**Research Proposal**

**Submitted By: Syeda Annam Zohra Bokhari**

**Roll No: 23 I 7334**

**Batch: MS BA 23 Fall**

**Supervised By: Dr Abdul Wahab**

**Co Supervised By: Kashif Raza**

**Title:** Strategic Placement of Cardiac Centres: Leveraging Machine Learning for Population Health Management

**Introduction**

There are many problems in the healthcare sector of Khyber Pakhtunkhwa (KPK), Pakistan including, health needs burden, nonavailability of health facilities, and inequalities in health services. Being the most populated province of Pakistan, KPK has been through major demographic changes that require strategic planning for healthcare, mainly infrastructure development. Equitable access to health services is recognized by the World Health Organization (WHO) as one of the mainstays of health services delivery. Therefore, to meet the public health needs in KPK, it is essential to determine optimal locations for where new hospitals in KPK should be built.

Some recent studies have shown that there is a high prevalence of diseases such as cardiovascular disease and ischemic heart disease (IHD) among the people of KPK, which certainly increases the health risk of the population. The rise of Cardiovascular diseases is amplified by post pandemic along with infectious diseases. According to Global burden of disease 2019 study Pakistan’s age standardized incidence of CVD stood at 918.18 per 100000 people that is higher than the global average of 684.33 per 100000 people. CVD induced death rate is 357.88 per 100000 people compared to global average of 239.85. These facts were recorded by Zainab Samad et.al (2024, April 25) (Samad & Hanif, 2023).

According to Khan et al. (2019), IHD is present with a high prevalence in KPK (S. B. Khan, Habib, Jibran, & Ali, 2019). As the data show, certain districts are hit down far more than others and need medical care urgently. Health authorities can effectively prioritize hospital locations in areas with the highest burden of disease by mapping the prevalence of IHD. Thus, this localized approach makes sure that the healthcare facilities are accessible as well as able to tackle those health issues for which it is known.

Epidemiological evaluations also suggest that there are districts within the province that are affected by these conditions at a higher rate, highlighting the time-sensitive nature of the healthcare projects to deliver healthcare facilities with access to these facilities and comprehensive care. In addition, socioeconomic factors, including health insurance coverage, have indirect effects on hospital utilization rates and thus also impact strategic hospital placement decisions to improve healthcare access for underserved populations in Dallas County.

This literature review aims at synthesizing different studies seeking methodologies to find optimal locations for hospitals in the KPK in terms of population coverage, epidemiological data, and health service utilization patterns. Policymakers can use tools like Geographic Information Systems (GIS) and multi-criteria decision-making models to better research the needs of the population and more strategically develop new hospitals in areas as a priority. This review would contribute to the ongoing discussions on healthcare planning in KPK and discuss evidence-based guidelines for the improvement of health service delivery in the province.

A study by Marta Dell’Ovo (July 2018) addresses the complex challenge of selecting optimal sites for healthcare facilities by developing transparent, multicriteria and spatially informed decision-making framework. This incorporated GIS technology into MCDA creating MC-SDSS (Dell’Ovo, Capolongo, & Oppio, 2018).

A study by Qutaiba. A Nsaif et.al (2020) uses GIS based multi criteria analysis to identify optimal sites for a new hospital in Baquba district considering factors like proximity to main roads and population density. By creating a suitability map it classifies/categorizes potential sites into excluded low and high suitability zones helping decision makers in appropriate resource allocation (Nsaif, Khaleel, & Khateeb, 2020).

Cardiac facilities need to be strategically deployed to reduce access to healthcare inequalities and provide timely intervention which could make a dramatic difference to patient outcomes. Considering the high non-communicable disease burden in the region, it is necessary to determine the distribution of healthcare resources and to know the needed population where they could best be deployed. This proposal includes a complete examination of the existing healthcare landscape and gaps and the prediction of the birth of new cardiac care centers using predictive analytics in the synopsis.

**Literature Review:**

There is no epidemiological profile of the area where new hospitals need to be set up in KPK, Pakistan, where either population is growing or spreading in an area beyond the capacity of existing hospitals. This literature review synthesizes findings from other research studies, which created the basis for the proposed optimal framework for assessment of the need of a hospital through population coverage and the health services need recommendations set by the World Health Organization (WHO).

