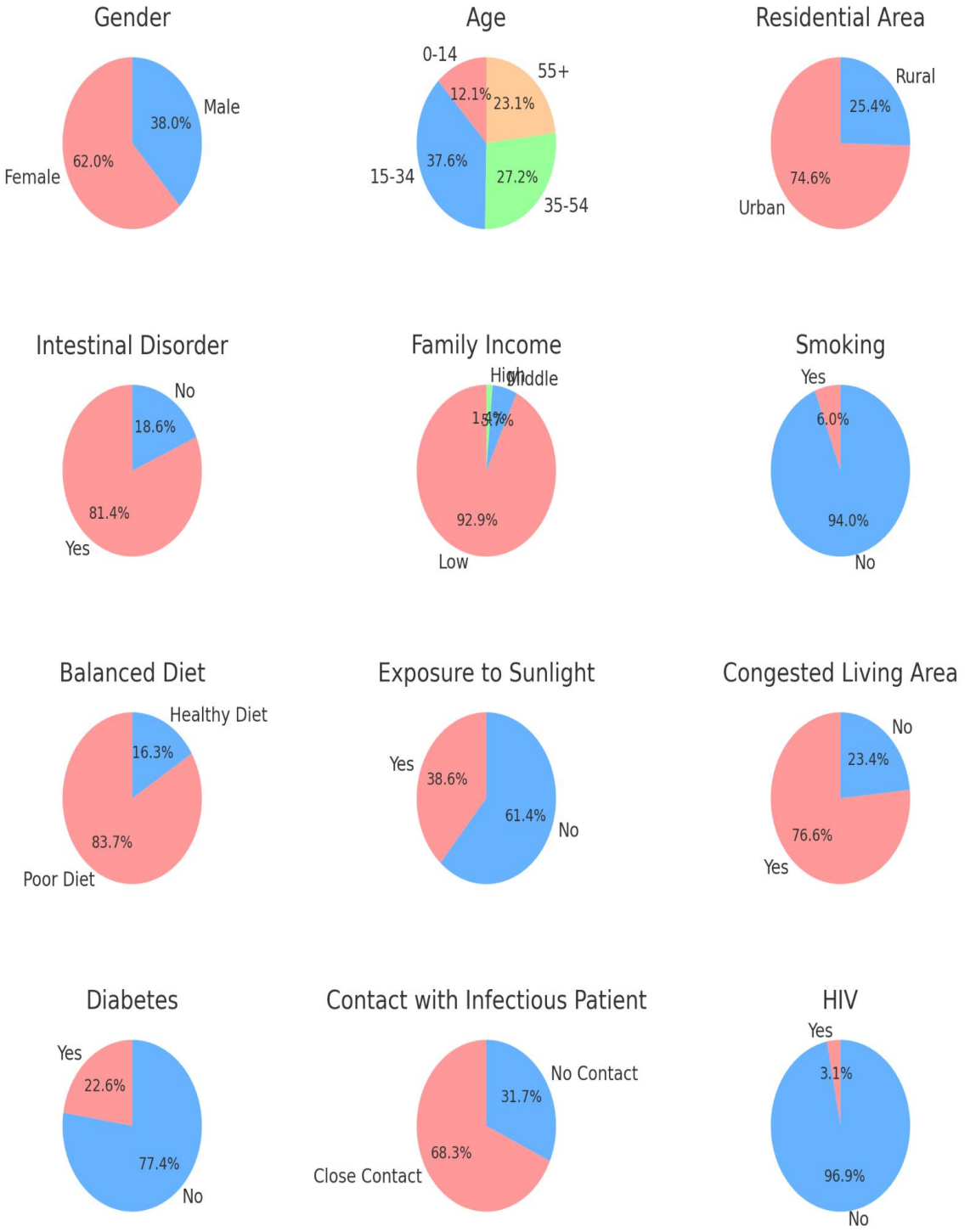


Figure 3 Socio-demographic and Health-related Factors Among TB Positive Cases

Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

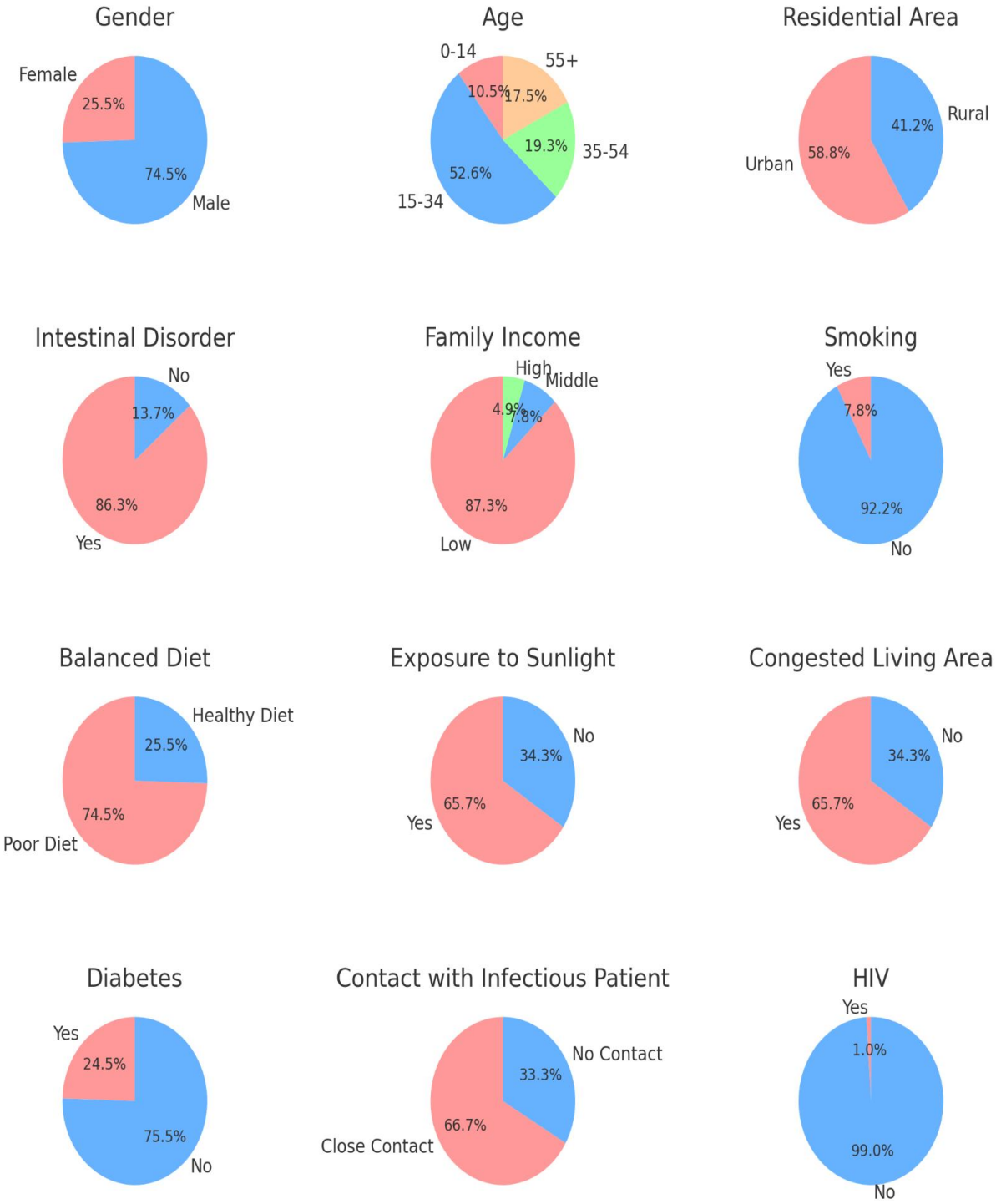


Figure 4 Socio-demographic and Health-related Factors Among TB Negative Cases

Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

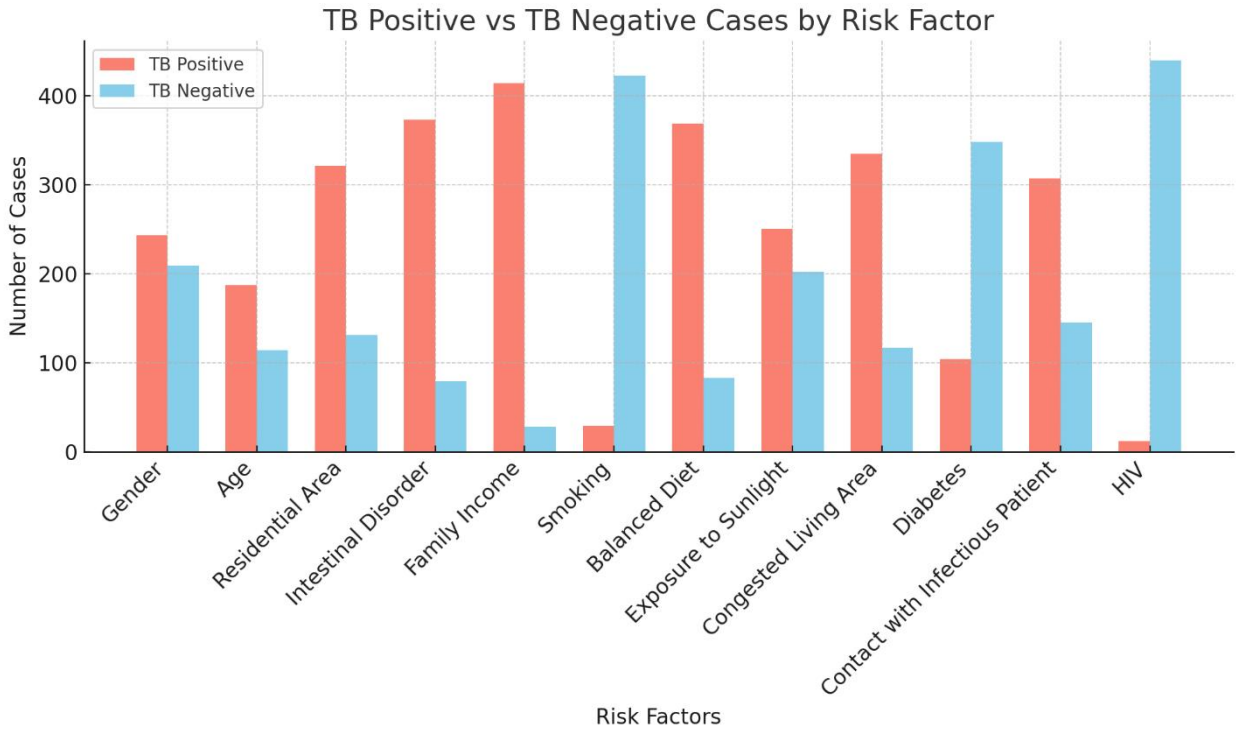
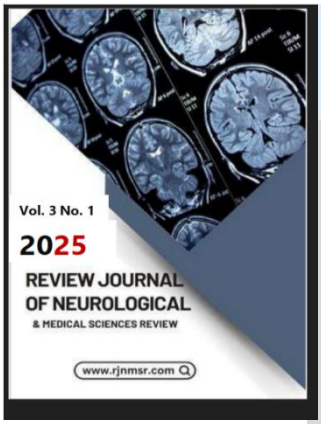


Figure 5 Comparison of TB Positive vs TB Negative Cases Across Various Risk Factors.

Figure 4 displays pie charts showing the distribution of various socio-demographic and health-related factors among individuals who are TB-negative. In this group, males represent the majority (74.5%), and most individuals are between the ages of 15-34 (52.6%). A large proportion (58.8%) live in urban areas. Additional factors include a low prevalence of intestinal disorders (86.3% without the condition), low family income (87.3%), and a poor diet (74.5%). A significant number of individuals are exposed to sunlight (65.7%), while a smaller percentage live in congested areas (65.7%). Only 7.8% of individual's smoke, and 66.7% had close contact with infectious patients. The prevalence of diabetes is low (24.5%), and almost none of the individuals are HIV positive (1%).

The bar graph in figure 5 compares the number of TB-positive and TB-negative cases across various risk factors. For gender, more males are TB-negative, while age shows TB-positive cases peaking for younger individuals. Urban areas have more TB-negative cases, while rural areas are more associated with TB-positive cases. Those with intestinal disorders and low family income show higher TB-positive cases. Smoking, poor diet, and living in congested areas are also associated with more TB-positive cases. TB-negative cases are more common among individuals with no exposure to sunlight, no contact with infectious patients, and those with no diabetes or HIV, although TB-negative cases dominate among non-HIV individuals.



