Remittances Review
April, 2024
Volume: 9, No: 2, pp.3397-3412
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Received: 05 February 2024, Accepted: 05 April 2024
DOI: https://doi.org/10.33282/rr.vx9i2.174

Efficient Mechanism for the Prediction and Analysis of Disease Outbreak – Study based on Urban Area of Pakistan

Muhammad Zulqarnain Siddiqui[1] Associate Professor,
Department of Computer Science, Saifee Burhani University, Karachi
zulqarnain@sbu.edu.pk

M Zamin Ali Khan [2]
Professor/HOD Department of Computer Science, Iqra University, Karachi
muhammad.zamin@iqra.edu.pk

Khalid Bin Muhammad[3]
Assistant Professor, College of Computer Science and Information Systems, IOBM, Karachi
kbmcbm@iobm.edu.pk

Muhammad Asghar Khan[4]
Assistant Professor, College of Computer Science and Information Systems, IOBM, Karachi
muhammad.asghar@iobm.edu.pk

Asim Iftikhar[5]
Assistant Professor, College of Computer Science and Information Systems, IOBM, Karachi
asim.iftikhar@iobm.edu.pk

Javeria Barkat[6]
Senior Lecturer, Department of Computer Science, Iqra University, Karachi
javeria.barkat@iqra.edu.pk

ABSTRACT

This research project aims to revolutionize the availability of disease information by developing a state-of-the-art mobile application. The main goal is to provide users with predictive information about disease occurrence in certain areas. Using advanced data analysis techniques and machine learning algorithms, the application aims to enable people to proactively predict future diseases, facilitating informed decision making and proactive measures for personal health. The initiative goes beyond mapping and forecasting to emphasize proactive actions and community participation. An important part is the development of educational resources and community-based campaigns to increase awareness of disease prevention, symptom recognition and access to health services. The goal of this collaborative effort is to

empower communities to take proactive steps to prevent disease and maintain health. At the heart of the mobile application is a comprehensive mapping function that visually presents diseases and trends in different parts of the city. Using various data sources such as historical records, user-generated reports, the application undergoes careful data processing and analysis. These processes reveal valuable patterns, paving the way for the development of accurate machine learning models to predict disease in specific regions. This predictive capability not only benefits individuals by helping them take precautions, but also helps public health agencies and healthcare providers optimize resource allocation and target interventions. User experience continues to be critical to application and development. The user interface prioritizes intuitiveness, user-friendliness and aesthetic appeal. Users can interact with the map, explore different areas of the city and use filter options to focus on specific diseases. High-risk areas and related health advisories, informing users and enabling rapid action.

Keywords
Disease Mapping, Predictive Modeling, Infectious Diseases, Data Analysis. Healthcare Management

1. INTRODUCTION

The incessant quest for technological progress and its implementation in the healthcare sector has initiated a period of groundbreaking approaches with the objective of improving the health and welfare of the public. The creation of a mobile application that offers users an all-encompassing comprehension of disease prevalence in different urban areas serves as evidence of the dynamic nature of health technology. The objective of this study is to examine the complex design, development, and features of this innovative mobile application in order to shed light on its crucial contribution to the transformation of public health awareness and proactive health management.

The fundamental objective of this endeavor is to tackle the urgent requirement for easily obtainable and implementable health information among urban residents. By capitalizing on the capabilities of machine learning algorithms and data analysis, the mobile application endeavors to provide a dynamic and interactive environment. The aforementioned platform furnishes users with up-to-date information regarding the prevalence and dispersion of diseases across various urban environments. Through the integration of sophisticated technologies and intuitive interfaces, the application aims to provide users with up-to-date and relevant health-related information.

The undertaking holds great importance due to its potential to provide individuals with predictive disease analytics, which would empower them to proactively manage their health and make well-informed decisions. By utilizing the app's user-friendly interface, individuals are able to effortlessly traverse the urban map while also perceiving the prevalence of diseases in different areas via color-coded depictions. This visual representation functions as an essential instrument for promoting public health consciousness and cultivating a proactive stance towards healthcare management.

The primary objective of this report is to provide an exhaustive analysis of the various functionalities of the mobile application and its profound implications for public health. Through a detailed analysis of the complexities inherent in its design and development, it seeks to highlight the transformative potential of this innovative tool. Moreover, it strives to emphasize the application's function in fostering a climate of health awareness and proactive health conduct among its users, thus making a substantial contribution to the overarching objective of enhancing health outcomes within the community.

By shedding light on the app's predictive disease analysis and mapping capabilities, this research aims to underscore its potential to revolutionize health-related decision-making and empower individuals to take charge of their well-being.

The background for this research project originates from the inherent challenges in accessing timely and predictive disease information, particularly in urban areas. Recognizing the limitations of traditional disease tracking methods, the endeavor to develop a state-of-the-art mobile application emerged. Urban environments present unique complexities in disease management, prompting the need for innovative solutions. This choice of topic was motivated by the potential of technology, specifically advanced data analysis and machine learning, to provide proactive disease predictions. The goal is to bridge the gap between available health information and actionable insights, aiming to empower individuals and communities towards proactive health management. Emphasizing community participation and education, the project aspires to revolutionize disease prevention strategies and healthcare resource allocation, contributing to improved public health outcomes in urban settings.