Jeddah City is highly urbanized; therefore  Murad et al. (2024) apply MCDM and P‐Median problem modeling to find the optimal locations to locate public hospitals and thereby identify the best sites (Murad et al., 2024). Second, the above-mentioned methodologies could easily be shifted to KPK for the purpose of proximity and accessibility to populations in need. Factors related to transportation, demography, and health needs are identified as important in the study, and the conclusion is that such criteria should be used for determining the hospital site in KPK. Through MCDM, a systematic assessment of multiple site options is performed in terms of weighted criteria, assuring that new facilities satisfy a wide variety of community demands.

According to Khan et al. (2019), IHD is present with a high prevalence in KPK (S. B. Khan et al., 2019). As the data show, certain districts are hit down far more than others and need medical care urgently. Health authorities can effectively prioritize hospital locations in areas with the highest burden of disease by mapping the prevalence of IHD. Thus, this localized approach makes sure that the healthcare facilities are accessible as well as able to tackle those health issues for which it is known.

A cross-sectional study within Peshawar city provided vital data regarding the frequency of cardiovascular diseases with the highest incidences among some of the groups carried out by Haq et al in 2021. Such epidemiological patterns need to be understood to be able to plan hospitals efficaciously. The next step policymakers can take is to use these areas as areas of higher rates of cardiovascular diseases to identify areas to place hospitals in that have a higher potential to improve public health outcomes in such areas. The results of this study strengthen the case that hospital locations should coincide with the epidemiological landscape. (Haq, 2021)

Iqbal et al. (2022) in another study, found various risk factors like lifestyle and genetic predisposition in the region, associated with IHD in KPK. If these risk factors are understood, health authorities can develop targeted interventions and strategically place hospitals away from high-risk populations. This study was cross-sectional, showing a snapshot of health challenges that can be mitigated by bringing up additional health facilities.

Different methodologies to select an appropriate place for creating hospital sites are discussed in Parvin et al. (2021), and their preference is based on a structured approach (Parvin, Ali, Hashmi, & Khatoon, 2021). Moreover, they outline the significance of geographic information systems (GIS) in examining population density; already existing healthcare facilities; and patient catchment areas. Such methodologies be implemented in KPK to facilitate data-driven decisions for new hospital locations that will maximize the service coverage and accessibility

The rise of Cardiovascular diseases is amplified by post pandemic along with infectious diseases. According to Global burden of disease 2019 study Pakistan’s age standardized incidence of CVD stood at 918.18 per 100000 people that is higher than the global average of 684.33 per 100000 people. CVD induced death rate is 357.88 per 100000 people compared to global average of 239.85. These facts were recorded by Zainab Samad et.al (2024, April 25) (Samad & Hanif, 2023).

A study by Marta Dell’Ovo (July 2018) addresses the complex challenge of selecting optimal sites for healthcare facilities by developing transparent, multicriteria and spatially informed decision-making framework. This incorporated GIS technology into MCDA creating MC-SDSS (Dell’Ovo et al., 2018).

A study by Qutaiba. A Nsaif et.al (2020) uses GIS based multi criteria analysis to identify optimal sites for a new hospital in Baquba district considering factors like proximity to main roads and population density. By creating a suitability map it classifies/categorizes potential sites into excluded low and high suitability zones helping decision makers in appropriate resource allocation (Nsaif et al., 2020).

Ugochi Okengwu eta.al (June 2022) proposed a supervised machine learning approach to predict hospital locations in Africa by prioritizing specific local healthcare needs to enhance service delivery in developing countries (Okengwu, Memmert, Rein, & Osuigbo, 2022).

Asma H. Ahmed et.al (2016) developed a Multi Criteria Decision Support System by GIS and AHP to determine optimal hospital sites in Aswan, Egypt. Then classified them as good, bad and unsuitable sites for effective policy making (Ahmed, Mahmoud, & Aly, 2016).