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

Conclusions

This study aimed to identify the most significant risk factors for tuberculosis (TB) and evaluate the performance of various machine learning algorithms in predicting TB cases. Using data collected from 452 patients from Mardan International Hospital and Khyber Hospital Mardan across the Mardan division, we analyzed TB occurrence in relation to multiple risk factors, including diet, smoking, diabetes, exposure to sunlight, living conditions, income level, and close contact with TB patients, as well as demographic variables such as age, gender, and residence. The analysis showed that TB is significantly influenced by factors such as gender, residence, income, sunlight exposure, diet, and contact with infectious individuals. Notably, females were found to be more susceptible to TB than males, and patients living in urban areas had a higher incidence of TB compared to those in rural areas. Several machine learning algorithms such as logistic regression, k-nearest neighbors (k-NN), support vector machines (SVM), artificial neural networks (ANN), decision trees, and random forests (RF) were tested to evaluate their performance. Logistic regression performed well in terms of accuracy when using a higher number of features, while decision tree and random forest models exhibited strong performance in terms of specificity and Kappa statistics. The neural network model demonstrated the highest sensitivity in predicting TB cases. Based on the findings, the decision tree and random forest models were identified as the best-performing algorithms for predicting TB cases, particularly in terms of accuracy, specificity, and the Kappa statistic. However, the logistic regression model also proved to be a strong candidate in terms of overall accuracy. This study highlights the potential of machine learning algorithms to aid in the early prediction of TB, helping medical professionals better understand the factors contributing to TB incidence. Further research using larger datasets is recommended to refine these predictions and provide deeper insights into TB risk factors in Pakistan. These findings may ultimately inform public health strategies and interventions aimed at controlling the spread of TB.

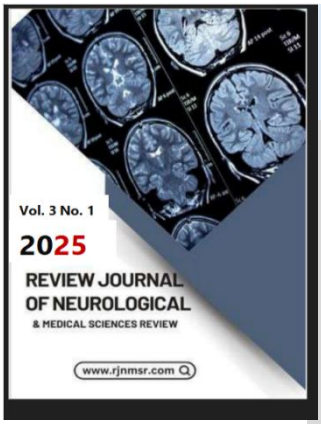
Author Contributions: All authors contributed equally to the conception, design, methodology, data analysis, writing, and review of this manuscript. Each author played an integral role in ensuring the accuracy, completeness, and scientific rigor of the study. All authors have read and approved the final version of the manuscript.

Data Availability Statement: Data will be provided upon request

Conflicts of Interest: The authors declare no conflicts of interest

References

1. M. H. Arsyad, I. Syafina, H. Hapsah, and H. Hervina, “Knowing and Understanding the Tuberculosis (Tb) Disease of the Lung (Literature Review),” Int. J. Nat. Sci. Stud. Dev., vol. 1, no. 2, pp. 56–85, 2024.
2. W. H. Organization, “National tuberculosis prevalence surveys 2007-2016,” 2021.
3. W. H. Organization, Global tuberculosis report 2024. World Health Organization, 2024.