The primary objectives of this research project are to develop a mobile application integrating advanced data analysis and machine learning algorithms for predicting disease occurrences in specific urban areas. This initiative aims to empower users with proactive health insights, foster community participation in disease prevention, and optimize resource allocation for healthcare interventions through accurate disease prediction models. Additionally, this research project seeks to enhance user experience by prioritizing intuitive interface design and efficient data visualization features within the application.

2. Literature Review

A notable occurrence of typhoid fever emerged in Ghannouche, a municipality situated in the southern region of Tunisia, in July 2016. The affected population constituted 16.24% of the population under investigation. Transmission pathways and risk factors for effective control measures were the objectives of the study. Males (56%) and children under the age of ten (38.1%) comprised the majority of the 102 cases identified; fever (95%) and diarrhea (57%) were the most commonly observed symptoms. Various housing types, youth, and inadequate education were found to be significantly correlated with heightened risk. Typhoid fever was also associated with the consumption of softened water, home-grown raw fruits and vegetables, and improper refuse disposal. Two asymptomatic carriers and non-compliant softened water samples were discovered during the investigation, highlighting the criticality of water and food safety in the prevention of such outbreaks. Implementation of recommendations to prevent and control the disease's further spread. [1]

The article discusses the rise in the prevalence of mental health disorders such as anxiety and depression, especially since the outbreak of the COVID-19 pandemic. The article also highlights the limitations of traditional diagnostic practice and low help-seeking behavior due to various factors. The use of machine learning models in detecting anxiety and depression through social media data is explored. The article reports the most common machine learning models identified and the promising findings for AI-technology based predictive models in the area of mental health. The article also

discusses the value of insights into mental health risk factors and population-level disclosure of mental health symptoms on social media platforms that ordinarily may not be disclosed. Finally, the article emphasizes the need for social media studies to be collated and reviewed following PRISMA guidelines. [2]

This systematic review analyzed 90 studies on machine learning techniques used in predicting type 2 diabetes. The review found that the structure of the dataset is crucial for model accuracy, and tree-based models showed optimal performance. However, complementary techniques such as K nearest neighbors and Support vector machines are needed to balance data and reduce dimensionality. Deep Neural Networks can handle big data but require datasets with over 70,000 observations. The review also recommended reporting at least three performance metrics, including AUC (ROC), to reduce heterogeneity in the performance comparison. The lack of transparency in feature selection remains a challenge to model interpretability. The review provides insight into the opportunity areas in diabetes prediction using machine learning techniques. [3]

The author proposes a machine learning-based approach for the identification and prediction of chronic diseases. The study focuses on the use of electronic health records (EHRs) for the prediction of chronic diseases such as diabetes, hypertension, and chronic kidney disease.

The paper begins by discussing the prevalence and impact of chronic diseases on healthcare systems, highlighting the need for early identification and intervention. The author then presents a review of relevant literature on machine learning-based approaches for disease prediction, identifying key challenges and limitations of current methods.

Next, the author presents their proposed framework for disease prediction, which involves data preprocessing, feature selection, model training and testing, and prediction of disease risk. The study uses a dataset of patients' EHRs, which includes demographic information, medical history, laboratory results, and medication records.

The results of the study demonstrate the effectiveness of the proposed machine learning approach for disease prediction, with high accuracy rates for the identification of patients at risk for chronic diseases. Overall, the study highlights the potential of machine learning in improving disease identification and prediction, and it’s potential to revolutionize healthcare delivery. [4]

It is a comprehensive review that summarizes the recent developments and advancements in machine learning-based disease diagnosis. The paper starts with an introduction to machine learning and its applications in healthcare, followed by a discussion of the challenges and opportunities associated with using machine learning for disease diagnosis.

The authors then review various machine learning techniques and algorithms, including support vector machines, decision trees, random forests, and deep learning, and

discuss their applications in disease diagnosis.

The authors highlight the potential of machine learning algorithms to improve the accuracy and efficiency of disease diagnosis, as well as the challenges and limitations in their implementation. They discuss the various types of machine learning algorithms and their applications in different disease diagnosis areas, such as cancer, cardiovascular disease, neurological disorders, and infectious diseases.

Overall, this paper provides a comprehensive overview of the current state of the art in machine learning-based disease diagnosis and highlights the potential of this technology to revolutionize healthcare by providing accurate and personalized diagnosis and treatment. [5]

The aims to predict the spread and severity of COVID-19 using machine learning techniques. The authors collected data from various sources such as the World Health Organization (WHO), John Hopkins University (JHU), and various government websites. They then used machine learning algorithms such as support vector machines (SVM), random forests, and artificial neural networks (ANN) to predict the number of confirmed cases, deaths, and recoveries.