Kahled Yousef Almansi et.al (2022) compared Machine learning and MCDA methods for hospital site suitability in Malacca, Malaysia using environmental, topographic and geodemographic factors. Results were significant in both (80 % prediction accuracy in MLP and 83 % in AHP) (Almansi et al., 2022).

Shaukat et al. (2024) seeks to achieve alignment of the new hospital site selection with the Sustainable Development Goals (SDGs) to make sure that healthcare in KPK contributes to those larger health objectives (Shaukat et al., 2024). Integrating health planning which incorporates social determinants of health are underscored. The suitability of SDG target for KPK to improve equitable health service delivery and improve health outcomes for socio-economic groups in KPK will be facilitated by having new hospitals located in places regarding the target.

As Qureshi et al. (2023) did in their systematic review, they discuss how understanding disease patterns is necessary to develop healthcare infrastructure (Qureshi et al., 2023). By determining high-burden areas, new hospitals in KPK can be established in such a way that they are handling the most critical health issues and bringing more people to the required care.

Kiran et al. (2024) conducted another study on IHD, which indicated the prevalence and the risk factors of IHD further calling for targeted healthcare strategies (Kiran, Khan, Khan, & Liaquat, 2024). Strategic allocation of new hospitals to high-risk regions can facilitate better health management, disease prevention, and identifying high-risk populations.

In addition to the previous, Khan et al. (2024) highlights the importance of the epidemiological context of IHD in Pakistan and offer the data to determine the hospital site to select (W. A. Khan, Raj, Khan, & Khan, 2024). The insights can help guide healthcare policymakers in choosing where to place facilities to optimally serve populations at high risk of heart disease.

Khurshid et al. (2023) provides another review of major trends in cardiovascular disease epidemiology and provide another view on disparities in health care across different regions (Khurshid, Awais, & Malik, 2023). In KPK, we suggest that new hospital locations should focus on areas with the highest cardiovascular disease burden.

In this systematic review, Iftikhar et al. (2023) collate data on patterns of cardiovascular disease in Pakistan (Iftikhar et al., 2023). Healthcare planners can thereby target hospital placement effectively by identifying high-prevalence areas to be able to allocate resources where resources are most needed.

Shah et al. (2024) endeavor to determine how geographic and socio-economic factors affect health service utilization in KPK (Shah & Shah, 2024). This helps indicate that there are numerous parts of the world that continue to go without any sort of primary healthcare facilities and thus need to implement more. This research will provide policymakers with insights on which hospital sites to choose in order to ensure equity in access to healthcare services for all population groups.

Collectively the reviewed literature emphasizes the need for data-based approaches for identifying the best hospital locations in KPK, Pakistan. Health facility assessments, epidemiological data, and previously established frameworks for hospital classification and site selection, when used together, can greatly expand access to systems for improving health outcomes. Placing new healthcare facilities in places where the greatest need is, KPK can enhance public health, where global health initiatives and goals for sustainable development are met. The plan is to use these strategies effectively, working on the road map to address healthcare challenges in the region.

**Problem Statement**

Cardiovascular diseases are increasing in KPK while the healthcare infrastructure to meet the increasing demand for cardiac care is lacking and is insufficient. In Pakistan, CVD is the main cause of death and noncommunicable diseases constitute approximately 29% of its death toll (Gallup Pakistan, 2023). The health care system in the region is under-resourced and needs to strategically invest in cardiac care infrastructure to tackle the problem logically. The questions this research is meant to answer are:

**Research Questions**

* Are there enough cardiac care centers currently available in KPK to meet the population’s healthcare needs?
* Where should new cardiac centers be established to maximize accessibility and minimize cardiovascular disease-related mortality?
* How many new cardiac centres to be require for maximum population coverage?

**Objectives**

The primary objectives of this study are:

* To identify underserved populations in KPK that lack access to cardiac care.
* Proposing optimal sites for new cardiac care centers using machine learning models, ensuring maximum population coverage and improved healthcare accessibility.
* Proposing adequate number of new health facilities

**WHO Sustainability Development Goals:**

In this research our objective would be to address the SDG’s defined by WHO for equal provision of health facilities and solving healthcare industry issues. These are defined as follow:

* To align the study with the United Nations Sustainable Development Goals (SDGs), particularly SDG #3 ("Ensure healthy lives and promote well-being for all")
* And SDG #10 ("Reduce inequality within and among countries").Providing the guiding light for solving healthcare industry issues.