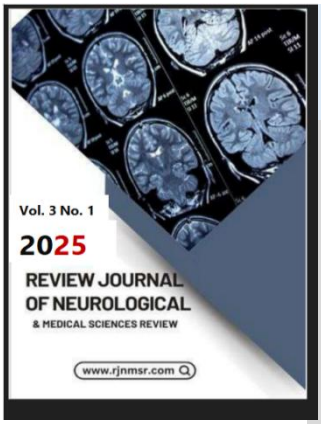


Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

4. B. Baral, K. Mamale, S. Gairola, C. Chauhan, A. Dey, and R. K. Kaundal, "Infectious diseases and its global epidemiology," in Nanostructured Drug Delivery Systems in Infectious Disease Treatment, Elsevier, 2024, pp. 1–24.
5. P. J. Dodd, C. M. Yuen, S. M. Jayasooriya, M. M. van der Zalm, and J. A. Seddon, "Quantifying the global number of tuberculosis survivors: a modelling study," Lancet Infect. Dis., vol. 21, no. 7, pp. 984–992, 2021.
6. R. Gopalaswamy, V. N. A. Dusthackeer, S. Kannayan, and S. Subbian, "Extrapulmonary tuberculosis—an update on the diagnosis, treatment and drug resistance," J. Respir., vol. 1, no. 2, pp. 141–164, 2021.
7. I. Pavord, N. Petousi, and N. Talbot, "Respiratory Disease," in Medicine for Finals and Beyond, CRC Press, 2022, pp. 155–202.
8. C. Giannessi et al., "Behçet's disease: a radiological review of vascular and parenchymal pulmonary involvement," Diagnostics, vol. 12, no. 11, p. 2868, 2022.
9. S. Mandal, P. Biswas, W. Ansar, P. Mukherjee, and J. J. Jawed, "Tuberculosis of the central nervous system: Pathogenicity and molecular mechanism," in A Review on Diverse Neurological Disorders, Elsevier, 2024, pp. 93–102.
10. E. I. Obeagu and E. C. Onuoha, "Tuberculosis among HIV patients: a review of Prevalence and Associated Factors," Int. J. Adv. Res. Biol. Sci, vol. 10, no. 9, pp. 128–134, 2023.
11. N. G. Etim, Y. Mirabeau, A. Olorode, and U. Nwodo, "Risk Factors of Tuberculosis and Strategies for Prevention and Control," Int J Innov. Healthc. Res, vol. 12, no. 1, pp. 1–13, 2024.
12. E. Cowan, M. R. Khan, S. Shastry, and E. J. Edelman, "Conceptualizing the effects of the COVID-19 pandemic on people with opioid use disorder: an application of the social ecological model," Addict. Sci. Clin. Pract., vol. 16, pp. 1–6, 2021.
13. G. Cáceres, R. Calderon, and C. Ugarte-Gil, "Tuberculosis and comorbidities: treatment challenges in patients with comorbid diabetes mellitus and depression," Ther. Adv. Infect. Dis., vol. 9, p. 20499361221095830, 2022.
14. N. A. Pradhan, R. Najmi, and Z. Fatmi, "District health systems capacity to maintain healthcare service delivery in Pakistan during floods: a qualitative study," Int. J. Disaster Risk Reduct., vol. 78, p. 103092, 2022.
15. S. A. H. Zaidi, M. Shahbaz, F. Hou, and Q. Abbas, "Sustainability challenges in public health sector procurement: an application of interpretative structural modelling," Socioecon. Plann. Sci., vol. 77, p. 101028, 2021.
16. A. A. Khan et al., "PREVALENCE OF RIFAMPICIN RESISTANCE AND PROBES IDENTIFICATION OF 81BP RRDR RPO-β GENE IN PULMONARY TUBERCULOSIS POPULATION OF DISTRICT BANNU, PAKISTAN.," Gomal J. Med. Sci., vol. 20, no. 2, 2022.
17. B. T. Shaikh, A. K. Laghari, S. Durrani, A. Chaudhry, and N. Ali, "Supporting tuberculosis program in active contact tracing: a case study from Pakistan," Infect. Dis. Poverty, vol. 11, no. 02, pp. 72–76, 2022.
18. G. N. Kazi, K. B. Mohamud, A. Quadir, S. K. Shah, and Z. ul Haq, "Tuberculosis control in Pakistan: A decade (2011-2020) in review," Pakistan J. Public Heal., vol. 12, no. 1, pp. 17–22, 2022.