The authors found that the SVM algorithm was the most accurate in predicting the number of confirmed cases, deaths, and recoveries. They also found that the ANN algorithm was the most accurate in predicting the severity of the pandemic. The authors concluded that machine learning algorithms can be useful in predicting the spread and severity of pandemics such as COVID-19.

Overall, the paper provides valuable insights into the use of machine learning in predicting the spread and severity of infectious diseases such as COVID-19. The authors' approach is thorough and comprehensive, and their findings suggest that machine learning techniques can be an effective tool in forecasting and predicting pandemics.[6]

It provides a comprehensive analysis of the application of machine learning techniques for the prediction and management of infectious diseases. The study reviewed 80 articles published between 2016 and 2021 that utilized machine learning algorithms to predict the outbreak, spread, and severity of infectious diseases.

The review revealed that machine learning models have shown promising results in predicting outbreaks and the spread of infectious diseases such as COVID-19, Ebola, and Dengue fever. The models were used to predict transmission rates, identify high-risk populations, and optimize resource allocation for disease control.

The study also highlighted the challenges associated with using machine learning for infectious disease prediction. These challenges include data quality issues, ethical concerns related to the use of personal data, and the need for real-time data.

Overall, the paper suggests that machine learning has the potential to significantly improve the prediction and management of infectious diseases, but more research is

needed to overcome the challenges associated with its implementation. [7]

The text discusses the challenge of machine-learning models in medical imaging tasks when faced with scenarios differing from the training dataset. The authors propose a strategy called REMEDIS (Robust and Efficient Medical Imaging with Self-supervision) to address the 'out of distribution' performance issue. REMEDIS combines supervised transfer learning on natural images with contrastive self-supervised learning on medical images, requiring minimal taskspecific customization. The approach significantly improves model robustness and training efficiency. In various diagnostic-imaging tasks across different domains, REMEDIS outperforms strong supervised baseline models, achieving up to an 11.5% improvement in diagnostic accuracies. In out-of-distribution scenarios, REMEDIS requires only 1-33% of the data for retraining to match the performance of supervised models using all available data. The strategy shows promise in accelerating the development of machine-learning models for medical imaging. [8]

The research focuses on the increasing prevalence of obesity-related conditions. Utilizing a Positive and Unlabeled (PU) machine learning method, the study achieved a 98% sensitivity in identifying obese patients without a specific diagnosis code. Notably, around 18% of patients without a formal diagnosis were predicted to be obese, consistent with WHO reports. This underscores the machine learning's potential in identifying obesity and related health issues in healthcare records, providing valuable insights for effective intervention and management. [9]

This study addresses the increasing global impact of breast cancer, a leading cause of death among women. Recognizing the critical need for accurate prediction and early diagnosis, the research focuses on leveraging machine learning techniques for these purposes. Five algorithms— Support Vector Machine (SVM), Random Forest, Logistic Regression, Decision Tree (C4.5), and K-Nearest Neighbors (KNN)—are applied to the Breast Cancer Wisconsin Diagnostic dataset. The study includes a thorough performance evaluation and comparison of these classifiers. The primary goal is to predict and diagnose breast cancer, emphasizing metrics like confusion matrix, accuracy, and precision. The results indicate that Support Vector Machine stands out as the most effective, achieving the highest accuracy at 97.2%. The entire analysis is conducted using the Anaconda environment with Python programming and the scikit learn library. This research contributes to the advancement of breast cancer diagnosis through machine learning, offering insights for improved treatment and patient survivability standards.[10]

The paper addresses the growing volume of critical healthcare data and its potential for early detection and prevention of life-threatening diseases, focusing on diabetes. A machine learningbased approach is proposed for the classification, early-stage identification, and prediction of diabetes. Three classifiers—random forest (RF), multilayer perceptron (MLP), and logistic regression (LR)—are utilized, with MLP outperforming others at 86.08% accuracy. For predictive analysis, long short-term

memory (LSTM), moving averages (MA), and linear regression (LR) are employed, with LSTM achieving a significant 87.26% accuracy. The study also introduces an IoTbased hypothetical diabetes monitoring system. Experimental evaluation using the PIMA Indian Diabetes dataset demonstrates the effectiveness of the proposed approach, positioning it as adaptable in various public healthcare applications. The research contributes to leveraging intelligent data analysis for disease prevention and early detection. [11]

The study addresses the need for rapid malaria diagnosis and explores a novel approach by developing a machine learning model using patient information extracted from PubMed abstracts spanning 1956 to 2019. Two datasets were constructed, one solely focused on parasitic diseases and another encompassing information about various diseases. Six machine learning models were compared, including support vector machine, random forest, multilayered perceptron, AdaBoost, gradient boosting, and CatBoost. The synthetic minority oversampling technique (SMOTE) was employed to tackle data imbalance issues. Results indicated that, for the solely parasitic disease dataset, random forest performed best, while gradient boosting excelled for the total dataset. Notably, random forest outperformed after applying SMOTE to the total dataset. Nationality emerged as a crucial feature in imbalanced data, while symptoms took precedence in balanced data with SMOTE. The findings underscore the successful application of machine learning techniques in predicting malaria using patient information. Keywords: Case reports; Diagnosis; Machine learning; Malaria; Patient information. [12]