**Scope of the Study**

This research will cover population density and areas with limited cardiac coverage in KPK. The study will analyze various factors such as population density, existing cardiac facilities and healthcare capacity, geographic dynamics of KPK district wise like road networks and highways using machine learning models to propose new cardiac care centers.

**Methodology**

1. **Data Collection:** The data will be collected from multiple sources to build, and validate a machine-learning model:
   * **Population Data & Healthcare Data::** Information from the Pakistan Bureau of Statistics on population size, age, and gender distribution across districts.

Data from the KPK Health Department, including the number of existing cardiac centers, bed capacity, and healthcare staff availability. Other Data sources include DHS data repository and Open data portal.

* + **Geospatial Data:** Geographic coordinates (latitude and longitude) of existing healthcare facilities and district boundaries to map potential areas for new cardiac centers. Moreover live data will be fetched from GIS site named as NextGIS <https://data.nextgis.com/> for KPK, Pakistan. It is offering various formats of datatypes. I will prefer CSV format.
  + **Road Networks & Accessibility:** Data from regional transportation authorities to account for road infrastructure and travel times to existing healthcare facilities. Moreover live data will be fetched from GIS site named as NextGIS <https://data.nextgis.com/> for KPK, Pakistan. It is offering various formats of datatypes. I will prefer CSV files for getting data values for Road networks.

The study is focused on following variables:

**Independent Variables:**

* + Population size per district
  + Age and gender distribution per district
  + Geographical area of each district
  + Number and capacity of current cardiac facilities
  + Road networks and travel distances
  + Population growth rates

**Dependent Variables:**

* + Number of new cardiac centers needed
  + Optimal locations for new cardiac centers
  + Optimal travel time and distance (Reduced in this term)

1. **Decision Criteria &Analytical Models:**

The literature review is abundant in research findings where a optimal solution for finding a hospital is used. This optimal solution included majorly a multiple decision criteria. This multiple Decision Criterio counts on the various factors as in my study. I would be using Population Density, Age Distribution pertaining to risk of Cardiac issues, Gender distribution, Growth Rate, Distance from Existing Cardiac facilities as the preliminary factors that would support the decision of placing a new facility as a blind spot. For ensuring the pointed facility is going to be an optimal I will go with the criteria that is based on Minimum Distance from Highways/Roads, Maximum Population Coverage as per WHO guidelines per 10000 population, Plain land.

**Geographic Information System (GIS) Software:** GIS will be used for spatial analysis and mapping of underserved areas. For the above targets I will employ GIS map analysis technique. The GIS is a significant tool to analyse maps and visualise spatially related data that assists planners and administrators to make accurate decisions. the GIS- based MCA will select a feasible site of a new hospitals KPK. Basically, the study will analyse different possible sites that will be divided into three categories; excluded, low suitability and high suitability. Interestingly, the study will identify the most suitable hospital location by investigating several factors such as existing hospitals and medical centres, distance to roads, population coverage.

This whole methodology was used Lina Zhou et.al with research paper entitled as “GIS-Based Multi-Criteria Analysis for Hospital Site Selection in Haidian District of Beijing” and also by

QUTAIBA A. NSAIF et.al with research entitled as “Integration Of Gis And Remote Sensing Technique For Hospital Site Selection In Baquba District”.

**Machine Learning Models:**

* + - **Support Vector Machine (SVM):** These classifications will utilize the levels of healthcare access, population density, and CVD incidence to classify districts using SVM. Classification will help to determine the regions requiring new centers of cardiology.
    - **Clustering Techniques/ Location-Allocation Models:** We propose that clustering algorithms like K-means will group together districts whose healthcare needs, population characteristics, and access constraints are similar and can then help find clusters of underserved populations where new cardiac center placements can be made. LAM models will identify optimal locations for healthcare facilities based on population distribution and access requirements.