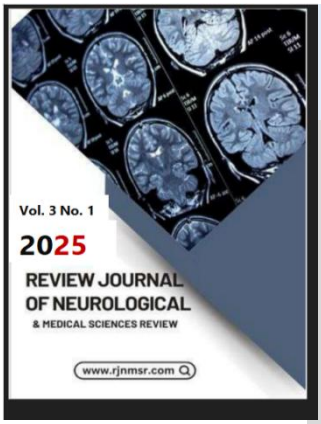


Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

19. S. Saleem, "Power, politics, and public health: understanding the role of healthcare expenditure in shaping health outcomes in Pakistan for policy enhancement," Politica, vol. 2, no. 1, pp. 58–72, 2023.
20. F. Malik and J. Creswell, "Innovative approaches to end TB in Pakistan: a review of TB REACH projects from 2010 to 2020," Pakistan J. Public Heal., vol. 11, no. 2, pp. 62–73, 2021.
21. A. L. Samuel, "Some studies in machine learning using the game of checkers," IBM J. Res. Dev., vol. 44, no. 1.2, pp. 206–226, 2000.
22. B. Lantz, Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd, 2019.
23. I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN Comput. Sci., vol. 2, no. 3, p. 160, 2021.
24. V. Nasteski, "An overview of the supervised machine learning methods," Horizons. b, vol. 4, no. 51–62, p. 56, 2017.
25. M. Usama et al., "Unsupervised machine learning for networking: Techniques, applications and research challenges," IEEE access, vol. 7, pp. 65579–65615, 2019.
26. J. Lin, Z. Ma, R. Gomez, K. Nakamura, B. He, and G. Li, "A review on interactive reinforcement learning from human social feedback," IEEE Access, vol. 8, pp. 120757–120765, 2020.
27. A. A. Ahmed, S. Sayed, A. Abdoulhalik, S. Moutari, and L. Oyedele, "Applications of machine learning to water resources management: A review of present status and future opportunities," J. Clean. Prod., p. 140715, 2024.
28. S. Dhanabal and S. Chandramathi, "A review of various k-nearest neighbor query processing techniques," Int. J. Comput. Appl., vol. 31, no. 7, pp. 14–22, 2011.
29. C. Banapuram, A. C. Naik, M. K. Vanteru, V. S. Kumar, and K. K. Vaigandla, "A Comprehensive Survey of Machine Learning in Healthcare: Predicting Heart and Liver Disease, Tuberculosis Detection in Chest X-Ray Images," SSRG Int. J. Electron. Commun. Eng., vol. 11, no. 5, pp. 155–169, 2024.
30. Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," Expert Syst. Appl., vol. 237, p. 121549, 2024.
31. A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," Reliab. Eng. Syst. Saf., vol. 233, p. 109126, 2023.
32. J. S. Bowers et al., "Deep problems with neural network models of human vision," Behav. Brain Sci., vol. 46, p. e385, 2023.
33. O. Hrizi et al., "Tuberculosis disease diagnosis based on an optimized machine learning model," J. Healthc. Eng., vol. 2022, no. 1, p. 8950243, 2022.
34. Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. Health & Social Care in the Community, 2023(1), 1406060.
35. Qureshi, M., Khan, S., Bantan, R. A., Daniyal, M., Elgarhy, M., Marzo, R. R., & Lin, Y. (2022). Modeling and forecasting monkeypox cases using stochastic models. Journal of Clinical Medicine, 11(21), 6555.



Vol. 3 No. 1 (2025): January - March
https://rjnmsr.com/index.php/rjnmsr/about

Review Journal of Neurological & Medical Sciences Review

E(ISSN) : 3007-3073
P(ISSN) : 3007-3065

36. Iftikhar, H., Daniyal, M., Qureshi, M., Tawiah, K., Ansah, R. K., & Afriyie, J. K. (2023). A hybrid forecasting technique for infection and death from the mpox virus. Digital Health, 9, 20552076231204748.
37. Daniyal, M., Qureshi, M., Marzo, R. R., Aljuaid, M., & Shahid, D. (2024). Exploring clinical specialists' perspectives on the future role of AI: evaluating replacement perceptions, benefits, and drawbacks. BMC Health Services Research, 24(1), 587.
38. Qureshi, M., Daniyal, M., & Tawiah, K. (2022). Comparative Evaluation of the Multilayer Perceptron Approach with Conventional ARIMA in Modeling and Prediction of COVID-19 Daily Death Cases. Journal of Healthcare Engineering, 2022(1), 4864920.
39. Qureshi, M., Ishaq, K., Daniyal, M., Iftikhar, H., Rehman, M. Z., & Salar, S. A. (2025). Forecasting cardiovascular disease mortality using artificial neural networks in Sindh, Pakistan. BMC Public Health, 25(1), 34.
40. Hussain, I., Qureshi, M., Ismail, M., Iftikhar, H., Zywiołek, J., & López-Gonzales, J. L. (2024). Optimal features selection in the high dimensional data based on robust technique: Application to different health database. Heliyon, 10(17).
41. Yousfani, K., Qureshi, M., Daniyal, M., & Ismail, M. (2024). Comparative Study of Machine Learning (ML) and Conventional Time Series Methodologies in Modelling the Exports Trade of Pakistan. Indus Journal of Social Sciences, 2(2), 349-367.