The study explores the potential of applying deep learning convolutional neural network approaches to predict the geographic origin of Plasmodium falciparum and Plasmodium vivax infections using whole genome sequencing (WGS) data. With datasets from multiple countries, the analysis of population structure and ancestry using high-quality single nucleotide polymorphisms (SNPs) revealed clustering at the country level. Classification methods demonstrated the lowest distance errors, achieving over 90% accuracy at a country level for both species. The findings underscore the utility of machine learning for geo-classification of malaria parasites, emphasizing the importance of timely and extensive WGS data for enhanced disease control activities. [13]

This research focuses on the escalating water quality issues in Pakistan due to rapid population growth. Using patient data from Ayub Medical Hospital (2017-2020), the study employs machine learning models, with Random Forest proving superior in predicting waterborne diseases, achieving 60% accuracy for malaria and 77% for typhoid cases through tenfold cross-validation.

The study delves into the significance of input features, revealing age, history, and test results as pivotal factors in waterborne disease prediction. By employing the random forest feature selection technique, the research provides valuable insights for health departments, suggesting a potential reduction in waterborne diseases by focusing on specific risk factors.

In conclusion, the study underscores the efficacy of machine learning, particularly Random Forest, in predicting and mitigating waterborne diseases, emphasizing the importance of targeted input features for accurate and actionable predictions in public health efforts. [14]

The lack of accurate typhoid diagnostics hampers regional burden estimates and affects government interventions. Improved diagnostic capacity, including access to advanced test kits, is crucial for effective disease control and vaccine implementation in endemic regions. Multidrug resistance, facilitated by global travel, poses challenges in clinical management. The review explores diagnostic gaps, ongoing developments, and challenges, emphasizing the need for affordable, efficient tools. Current investigations include RDTs with selected antigens, metabolomic platforms, and machine-learning-based gene expression signatures. PCR-based detection shows promise, but existing tools fall short in resource-poor settings. Molecular approaches for environmental surveillance offer potential for public health monitoring. Balancing low-cost simplicity and advanced technology is crucial for the future diagnostic landscape, combining high-tech surveillance with improved RDTs for individual patient management. [15]

This Systematic Literature Review (SLR) focuses on dengue modeling, which has gained prominence due to the need for early prediction and decision-making to control the disease. The review categorizes modeling approaches into diagnostic, epidemic, and intervention, emphasizing prediction, prescription, and optimization models. Sixty-four articles were analyzed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology.

Results indicate that logistic regression is predominantly used (59.1%) for diagnosing dengue. In the epidemic approach, spatial analysis is commonly performed using linear regression (17.4%).
For intervention modeling, the General Linear Model is prevalent, accounting for 70%.

The study concludes that cause-effect models can enhance dengue diagnosis and comprehension. Models addressing uncertainty are crucial due to low data quality in healthcare. Additionally, decentralization of data through federated learning is recommended to reduce computational costs and facilitate model building while ensuring data security. [16]

This review paper emphasizes the global public health challenge posed by Dengue disease and explores the application of various machine learning (ML) algorithms for predicting Dengue spread. ML techniques, including Self-Organizing Maps, Decision Trees, and neural networks, demonstrate high predictive accuracy, enabling rapid assessment for targeted interventions. Collaboration between health and data scientists is crucial for innovative methods in diagnosis, treatment, and prediction. The paper advocates for incorporating geographic variables and local data within the ML framework to enhance accuracy. Ultimately, ML approaches may inform policies and contribute to Dengue eradication efforts, addressing the rising number of cases worldwide. [17]

This study employs machine learning to enhance predictive models for monitoring dengue transmission rates, using vector indices and meteorological data. Ensemble ML methods, including XG Boost, AdaBoost, and Random Forest, outperform other algorithms, with XGBoost showing the highest AUC, accuracy, and F1 score. Removing the least important variable, the container index, improves model performance by at least 6% in AUC and F1 score. The study emphasizes the increasing global magnitude of dengue transmission and contributes to a more comprehensive understanding through a robust dataset, setting the stage for future research on predictive models and early warning systems. [18]

This study employs Machine Learning techniques, including logistic regressions, linear discriminant analyses, Naive Bayes, decision tree, and random forest classifiers, to assess dengue fever risk in São Luís do Maranhão, Brazil. Utilizing data from 2014 to 2020, including cases per neighborhood and climatic, territorial, and environmental variables, oversampling techniques address the unbalanced database. The Random Forest classifier, trained with DBSMOTE oversampling, achieves the best results, with a 75.1% AUC, 75.43% sensitivity, and 60.53% specificity. The broader context discusses the application of machine learning algorithms in analyzing and predicting communicable diseases, focusing on the dengue virus's expansion, vector behavior, and spatial transmission dynamics. [19]