The above two machine learning models will again help us in identifying the optimal location sites based on decision criteria. SVM will be used to label populations or districts with some labels proposing that they are either optimal facilities or not as Yes/No or 1/0. A research paper entitled as “Hospital Site Suitability Assessment Using Three Machine Learning Approaches: Evidence from the Gaza Strip in Palestine” by Kahled Yousaf Almansi et.al used SVM approach to nominate new hospital sites randomly and then selecting few of those based on conditioning factors used as most suitable or not suitable.

While in clustering or LAM we will use P-Median technique to classify the similar populations or districts with similar characteristic based on criteria set based on factors as similar groups. This will make us identify or highlight the best hospital location sites based on criteria and suitability scale. This method is used by Abdulkader Murad et.al in research paper entitled as “Optimizing health service location in a highly urbanized city: Multi criteria decision making and P-Median problem models for public hospitals in Jeddah City, KSA”. Similar work is evident from research entitled as “Study on the Location of Private Clinics Based on K-Means Clustering Method and an Integrated Evaluation Model” by Xiaojia Wang et.al.

Using predictive analytical approach through SVM we will present the solution for objective that states predicting the number of new facilities. For the objective with focus on identifying optimal location sites and undeserved population/districts I will use location analysis model, classification and cluster models respectively.

1. **Proposed Framework:**
   * **Data Preprocessing:** Data will be cleaned, organized, and integrated from multiple sources. Missing values will be appropriately handled to ensure data quality.
   * **Model Development:** Collected data will be used to train SVM and clustering models to predict the optimal location of new cardiac centers.
   * **Validation:** The accuracy of the models will be validated using cross-validation techniques and validated based on their ability to predict high-demand health services areas. The training set will be of 80 percent data and testing will be of 20 percent.
   * **Recommendation:** The accuracy of the models will be validated using cross-validation techniques and validated based on their ability to predict high-demand health services areas.

A diagram of data processing

Description automatically generated

Figure # 01 Summarizing Work Flow

**Significance of Study**

The study uses machine learning and data analytics to enable us to derive actionable insights into the strategic location of cardiac centers. The study will focus on the following areas of improvement:

* Efficient Resource Allocation: With the suggestion of new facilities the healthcare staff will be accessible to people more efficiently. Instead of overburdened existing sites new sites will share the resources efficiently. Healthcare industry facing the issue of cardiac disease burden and cost will be reduced.
* Healthcare Accessibility And Equity: With the new sites the access to Cardiac facilities whether these are about awareness, procedures or interventions will enhance. The cost of expensive emergent procedures will reduce and health will be promoted
* Data Driven Decision Making: Data driven approach will identify the optimal solution more accurately. The decisions are well informed and reliable.
* Cost Efficiency & Fiscal Responsibility: The strategic placement of hospital will not waste monetary values by investing on wrong location sites randomly selected. The resources and human workforce will also be placed with informed decision. The reduction of CVD procedures will reduce emergency costs beforehand.
* De Burden of CVD Disease: The disease load on healthcare machinery will reduced due to better planning and decision making.
* Improved Population Health: Population health will be improved.
* Attracting Potential Partnership and Funds: Through data driven decision making the right partnerships or funding bodies will be easily convinced to collaborate or sponsor the construction cost in the future reducing cost burden on Government machinery.

The findings will help not only the healthcare sector in KPK, but the findings will also be a model for such kinds of interventions in other parts of Pakistan as broader view.

**Conclusion**

Finally, this research proposal would like to strategically use machine learning and data analytics to uncover the best places for facilities to house cardiac care in KPK. This study tries to improve healthcare infrastructure to meet urgent needs, increase access to these essential services, lower the mortality linked to cardiovascular diseases, and contribute to the population's health management in that region.

**References**

Ahmed, A. H., Mahmoud, H., & Aly, A. M. M. (2016). Site suitability evaluation for sustainable distribution of hospital using spatial information technologies and AHP: A case study of Upper Egypt, Aswan City. *Journal of Geographic Information System, 8*(05), 578.