This perspective article addresses the ongoing challenge of persistently high HIV incidence rates despite efforts to reduce HIV-related mortality. The UNAIDS Fast-track targets and preventative strategies like pre-exposure prophylaxis are crucial, and accurate risk prediction is essential for their success. Machine learning and artificial intelligence algorithms, known for capturing complex data interactions, prove effective in predicting HIV infection risk. However, the interpretability of these algorithms poses a challenge. The article introduces machine learning, discusses considerations in variable selection and algorithm evaluation, and explores emerging tools like Shapely Additive Explanations for model understanding in the context of HIV. Additionally, it highlights potential public health and clinical applications for risk assessment models, emphasizing integration synergies with algorithms predicting sexually transmitted infections and tuberculosis.[20]

3. METHODOLOGY

In this research efficient methodology is utilized to execute, which includes a comprehensive data collection strategy. Information is gathered from a variety of sources, including hospital laboratories and clinics, to acquire valuable insights using this method. The comprehensive dataset comprises historical records, demographic information, and environmental factors that exhibit a strong correlation with disease incidence. Prior to exploratory analysis, a preprocessing phase ensures the data's quality and consistency by uncovering critical correlations and patterns that are necessary for making informed decisions. Sophisticated machine learning algorithms are employed within the methodology to generate predictive models. These models ascertain significant determinants that contribute to the development of diseases by employing stringent methodologies for

feature selection and engineering. The integration of these predictive models and advanced visualization functionalities into a user-friendly mobile application developed with React Native establishes the technical framework. Constant adaptation of the application's user interface to user requirements is ensured through the continuous development of the interface, which is fueled by ongoing user feedback and comprehensive usability tests. Furthermore, the methodology incorporates

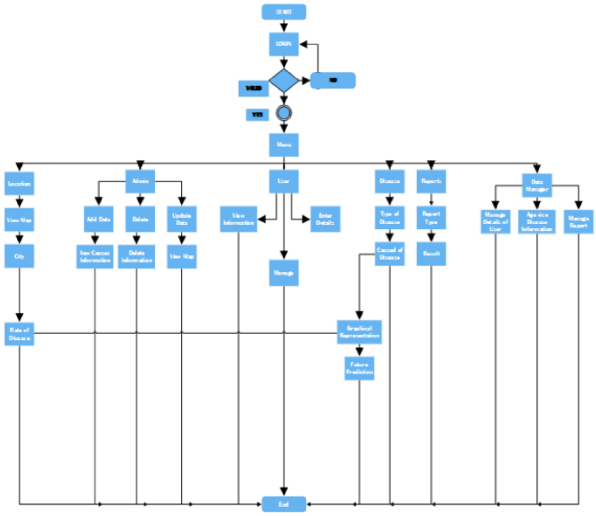

figure 3.1 **Process Flow Diagram**

Fig **3.1** illustrated the ordered progression of the research project starting from its inception. The user, administrator, or data manager will initially log in with their identification and password; if the login is successful, they will proceed; otherwise, they will be redirected to the login page. Then, the location option will appear on the main menu, accompanied by a map from which the user will select the country, followed by the city, and finally the disease rate will be displayed. Administrators will now have the ability to add, modify, or remove their information or details. Users will have the ability to access their information and provide additional details.

educational and community engagement initiatives alongside technical components. These endeavors are designed to foster consciousness and stimulate active participation, thereby augmenting the impact of the application. The enduring significance and flexibility of the application are validated through ongoing monitoring, evaluation, and documentation procedures, which facilitate consistent improvement and a lasting impact on urban disease prevention and public health consciousness.

The development process for the Disease Mapping App involves several key steps.

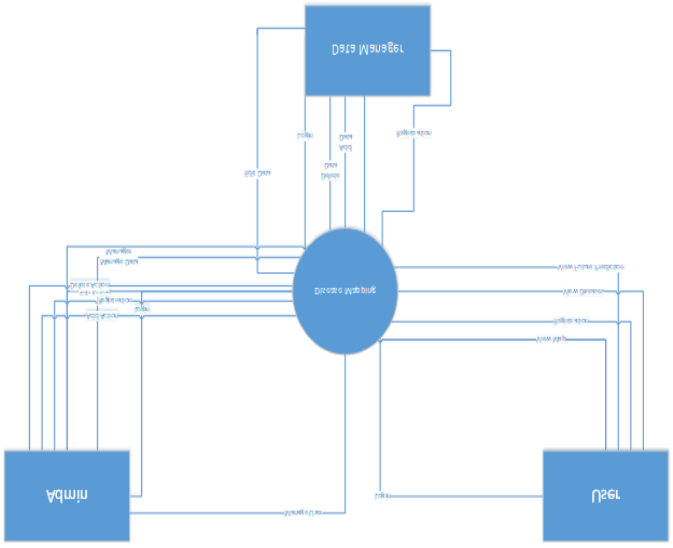

Figure 3.2 Data Flow Diagram

Figure **3.2.** Showed the flow of data of admin, user and data manager with the actions which they can perform for example the user can login, view map, register themselves, view disease and view future description.

Remittances Review
April, 2024
Volume: 9, No: 2, pp.3397-3412
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

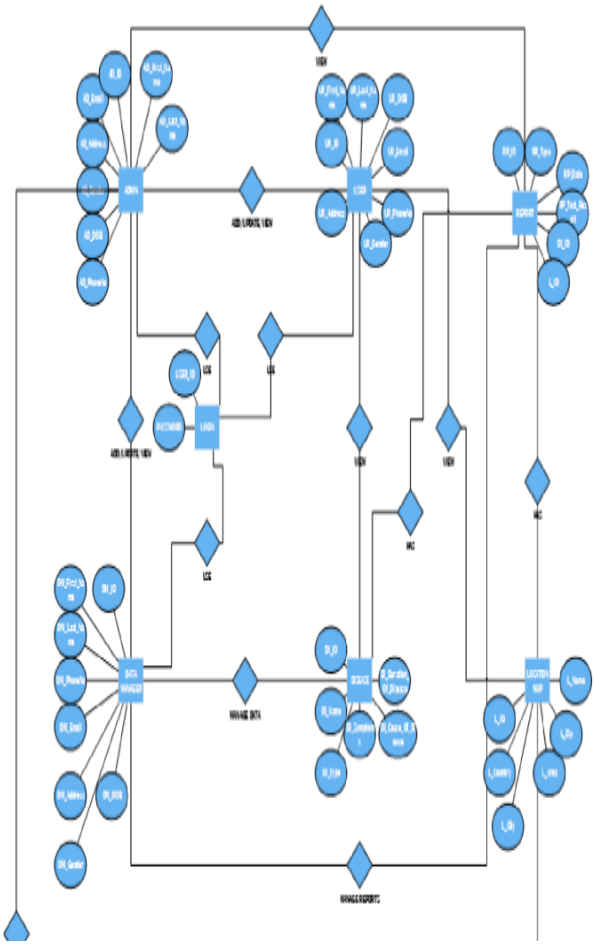

Figure 3.3. ERD Diagram

Figure **3.3.** identified the relationships between the six entities in the ERD diagram. Administrator, user, data manager, report, disease, and map are all entities in our system. Additionally, the potential attributes of each entity have been delineated.

3.2 Key Features

The development process for the Disease Mapping App involves several key steps.

3.2.1 Research and Data Collection

- Collect complete data on the prevalence of diseases from reputable sources spanning multiple regions of the city.
- Determine significant diseases that impact various regions, taking into account historical trends and patterns.

3.2.2. Data Analysis and Visualization

- Apply data analysis methodologies to methodically organize and visually represent disease patterns along with their spatial distribution.
- Using Geographic Information System (GIS) tools, generate disease prevalence-based visual representations (maps).

3.2.3. Machine Learning Model Development

- Construct predictive models for future disease cases utilizing machine learning algorithms (e.g., Random Forest, Decision Trees).
- To improve the accuracy of predictions, train these models utilizing historical disease data and geographical features.

3.2.4. Integration with Mobile Application

- Construct a mobile application framework that incorporates machine learning models in order to furnish users with prognostic insights pertaining to disease outbreaks.

3.2.5. App Interface Design

Construct a simple user interface that includes:

- City Map Display: Utilizing color-coded representations to emphasize regions displaying different diseases.
- Disease Data: Providing comprehensive information regarding the prevalence of specific diseases in urban regions.
- User Interaction: Enabling user engagement with the disease data and map to enhance knowledge and gain deeper insights.

3.2.6. User Testing and Feedback

- In order to obtain feedback on the usability, functionality, and information presentation of the application, it is recommended to organize user testing sessions.
- Incorporate enhancements and suggestions from users to enhance the functionality and user experience of the application.

3.2.7. Limitations and Improvement Considerations

- Recognize possible limitations, including user interface preferences, data precision, and model prediction accuracy.

- Consistently enhance and revise the application in accordance with evolving disease data, user input, and technological developments.

3.2.8. Security and Privacy Measures

- Throughout app usage, implement strict security measures to safeguard sensitive user data and ensure user privacy.