Almansi, K. Y., Shariff, A. R. M., Kalantar, B., Abdullah, A. F., Ismail, S. N. S., & Ueda, N. (2022). Performance evaluation of hospital site suitability using multilayer perceptron (MLP) and analytical hierarchy process (AHP) models in Malacca, Malaysia. *Sustainability, 14*(7), 3731.

Dell’Ovo, M., Capolongo, S., & Oppio, A. (2018). Combining spatial analysis with MCDA for the siting of healthcare facilities. *Land use policy, 76*, 634-644.

Haq, I. (2021). Epidemlogical Assessment and Frequency of Cardiovascular Diseases in Peshawar KP, Pakistan: A Cross-Sectional study. *Annals of the Romanian Society for Cell Biology, 25*(7), 1377-1388.

Iftikhar, H., Khan, M., Khan, Z., Khan, F., Alshanbari, H. M., & Ahmad, Z. (2023). A comparative analysis of machine learning models: a case study in predicting chronic kidney disease. *Sustainability, 15*(3), 2754.

Khan, S. B., Habib, S. A., Jibran, M. S., & Ali, I. (2019). Trends of Coronary Artery Disease in Khyber Pakhtunkhwa, Pakistan: A Retrospective Study. *Khyber Medical University Journal, 11*(4), 214-217.

Khan, W. A., Raj, H., Khan, S., & Khan, F. R. (2024). The Cost-Effectiveness of Early Invasive Procedures for Acute Coronary Syndrome in Low-Income Regions: A Prospective Cohort Study in Pakistan. *Cureus, 16*(8), e68266.

Khurshid, R., Awais, M., & Malik, J. (2023). Electrophysiology practice in low-and middle-income countries: An updated review on access to care and health delivery. *Heart Rhythm O2, 4*(1), 69-77.

Kiran, A., Khan, M. K., Khan, M. D., & Liaquat, F. (2024). Enhancing Cardiovascular Disease Risk Prediction Using Resampling and Machine Learning.

Murad, A., Faruque, F., Naji, A., Tiwari, A., Qurnfulah, E., Rahman, M., & Dewan, A. (2024). Optimizing health service location in a highly urbanized city: Multi criteria decision making and P-Median problem models for public hospitals in Jeddah City, KSA. *Plos one, 19*(1), e0294819.

Nsaif, Q. A., Khaleel, S. M., & Khateeb, A. H. (2020). Integration of GIS and remote sensing technique for hospital site selection in Baquba district. *J Eng Sci Technol, 15*(3), 1492-1505.

Okengwu, U., Memmert, D., Rein, R., & Osuigbo, E. (2022). Predictive hospital site selection model using machine learning techniques. *Scientia Africana, 21*(1), 159-174.

Parvin, F., Ali, S. A., Hashmi, S. N. I., & Khatoon, A. (2021). Accessibility and site suitability for healthcare services using GIS-based hybrid decision-making approach: a study in Murshidabad, India. *Spatial Information Research, 29*, 1-18.

Qureshi, K. A., Parvez, A., Fahmy, N. A., Abdel Hady, B. H., Kumar, S., Ganguly, A., . . . Parkkila, S. (2023). Brucellosis: epidemiology, pathogenesis, diagnosis and treatment–a comprehensive review. *Annals of medicine, 55*(2), 2295398.

Samad, Z., & Hanif, B. (2023). Cardiovascular Diseases in Pakistan: Imagining a Postpandemic, Postconflict Future. *Circulation, 147*(17), 1261-1263.

Shah, D. S. A. H., & Shah, S. S. (2024). Strategy for optimizing human capital export from Pakistan: A game-theoretic approach with a focus on Khyber Pakhtunkhwa. *Available at SSRN 4681033*.

Shaukat, M., Imping, A., Rogge, L., Khalid, F., Ullah, S., Ahmad, F., . . . De Allegri, M. (2024). Un/met: a mixed-methods study on primary healthcare needs of the poorest population in Khyber Pakhtunkhwa province, Pakistan. *International Journal for Equity in Health, 23*(1), 190.