4. RESULT

In this research, the health data for Karachi in the present year reveal important trends related to a range of diseases. A total of 14,890 cases of malaria have been reported, of which 7,624 were confirmed positives. Monthly variances were noted, with September showing the maximum number of reported cases, which incited more research. The gender distribution of the cases revealed that 51.16% were male and 48.84% were female. Additionally, 40% of the cases involved individuals under the age of 18, which underscores the importance of implementing targeted interventions, especially in September. A total of 9,500 documented cases of diabetes were identified, which corresponds to a prevalence rate of 71.05%. December had the maximum number of reported cases, with males having a higher prevalence rate of 55.68%. Significantly, 65% of diabetes cases were observed in individuals aged 40 years and older, underscoring the importance of age in disease occurrence. The gender distribution of the 1,068 dengue cases in Karachi was as follows: individuals aged 15 to 75 were primarily affected, with females comprising 18.9% and males comprising 81.1%. Indicating that 75% of the 25,932 dengue cases reported by the WHO occurred between June and September, this critical period requires immediate and effective measures to contain the disease's transmission. Regarding of sexually transmitted diseases, a total of 9,029 reported cases of HIV resulted in a positive rate of 2.15%. The findings revealed that 61% of the cases involved females, with the majority transpiring among individuals aged 1 to 26 years; this underscores the criticality of implementing focused interventions. An observable temporal pattern emerged, with October being particularly significant. This indicates that during identified high-risk periods, increased vigilance and interventions are required. In fig typhoid cases revealed demographic and seasonal patterns. In Fig 4.2.1(d) 36.8% positive cases, the spring season produced the most, followed by the summer with 25.6% cases. There was significant variation in age distribution, extending from one to sixty years, with an average age of 18.76 years.In Fig 4.2.1(a) males was greater as compared to females. In fig 4.2.1(b) Orangi, Jamshed, and Lyari were identified through geographic analysis as the region most significantly affected, thereby giving crucial data for the creation of targeted public health interventions in particular geographic regions and demographic groups

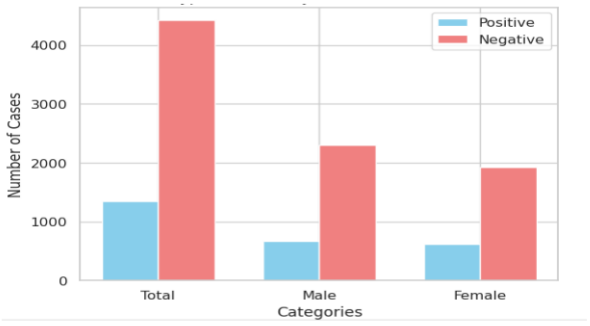

Figure 4.1

In **Fig 4.** The graph depicts the distribution of positive and negative typhoid cases by gender. The figure shows 4,000 good and

1,000 negative cases. Positive instances occur at a higher rate in both male and female populations than negative cases. However, the offered graphic does not provide precise breakdowns for male and female positive and negative cases.

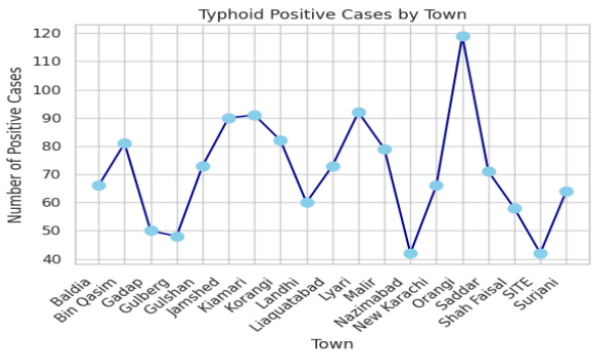

Figure 4.2

In **fig 4.2** Orangi, Jamshed, and Lyari were identified through geographic analysis as the region most significantly affected, thereby giving crucial data for the creation of targeted public health interventions in particular geographic regions and demographic groups.

5. CONCLUSION

The Disease Mapping and Prediction research project is an innovative initiative that aims to improve healthcare management and address health concerns by utilizing technology. The combination of remarkable technological advancements and the imminent threat of infectious diseases highlights the crucial requirement for innovative solutions. This program aims to transform disease surveillance and prediction by using a mobile application that incorporates sophisticated data analysis, machine learning, and user involvement.

The methodology devised for this research project demonstrates a thorough and systematic approach. Starting with the initial phase of data collecting, which entails collaborating with hospitals, clinics and Laborites to gather a comprehensive dataset encompassing a range of diseases including Typhoid, HIV, Malaria, Dengue and Diabetes. Every step is carefully executed to establish a strong and reliable groundwork. This includes thorough data preparation to ensure accuracy and compatibility. The focus on exploratory data analysis, geographic mapping, and statistical modeling displays a dedication to obtaining significant insights from the gathered data. An important aspect of the research is the use of predictive modeling, demonstrated through a case study on the frequency of disease in Karachi. Typhoid was taken as a model disease to predict and analyze disease dataset. The Random Forest Regression model and Linear Regression Framework demonstrate their capacity to foresee future disease occurrences, offering useful insights for proactive health management. Libraries like Pandas, Numpy Matplotlib and Sklearn played a vital role in making predictive models efficient. The visual depiction of projected versus observed instances highlights the efficacy of the model in discerning patterns, seasonal fluctuations, and distinct local tendencies.

This research project exemplifies the potential of technology to revolutionize public health by visualizing a healthier and safer environment, where diseases lag us. The project's comprehensive approach, creative techniques, and dedication to ongoing enhancement pave the way for a future in which preventive healthcare becomes a tangible and attainable reality for all individuals.

REFERENCES

[1] Hechaichi, A., Bouguerra, H., Letaief, H., & Safer, M. (2023). Outbreak Investigation of typhoid Fever in the District of Gabes, South of Tunisia. Epidemiologia, 4(3) https://www.mdpi.com/2673-3986/4/3/23

[2] Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Abu Serhan, H., Abd-alrazaq, A. A., & Househ, M. (2022). Machine learning models to detect anxiety and depression through social media. Elsevier, 2, 100066 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9461333/

[3] Fregoso-Aparicio, L., Noguez, J., Montesinos, L., & García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes. Diabetology & Metabolic Syndrome, 13(1), 148 https://dmsjournal.biomedcentral.com/articles/10.1186/s13098-021-00767-9#Sec1

[4] Alanazi, R. (2022). Identification and Prediction of Chronic Diseases Using Machine Learning Approach. Journal of Healthcare Engineering, 2826127. https://www.hindawi.com/journals/jhe/2022/2826127/

[5] Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022).Machine-Learning-Based Disease Diagnosis. Healthcare (Basel), 10(3) https://pubmed.ncbi.nlm.nih.gov/35327018/

[6] Painuli, D., Mishra, D., Bhardwaj, S., & Aggarwal, M. (2021).Forecast and prediction of COVID-19 using machine learning. 381(397) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8138040/

[7] Santangelo, O. E., Gentile, V., Pizzo, S., Giordano, D., & Cedrone, F. (2023). Machine Learning and Prediction of Infectious Diseases. Journal Name, 5(1) https://www.mdpi.com/2504- 4990/5/1/13

[8] Laura Culp, Jan Freyberg, & Basil Mustafa, 7 (2023), Machine learning for diagnostic imaging. https://www.nature.com/articles/s41551-023-01049-7

[9] Barnardo Valdivieso, Ruth Vilar ,Vicent Blanes-Selva, (2020).Machine Learning-Based Identification of Obesity from Positive and Unlabelled Electronic Health Records. 10.3233 https://pubmed.ncbi.nlm.nih.gov/32570505/

[10] Mohammad Amin Naji , Sanaa El Filali , Kawtar Aarika , (2021). Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, 191 https://www.sciencedirect.com/science/article/pii/S1877050921014629

[11] Umair Munner Butt, Mubashir Ali, Anees Baqir, (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8500744/

[12]You Won Lee, Jae Woo Choi , Eun-Hee Shin , (2021).Machine learning model for predicting malaria using clinical information. https://pubmed.ncbi.nlm.nih.gov/33290932/

[13] Wouter Deelder, Emilia Manko, Jody E. Phelan, (2022).Geographical classification of malaria parasites through applying machine learning to whole genome sequence data(12) https://www.nature.com/articles/s41598-022-25568-6

[14]Mushtaq Hussain , Mehmet Akif Cifci , Tayyaba Sehar , Said Nabi , Omar Cheikhrouhou , Hasaan Maqsood , Muhammad Ibrahim , Fida Mohammad,(2023) .Machine learning based efficient prediction of positive cases of waterborne diseases . BMC Med Inform Decis Mak, 23(1) https://pubmed.ncbi.nlm.nih.gov/36653779/

[15] Tamalee Roberts, Buddha Basnyat, Stephen Baker, Lee M Hampton, Sabine Dittrich , (2023).Current Perspectives and Future Outlooks for Product Development and Access, Diagnostics for Typhoid Fever ,10(1) https://academic.oup.com/ofid/article/10/Supplement_1/S17/7188897

[16] William Hoyos , Jose Aguilar , Mauricio Toro , (2021).Dengue models based on machine learning techniques, A systematic literature review, 119, 102157 https://www.sciencedirect.com/science/article/abs/pii/S0933365721001500

[17] Zafar Hussain, Imran Ahmed khan, Mudassar Hassan Arsala,(2023) .MACHINE LEARNING APPROACHES FOR DENGUE PREDICTION,Pakistan Geographical Review,78(1) https://www.researchgate.net/publication/372690422_MACHINE_LEARNING_APPROACHES_FOR_DENGUE_PREDICTION_A_REVIEW_OF_ALGORITHMS_AND_APPLICATIONS

[18] Song Quan Ong, Pradeep Isawasan, Ahmad Mohiddin Mohd Ngesom, Hanipah Shahar, As'malia Md Lasim & Gomesh Nair, (2023) .Predicting dengue transmission rates by comparing different machine learning models with vector indices and meteorological data, 13, 19129. https://www.nature.com/articles/s41598-023-46342-2

[19] Mateus Giesbrecht , Fernanda Paula Rocha, (2022). Machine learning algorithms for dengue risk assessment, Volume 41, article number 393. https://link.springer.com/article/10.1007/s40314-022-02101-z

[20] Joshua Fieggen ,Eli Smith ,Lovkesh Arora ,Bradley Segal ,(2022).The role of machine learning in HIV risk prediction, Volume 4 , 1062387 https://www.frontiersin.org/articles/10.3389/frph.2022.1062387/full